



Πανεπιστήμιο Πειραιώς

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

Θέματα Στατιστικής Θεωρίας Πληροφοριών  
και Αναλογισμού

Σαχλάς Π. Αθανάσιος

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Υποβλήθηκε στο

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

του Πανεπιστημίου Πειραιώς

Πειραιάς

Μάρτιος 2010





University of Piraeus

Department of Statistics and Insurance Science

Subjects on Statistical Information Theory  
and Actuarial Science

Sachlas P. Athanasios

PhD Thesis

Submitted to

Department of Statistics and Insurance Science

of the University of Piraeus

Piraeus

March 2010



Η διατριβή αυτή αφιερώνεται στη μνήμη του

«Μεράκια»

(1912 - 1995)



## Ευχαριστίες

Ολοκληρώνοντας τη διδακτορική μου διατριβή θα ήθελα να ευχαριστήσω τον επιβλέποντά μου, ομότιμο καθηγητή κ. Τάκη Παπαϊωάννου για τη συνεχή καθοδήγησή του και την εμπιστοσύνη που μου επέδειξε. Οι γνώσεις αλλά και η εμπειρία του αποτέλεσαν σημαντικό εφόδιο για εμένα. Επιπλέον θα πρέπει να αναφέρω και το ενδιαφέρον που έδειξε για τη συνεργασία μας ακόμη και όταν βρισκόταν εκτός Ελλάδος. Επίσης θα ήθελα να ευχαριστήσω τα άλλα δυο μέλη της τριμελούς συμβουλευτικής επιτροπής, τον καθηγητή, πρόεδρο του τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, κ. Κλέωνα Τσίμπο και τον επίκουρο καθηγητή κ. Κωνσταντίνο Πολίτη, η συμβολή των οποίων υπήρξε καταλυτική για την ολοκλήρωση της διατριβής. Ιδιαίτερες ευχαριστίες πρέπει να απευθύνω επίσης στον καθηγητή, πρόεδρο του τμήματος Μαθηματικών του Πανεπιστημίου Ιωαννίνων, κ. Κωνσταντίνο Ζωγράφο, στον καθηγητή κ. Μάρκο Κούτρα, στην αναπληρώτρια καθηγήτρια κ. Μαρία Κατέρη και στον επίκουρο καθηγητή κ. Γεώργιο Πιτσέλη για τη συμμετοχή τους στην επταμελή επιτροπή. Οι εύστοχες παρατηρήσεις και διορθώσεις τους συνέβαλλαν στη βελτίωση του αρχικού χειμένου.

Θα πρέπει επίσης να ευχαριστήσω τους καθηγητές κ.κ. Κωνσταντινίδη Δημήτριο από το Τμήμα Στατιστικής, Αναλογιστικών και Χρηματοοικονομικών Σπουδών του Πανεπιστημίου Αιγαίου και Καρααγγελόπουλο Αλέξανδρο από το Τμήμα Μαθηματικών του Πανεπιστημίου Κύπρου, οι οποίοι με τα συνέδρια και τις ημερίδες που διοργάνωσαν μου έδωσαν την ευκαιρία πέραν του να παρουσιάσω την έρευνά μου, να γνωρίσω και να συζητήσω με τους καθηγητές κ.κ. Patrick Brockett και Adnan Awad.

Στη συνέχεια θα ήθελα να απευθύνω ευχαριστίες προς το τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς, το οποίο από το 1997 με «ανέχεται» στις τάξεις του ως προπτυχιακό φοιτητή, μεταπτυχιακό φοιτητή και υποψήφιο διδάκτορα, παρέχοντάς μου γνώσεις και υποστήριξη για τη συνέχιση των σπουδών μου.

Τέλος, ιδιαίτερες ευχαριστίες πρέπει να απευθύνω στην οικογένειά μου, τους γονείς και την αδερφή μου, η συμπαράσταση των οποίων σε διάφορα επίπεδα με βοήθησε να πραγματοποιήσω ένα από τα όνειρά μου.





# ΠΕΡΙΕΧΟΜΕΝΑ

<b>1</b>	<b>Εισαγωγή</b>	<b>1</b>
1.1	Εισαγωγή . . . . .	1
1.2	Δομή διατριβής . . . . .	2
<b>2</b>	<b>Στατιστική Θεωρία Πληροφοριών</b>	<b>5</b>
2.1	Εισαγωγή . . . . .	5
2.1.1	Σύντομη ιστορική αναδρομή της Θεωρίας Πληροφοριών . . . . .	6
2.1.2	Μέτρα Πληροφορίας . . . . .	9
2.1.3	Ιδιότητες Μέτρων Πληροφορίας . . . . .	23
2.2	Θεωρία Πληροφοριών και Στατιστική . . . . .	25
2.3	Θεωρία Πληροφοριών και Βιοστατιστική . . . . .	28
2.4	Θεωρία Πληροφοριών και Αναλογιστική Επιστήμη . . . . .	30
2.5	Πρόσφατη έρευνα στο πεδίο της Στατιστικής Θεωρίας Πληροφοριών . . . . .	33
<b>3</b>	<b>Μέτρα απόκλισης με μη πιθανοτικά μέτρα</b>	<b>35</b>
3.1	Εισαγωγή . . . . .	35
3.2	Το μέτρο απόκλισης των Kullback - Leibler με μη πιθανοτικά διανύσματα . . . . .	36
3.3	Αποκλίσεις δύναμης με μη πιθανοτικά διανύσματα . . . . .	44
3.4	Η διαφορά του Jensen με μη πιθανοτικά διανύσματα . . . . .	52
3.5	Συμπεράσματα Κεφαλαίου . . . . .	60
<b>4</b>	<b>Εξομάλυνση πινάκων θνησιμότητας μέσω αποκλίσεων</b>	<b>63</b>
4.1	Εισαγωγή . . . . .	63
4.2	Τι είναι εξομάλυνση . . . . .	63
4.3	Χαρακτηριστικά της εξομάλυνσης . . . . .	67
4.3.1	Ομαλότητα . . . . .	67
4.3.2	Καλή προσαρμογή . . . . .	68
4.4	Στατιστικά τεστ για την εξομάλυνση . . . . .	68
4.4.1	$X^2$ τεστ καλής προσαρμογής . . . . .	69
4.4.2	Έλεγχος μεμονωμένων τυποποιημένων αποκλίσεων . . . . .	70

4.4.3	Έλεγχος απόλυτων τυποποιημένων αποκλίσεων . . . . .	70
4.4.4	Έλεγχος αθροιστικών (συσσωρευμένων) αποκλίσεων . . . . .	71
4.4.5	Προσημικός έλεγχος . . . . .	71
4.5	Κατηγορίες μεθόδων εξομάλυνσης . . . . .	72
4.6	Εξομάλυνση μέσω μέτρων απόκλισης . . . . .	74
4.6.1	Εξομάλυνση μέσω του μέτρου απόκλισης των Kullback - Leibler . . . . .	76
4.6.2	Εξομάλυνση μέσω των αποκλίσεων δύναμης των Cressie - Read . . . . .	77
4.6.3	Εξομάλυνση μέσω του μέτρου απόκλισης του Jensen . . . . .	77
4.7	Ανάλυση των περιορισμών . . . . .	78
4.8	Συμπεράσματα Κεφαλαίου . . . . .	80
<b>5</b>	<b>Ελαχιστοποίηση μέτρων απόκλισης και Δυσικότητα</b>	<b>81</b>
5.1	Εισαγωγή . . . . .	81
5.2	Ελαχιστοποίηση αποκλίσεων . . . . .	82
5.3	Ελαχιστοποίηση αποκλίσεων στην Αναλογιστική Επιστήμη . . . . .	84
5.4	Το κανονικό MDI πρόβλημα . . . . .	85
5.5	Αρχή της Μέγιστης Εντροπίας . . . . .	86
5.6	Στοιχεία Μαθηματικού Προγραμματισμού . . . . .	88
5.6.1	Βέλτιστα και τοπικά βέλτιστα σημεία . . . . .	89
5.6.2	Πρόγραμμα Κυρτού Προγραμματισμού . . . . .	90
5.6.3	Προβλήματα τετραγωνικής βελτιστοποίησης . . . . .	91
5.6.4	Λαγκρανζιανή συνάρτηση και λαγκρανζιανή δυσικότητα . . . . .	92
5.7	Ελαχιστοποίηση του Διακριτού Μέτρου των Kullback - Leibler με γραμμικούς και τετραγωνικούς περιορισμούς . . . . .	95
5.8	Δυσική θεωρία για το πρόβλημα ελαχιστοποίησης της απόκλισης των Cressie - Read υπό τετραγωνικούς περιορισμούς . . . . .	99
5.9	Αποτελέσματα Λαγκρανζιανής δυσικότητας για τη διαφορά του Jensen . . . . .	102
5.10	Συμπεράσματα Κεφαλαίου . . . . .	107
<b>6</b>	<b>Εφαρμογή στην εξομάλυνση ποσοστών θνησιμότητας</b>	<b>109</b>
6.1	Εισαγωγή . . . . .	109
6.2	Ελαχιστοποίηση απόκλισης C - R υπό 5 περιορισμούς . . . . .	110
6.3	Ελαχιστοποίηση απόκλισης C - R υπό 6 περιορισμούς . . . . .	119
6.4	Ελαχιστοποίηση απόκλισης C - R υπό 4 περιορισμούς . . . . .	127
6.5	Ελαχιστοποίηση του εκτεταμένου μέτρου απόκλισης των Cressie and Read . . . . .	132
6.6	Ελαχιστοποίηση της διαφοράς του Jensen . . . . .	132
6.7	Χρήση άλλων μέτρων απόκλισης στην εξομάλυνση . . . . .	142
6.8	Συμπεράσματα Κεφαλαίου . . . . .	142

<b>7</b>	<b>Υπολογισμός εντροπίας για κατανομές απώλειας</b>	<b>145</b>
7.1	Εισαγωγή . . . . .	145
7.2	Επίδραση του πληθωρισμού . . . . .	148
7.3	Επίδραση της αποκοπής από κάτω . . . . .	149
7.4	Επίδραση της αποκοπής από επάνω . . . . .	154
7.5	Συμπεράσματα . . . . .	156
<b>8</b>	<b>Συμπεράσματα Διατριβής</b>	<b>159</b>
8.1	Συμπεράσματα . . . . .	159
8.2	Μελλοντική Έρευνα . . . . .	162
	<b>Βιβλιογραφία</b>	<b>165</b>
<b>A</b>	<b>Αποτελέσματα Δυικής Θεωρίας για την Ελαχιστοποίηση της Απόκλισης των Cressie and Read</b>	<b>177</b>
<b>B</b>	<b>Υπολογισμός της εντροπίας</b>	<b>181</b>
B.1	Υπολογισμός της εντροπίας με αποκοπή από κάτω . . . . .	181
B.1.1	Εκθετική Κατανομή . . . . .	181
B.1.2	Κατανομή Pareto . . . . .	182
B.1.3	Κατανομή Weibull . . . . .	183
B.1.4	Κατανομή Gamma . . . . .	184
B.1.5	Transformed Gamma Κατανομή . . . . .	185
B.1.6	Loggamma Κατανομή . . . . .	187
B.1.7	Lognormal Κατανομή . . . . .	189
B.1.8	Κατανομή Burr . . . . .	190
B.1.9	Γενικευμένη Pareto Κατανομή . . . . .	190
<b>Γ</b>	<b>Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών</b>	<b>193</b>
Γ.1	Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με το μέτρο των Cressie and Read και 5 περιορισμούς . . . . .	193
Γ.1.1	Σύνολο Δεδομένων L85 . . . . .	193
Γ.1.2	Σύνολο Δεδομένων HK01M . . . . .	194
Γ.1.3	Σύνολο Δεδομένων HK01F . . . . .	196
Γ.2	Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με το μέτρο των Cressie and Read και 6 περιορισμούς . . . . .	197
Γ.2.1	Σύνολο Δεδομένων L85 . . . . .	197
Γ.2.2	Σύνολο Δεδομένων HK01M . . . . .	198
Γ.2.3	Σύνολο Δεδομένων HK01F . . . . .	200

Γ.3	Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με το μέτρο των Cressie and Read και 4 περιορισμούς . . . . .	201
Γ.3.1	Σύνολο Δεδομένων L85 . . . . .	201
Γ.3.2	Σύνολο Δεδομένων HK01M . . . . .	202
Γ.3.3	Σύνολο Δεδομένων HK01F . . . . .	203
Γ.4	Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με το εκτεταμένο μέτρο των Cressie and Read . . . . .	205
Γ.4.1	Σύνολο Δεδομένων L85 . . . . .	205
Γ.4.2	Σύνολο Δεδομένων HK01M . . . . .	206
Γ.4.3	Σύνολο Δεδομένων HK01F . . . . .	207
Γ.5	Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με τη διαφορά του Jensen και 5 περιορισμούς . . . . .	209
Γ.5.1	Σύνολο Δεδομένων L85 . . . . .	209
Γ.5.2	Σύνολο Δεδομένων HK01M . . . . .	210
Γ.5.3	Σύνολο Δεδομένων HK01F . . . . .	211
Γ.6	Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με τη διαφορά του Jensen και 6 περιορισμούς . . . . .	213
Γ.6.1	Σύνολο Δεδομένων L85 . . . . .	213
Γ.6.2	Σύνολο Δεδομένων HK01M . . . . .	214
Γ.6.3	Σύνολο Δεδομένων HK01F . . . . .	215
Γ.7	Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με τη διαφορά του Jensen και 4 περιορισμούς . . . . .	217
Γ.7.1	Σύνολο Δεδομένων L85 . . . . .	217
Γ.7.2	Σύνολο Δεδομένων HK01M . . . . .	218
Γ.7.3	Σύνολο Δεδομένων HK01F . . . . .	219

# ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

2.1	Πυραμίδα Shannon - Kullback - Lindley - Jaynes . . . . .	10
2.2	Τρίγωνο Fisher - Shannon - Kullback . . . . .	11
6.1	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (L85 - 5 περιορισμοί) . . . . .	112
6.2	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (HK01M - 5 περιορισμοί) . . . . .	113
6.3	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (HK01F - 5 περιορισμοί) . . . . .	114
6.4	Ομαλότητα $S$ έναντι $\lambda$ (L85) . . . . .	116
6.5	Ομαλότητα $S$ έναντι $\lambda$ (HK01M) . . . . .	116
6.6	Ομαλότητα $S$ έναντι $\lambda$ (HK01F) . . . . .	117
6.7	Καλή προσαρμογή $F$ έναντι $\lambda$ (L85) . . . . .	118
6.8	Καλή προσαρμογή $F$ έναντι $\lambda$ (HK01M) . . . . .	118
6.9	Καλή προσαρμογή $F$ έναντι $\lambda$ (HK01F) . . . . .	119
6.10	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (L85 - 6 περιορισμοί) . . . . .	121
6.11	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (HK01M - 6 περιορισμοί) . . . . .	122
6.12	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (HK01F - 6 περιορισμοί) . . . . .	123
6.13	Καλή προσαρμογή $F$ έναντι $\lambda$ (L85- 6 περιορισμοί) . . . . .	123
6.14	Καλή προσαρμογή $F$ έναντι $\lambda$ (HK01M- 6 περιορισμοί) . . . . .	124
6.15	Καλή προσαρμογή $F$ έναντι $\lambda$ (HK01F- 6 περιορισμοί) . . . . .	125
6.16	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (L85 - 5 και 6 περιορισμοί και $\lambda = 2/3$ )	126
6.17	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (HK01M - 5 και 6 περιορισμοί και $\lambda = 2/3$ )	126
6.18	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (HK01F - 5 και 6 περιορισμοί και $\lambda = 2/3$ )	127
6.19	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (L85 - 4 περιορισμοί) . . . . .	128
6.20	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (HK01M - 4 περιορισμοί) . . . . .	129
6.21	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές (HK01F - 4 περιορισμοί) . . . . .	130
6.22	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές με το εκτεταμένο μέτρο CR (L85 - 5 περιορισμοί) . . . . .	134
6.23	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές με το εκτεταμένο μέτρο CR (HK01M - 5 περιορισμοί) . . . . .	134
6.24	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές με το εκτεταμένο μέτρο CR (HK01F - 5 περιορισμοί) . . . . .	137

6.25	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές μέσω Jensen difference (L85) . . . . .	138
6.26	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές μέσω Jensen difference (HK01M) . . .	139
6.27	Αδρές $u_x$ και εξομαλυμένες $v_x$ τιμές μέσω Jensen difference (HK01F) . . . .	140

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΠΑ

# ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

2.1	Ειδικές επιλογές της συνάρτησης $\phi$ . . . . .	16
2.2	$\phi$ - εντροπίες και $(h, \phi)$ - εντροπίες . . . . .	21
6.1	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων L85 (5 περιορισμοί) . . . . .	111
6.2	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01M (5 περιορισμοί) . . . . .	112
6.3	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01F (5 περιορισμοί) . . . . .	113
6.4	Τιμές των μέτρων ομαλότητας και προσαρμογής (5 περιορισμοί) . . . . .	115
6.5	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων L85 (6 περιορισμοί) . . . . .	120
6.6	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01M (6 περιορισμοί) . . . . .	121
6.7	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01F (6 περιορισμοί) . . . . .	122
6.8	Τιμές των μέτρων ομαλότητας και προσαρμογής (6 περιορισμοί) . . . . .	124
6.9	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων L85 (4 περιορισμοί) . . . . .	128
6.10	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01M (4 περιορισμοί) . . . . .	129
6.11	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01F (4 περιορισμοί) . . . . .	130
6.12	Τιμές των μέτρων ομαλότητας και προσαρμογής (4 περιορισμοί) . . . . .	131
6.13	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων L85 με το εκτεταμένο μέτρο CR και 5 περιορισμούς . . . . .	133
6.14	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01M με το εκτεταμένο μέτρο CR και 5 περιορισμούς . . . . .	133
6.15	Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01F με το εκτεταμένο μέτρο CR και 5 περιορισμούς . . . . .	135
6.16	Τιμές των μέτρων ομαλότητας και προσαρμογής για την εξομάλυνση με (i) το εκτεταμένο CR μέτρο απόκλισης και 5 περιορισμούς και (ii) τη μέθοδο W-H (χωρίς περιορισμούς) για τα σύνολα δεδομένων L85, HK01M, HK01F . . . . .	136
6.17	Εξομάλυνση μέσω Jensen difference για το σύνολο δεδομένων L85 . . . . .	138
6.18	Εξομάλυνση μέσω Jensen difference για το σύνολο δεδομένων HK01M . . . . .	139
6.19	Εξομάλυνση μέσω Jensen difference για το σύνολο δεδομένων HK01F . . . . .	140
6.20	Τιμές των μέτρων ομαλότητας και προσαρμογής (Jensen difference) . . . . .	141
7.1	Εκφράσεις εντροπίας . . . . .	147
7.2	Εντροπία υπό αποκοπή από κάτω . . . . .	152

7.3	Άνω φράγμα της εντροπίας υπό αποκοπή από κάτω . . . . .	153
7.4	Εντροπία υπό αποκοπή από επάνω . . . . .	157

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΑΙΑ



# ΠΕΡΙΛΗΨΗ

Στόχος της διατριβής αυτής είναι να συνδυάσουμε τη Στατιστική Θεωρία Πληροφοριών με επιστήμες όπως ο Αναλογισμός, η Βιοστατιστική και η Δημογραφία.

Συχνά στην πράξη χρειάζεται να αναθεωρούμε τις αρχικές εκτιμήσεις ποσοτήτων όπως τα ποσοστά θνησιμότητας, νοσηρότητας κλπ. με σκοπό την παραγωγή ομαλότερων εκτιμήσεων. Αυτό γίνεται μέσω μιας διαδικασίας που λέγεται εξομάλυνση (graduation ή smoothing). Στη διατριβή αυτή, διερευνούμε τη χρήση της οικογένειας των αποκλίσεων δύναμης (power divergence) τάξης  $\lambda$  που εισήγαγαν οι Read and Cressie - με κατάλληλους γραμμικούς και/ή τετραγωνικούς περιορισμούς με στόχο να βρούμε το καλύτερο μέτρο απόκλισης ώστε να πάρουμε την καλύτερη εξομάλυνση. Τα αποτελέσματα δείχνουν ότι μέτρα απόκλισης με μη πιθανοτικά διανύσματα (non-probability vectors), όπως συμβαίνει με τα ποσοστά θνησιμότητας, ικανοποιούν, υπό ορισμένες συνθήκες, μερικές από τις ιδιότητες των μέτρων απόκλισης ή πληροφορίας όπως αυτές ορίζονται στη Στατιστική Θεωρία Πληροφοριών. Οι αποκλίσεις δύναμης δίνουν αποτελέσματα ισοδύναμα με αυτά άλλων συχνά χρησιμοποιούμενων μεθόδων εξομάλυνσης. Παρουσιάζεται επίσης μια αριθμητική διερεύνηση για τη σύγκριση των μεθόδων και την εύρεση της δύναμης  $\lambda$  του μέτρου απόκλισης, που δίνει την καλύτερη εξομάλυνση. Φαίνεται ότι η επιλογή  $\lambda > 0$  και πιο συγκεκριμένα η  $\lambda = 2/3$ , που προτάθηκε από τους Cressie and Read για λόγους στατιστικής ισχύος, είναι μια καλή επιλογή όσον αφορά την ομαλότητα και την προσαρμογή. Επιπλέον, μελετάται, στο ίδιο πλαίσιο, η διαφορά του Jensen που πρότειναν οι Burbea and Rao.

Η εντροπία του Shannon (Shannon, 1948), η οποία ορίζεται ως  $H(X) = -\sum_i p_i \ln p_i$  ή  $H(X) = -\int f(x) \ln f(x) dx$ , στην περίπτωση διακριτών και συνεχών κατανομών, αντίστοιχα, αποτελεί μια χρήσιμη έννοια στη Στατιστική Επιστήμη, καθώς είναι ένα μέτρο της αβεβαιότητας που σχετίζεται με τη τυχαία μεταβλητή  $X$ . Αποτελεί μια περιγραφική ποσότητα των κατανομών και ανήκει στην κλάση των μέτρων διασποράς, όπως η διακύμανση και η τυπική απόκλιση. Ένα κεφάλαιο της διατριβής αυτής, αφιερώνεται στη διερεύνηση της επίδρασης και των ιδιοτήτων που έχουν στην εντροπία του Shannon, για τις πιο γνωστές κατανομές απώλειας, ο πληθωρισμός (inflation), η αποκοπή από επάνω (truncation from above) και η αποκοπή από κάτω (truncation from below), περιπτώσεις οι οποίες συχνά εμφανίζονται στην Αναλογιστική Επιστήμη. Επιπλέον, παρουσιάζουμε αναλυτικές εκφράσεις για την εντροπία των κατανομών αυτών.

# РАНЕЕЗНАМО ПЕРПАА

# ABSTRACT

The aim of this dissertation is to combine Statistical Information Science with Actuarial Science, Biostatistics and Demography.

Frequently we have to revise the initial estimates of death probabilities with the aim of obtaining smoother estimates through a procedure called graduation or smoothing. This thesis explores the use of measures of power divergence of order  $\lambda$  introduced by Cressie and Read, with proper linear and/or quadratic constraints with the aim of finding the divergence which results in the "best" fit. The results so far indicate that divergences with non-probability vectors in their arguments, as is the case with mortality rates, share, under some conditions, some of the properties of probabilistic or information theoretic divergences. The power divergences also give results equivalent to those obtained by other frequently used methods of graduation. A numerical investigation is also presented. It seems that the choice of  $\lambda > 0$  and particularly the value of  $\lambda = 2/3$  proposed by Cressie and Read in the light of statistical power, is a good choice as far as smoothness and goodness of fit are concerned. The Jensen difference proposed by Burbea and Rao is studied as well, under the same framework.

Shannon's entropy (Shannon, 1948), which is defined as  $H(X) = -\sum_i p_i \ln p_i$  or  $H(X) = -\int f(x) \ln f(x) dx$ , in the case of discrete or continuous distributions, respectively, constitutes a useful notion in Statistics, as it is a measure of the uncertainty related to a random variable  $X$ . It is a descriptive measure of distributions, belonging to the class of dispersion measures, such as the variance and the standard deviation. A chapter of this dissertation is devoted to the investigation of the effect and the properties of inflation, truncation from above and truncation from below on Shannon's entropy, for the most popular loss distributions. These situations frequently appear in Actuarial Science. Moreover, analytic expressions for the entropy in the above cases are also presented.

# РАНЕЕЗНАМО ПЕРПАА

# ΚΕΦΑΛΑΙΟ 1

## ΕΙΣΑΓΩΓΗ

### 1.1 Εισαγωγή

Ο κύριος σκοπός της παρούσας διδακτορικής διατριβής είναι η σύνθεση - σύνδεση της Στατιστικής Θεωρίας Πληροφοριών με άλλες επιστήμες. Δεν είναι λίγες οι περιπτώσεις που έννοιες και τεχνικές της Στατιστικής Θεωρίας Πληροφοριών βρίσκουν εφαρμογή παρέχοντας λύση σε διάφορα προβλήματα της Αναλογιστικής Επιστήμης, της Βιοστατιστικής και της Δημογραφίας. Οι εφαρμογές αυτές, μπορούν να ταξινομηθούν σε τρεις κατηγορίες ανάλογα με το μέτρο πληροφορίας που χρησιμοποιείται: την εντροπία, τα μέτρα απόκλισης και το μέτρο πληροφορίας του Fisher. Θα προσπαθήσουμε να συγκεντρώσουμε και να παρουσιάσουμε αυτές τις περιπτώσεις. Μια όχι τόσο πολύ γνωστή περίπτωση που χρησιμοποιείται η Στατιστική Θεωρία Πληροφοριών είναι η εξομάλυνση αναλογιστικών ποσοτήτων όπως είναι τα ποσοστά θνησιμότητας, η ένταση θνησιμότητας κ.α. Η περίπτωση αυτή θα μας απασχολήσει στο μεγαλύτερο μέρος αυτής της διατριβής. Συγκεκριμένα, θα παρουσιάσουμε την ελαχιστοποίηση μέτρων απόκλισης διαφορετικών από το μέτρο απόκλισης των Kullback-Leibler, το οποίο πρότεινε ο Brockett (1991).

Η μελέτη αυτή μας δίνει την ώθηση να μελετήσουμε τον ορισμό και τις ιδιότητες των μέτρων απόκλισης στην περίπτωση που έχουμε μη πιθανοτικά διανύσματα, δηλαδή όταν το άθροισμα των στοιχείων του διανύσματος δεν ισούται με 1, όπως συμβαίνει στην περίπτωση της αναλογιστικής εξομάλυνσης.

Η πληροφορία, με μια τεχνική έννοια του όρου, πρωτοεισήχθη στη Στατιστική από τον R.A. Fisher το 1925 σε εργασία σχετική με τη θεωρία της εκτίμησης. Βέβαια η έννοια της πληροφορίας είναι διάχυτη στη Στατιστική όπως τεκμηριώνεται από τη φράση: «τα δεδομένα παρέχουν ή περιέχουν πληροφορίες για ...». Η Θεωρία Πληροφοριών είναι ένας κλάδος της μαθηματικής θεωρίας των Πιθανοτήτων και της Μαθηματικής Στατιστικής. Ως τέτοια, χρησιμοποιείται σε ένα ευρύ φάσμα επιστημονικών πεδίων. Η Στατιστική Θεωρία Πληροφοριών σχετίζεται με τη στατιστική συμπερασματολογία και αποτελεί ένα από τα βασικά ενδια-

φέροντα των στατιστικών επιστημόνων. Η Στατιστική Θεωρία Πληροφοριών παρέχει μια ενοποίηση γνωστών αποτελεσμάτων και οδηγεί σε φυσικές γενικεύσεις και την παραγωγή νέων αποτελεσμάτων. Αρκετοί, εσφαλμένα, θεωρούν τη Στατιστική Θεωρία Πληροφοριών ως συνώνυμη με τη Θεωρία Επικοινωνιών (communication theory) αν και χρησιμοποιεί ορισμένες βασικές έννοιες αυτής.

## 1.2 Δομή διατριβής

Η δομή της παρούσας διδακτορικής διατριβής έχει ως εξής: Στο δεύτερο κεφάλαιο δίνουμε όσο το δυνατόν λεπτομερέστερα στοιχεία που αφορούν τη Στατιστική Θεωρία Πληροφοριών. Παρουσιάζουμε τα μέτρα πληροφορίας που έχουν προταθεί, τις εφαρμογές που αυτά βρίσκουν σε διάφορους επιστημονικούς τομείς καθώς και την πρόσφατη έρευνα στην περιοχή.

Στο Κεφάλαιο 3 διερευνούμε τις κυριότερες ιδιότητες των μέτρων απόκλισης (divergence measures) και πληροφορίας, στην περίπτωση που δεν έχουμε πιθανοτικά διανύσματα. Συγκεκριμένα, τα μέτρα που εξετάζουμε είναι η κατευθυνόμενη απόκλιση (directed divergence) των Kullback - Leibler, η απόκλιση των Cressie and Read, που πρότειναν οι Cressie and Read (1984) και η διαφορά του Jensen, που πρότειναν οι Burbea and Rao (1984).

Στο τέταρτο κεφάλαιο παρουσιάζουμε τα κυριότερα σημεία της εξομάλυνσης (graduation). Η εξομάλυνση είναι μια τεχνική που χρησιμοποιείται συχνά στην Αναλογιστική Επιστήμη, τη Βιοστατιστική και τη Δημογραφία, για την εκτίμηση ομαλών (smooth) εκτιμήσεων ποσοτήτων όπως ποσοστά θνησιμότητας, ποσοστά γονιμότητας, ποσοστά νοσηρότητας κλπ. Στη διατριβή αυτή θα παρουσιάσουμε την περίπτωση της εξομάλυνσης πινάκων θνησιμότητας. Θα παρουσιάσουμε τα βασικά χαρακτηριστικά της εξομάλυνσης, την ομαλότητα και την καλή προσαρμογή, τους στατιστικούς ελέγχους που χρησιμοποιούνται για τον έλεγχο της εξομάλυνσης και τις μεθόδους εξομάλυνσης.

Στο Κεφάλαιο 5 θα αναφερθούμε στον παραμετρικό και μη παραμετρικό τρόπο ελαχιστοποίησης των μέτρων απόκλισης. Επίσης θα αναφέρουμε την Ελάχιστη Διαχωριστική Πληροφορία (MDI) και τη μέθοδο της Μέγιστης Εντροπίας. Στη συνέχεια θα παρουσιάσουμε τα βασικότερα σημεία του Μαθηματικού Προγραμματισμού και κυριότερα του Κυρτού Μαθηματικού Προγραμματισμού και της Δυσκότητας κατά Lagrange (Λαγκρανζιανής Δυσκότητας). Αυτά χρειάζονται έτσι ώστε να παρουσιάσουμε διάφορα αποτελέσματα που αφορούν τη δυσκότητα για το πρόβλημα της ελαχιστοποίησης των μέτρων απόκλισης των Kullback - Leibler, των Cressie and Read, και της διαφοράς του Jensen, υπό τετραγωνικούς περιορισμούς.

Στο έκτο κεφάλαιο, παρουσιάζουμε μια αριθμητική διερεύνηση των μεθόδων εξομάλυνσης που στηρίζονται στη Στατιστική Θεωρία Πληροφοριών. Εξομαλύνουμε τρία σύνολα δεδομένων - πίνακες θνησιμότητας - ελαχιστοποιώντας τα μέτρα απόκλισης υπό διάφορους κατάλληλους περιορισμούς και συνδυασμούς περιορισμών και συγκρίνουμε τα αποτελέσματα

που προκύπτουν με αυτά που επιτυγχάνονται με άλλες γνωστές μεθόδους εξομάλυνσης.

Στο Κεφάλαιο 7, διερευνούμε τη χρήση της εντροπίας του Shannon ως μέτρο της αβεβαιότητας για κατανομές απώλειας (loss models or distributions) και γενικότερα στα αναλογιστικά μαθηματικά. Πιο συγκεκριμένα, μελετάμε την επίδραση που έχουν στον υπολογισμό της εντροπίας ο πληθωρισμός (inflation), η αποκοπή από επάνω (truncation from above) και η αποκοπή από κάτω (truncation from below). Οι περιπτώσεις αυτές εμφανίζονται συχνά στην πράξη. Η γενική, αλλά και η ειδική για κάθε κατανομή απώλειας, αναλυτική έκφραση δίνονται σε σχέση με την εντροπία των αρχικών δεδομένων. Αποδεικνύεται ότι ο πληθωρισμός αυξάνει την εντροπία, το οποίο σημαίνει ότι υπό πληθωρισμό έχουμε περισσότερη αβεβαιότητα σχετικά με την αρχική κατάσταση. Αντίθετα, η επίδραση της αποκοπής, είτε από επάνω είτε από κάτω, δεν είναι εύκολα ερμηνεύσιμη καθώς οι εκφράσεις της εντροπίας είναι αρκετά πολύπλοκες και διαφέρουν από μοντέλο σε μοντέλο. Σε μερικές περιπτώσεις δεν είναι δυνατόν να βρούμε μια αναλυτική έκφραση.

Στο όγδοο και τελευταίο κεφάλαιο παρουσιάζουμε τα κυριότερα αποτελέσματα που προέκυψαν από την έρευνά μας. Επιπλέον παρουσιάζουμε τις σχέψεις μας για μελλοντική έρευνα.

Στο τέλος της διατριβής υπάρχουν τρία παραρτήματα. Στο Παράρτημα Α παρουσιάζονται κάποια αποτελέσματα που αφορούν τη δυϊκότητα των μέτρων απόκλισης που μελετάμε. Στο Παράρτημα Β υπάρχουν αποτελέσματα σχετικά με τον υπολογισμό της εντροπίας των μοντέλων απώλειας ενώ στο Παράρτημα Γ παρουσιάζονται οι κώδικες που χρησιμοποιήθηκαν για να πάρουμε τα αριθμητικά αποτελέσματα που παρουσιάζονται στο Κεφάλαιο 6.

# РАНЕЕЗНАМО ПЕРПАА



## ΚΕΦΑΛΑΙΟ 2

# ΣΤΑΤΙΣΤΙΚΗ ΘΕΩΡΙΑ ΠΛΗΡΟΦΟΡΙΩΝ

### 2.1 Εισαγωγή

Σε κάθε στατιστικό πρόβλημα που εμπλέκεται δειγματοληψία, ο ερευνητής αναζητά την ποσότητα της πληροφορίας που δίνουν τα δεδομένα για την παράμετρο ή τις παραμέτρους του πληθυσμού (Kullback, 1959). Τεχνικά, πληροφορία σημαίνει το ποσό της αβεβαιότητας για μια άγνωστη ποσότητα, το οποίο μειώνεται από το αποτέλεσμα ενός πειράματος. Ο όρος αβεβαιότητα δεν σχετίζεται μόνο με την άγνωστη πυκνότητα  $f$  ή την παράμετρο  $\theta$  αλλά επίσης με το αποτέλεσμα ενός πειράματος τύχης (Papaioannou, 1985).

Πριν την εκτέλεση του πειράματος έχουμε μηδενική πληροφορία για την πυκνότητα  $f$  ή την παράμετρο  $\theta$ . Συλλέγοντας παρατηρήσεις από μια τυχαία μεταβλητή  $X$ , η οποία ακολουθεί κατανομή  $f$  ή  $f(X, \theta)$ , μειώνεται η αβεβαιότητα για την  $f$  ή την  $\theta$ . Δυο ομοιομόρφως κατανομημένες παρατηρήσεις παρέχουν την ίδια ποσότητα πληροφορίας. Στην εκτιμητική, στόχος μας είναι η εύρεση επαρκών στατιστικών συναρτήσεων, δηλαδή στατιστικών που περιέχουν όλη την πληροφορία για την  $f$  ή την  $\theta$  (Papaioannou, 1985).

Σύμφωνα με τον Soofi (1994), η πληροφορία ορίζεται ως μια αριθμητική ποσότητα που μετρά την αβεβαιότητα σχετικά με το αποτέλεσμα ενός πειράματος τύχης. Ο Fisher (1921) όρισε, στα πλαίσια της παραμετρικής εκτίμησης, τον αντίστροφο της διακύμανσης της δειγματικής κατανομής ενός εκτιμητή ως το μέτρο της σχετικής πληροφορίας που παρέχουν τα δεδομένα σχετικά με μια άγνωστη παράμετρο. Ιδιαίτερη έμφαση έχει δοθεί στη Θεωρία Πληροφοριών από την πλευρά της παροχής εναλλακτικών προσεγγίσεων / διατυπώσεων σε διάφορες παραδοσιακές στατιστικές αναλύσεις. Αυτό συνετέλεσε στο γεγονός, η Θεωρία Πληροφοριών να παρέχει ένα πλαίσιο για ενασχόληση με ένα ευρύ φάσμα στατιστικών προβλημάτων (Soofi, 1994).

Πολλοί θεωρούν ότι η Θεωρία Πληροφοριών είναι κλάδος της Θεωρίας Επικοινωνιών. Όμως είναι πολύ περισσότερα. Ενδεικτικά αναφέρουμε τις παρακάτω περιοχές εφαρμογής της Θεωρίας Πληροφοριών: Ηλεκτρική Μηχανική (Θεωρία Επικοινωνιών), Πληροφορική

(Kolmogorov Complexity), Φυσική (Θερμοδυναμική), Μαθηματικά (Θεωρία Πιθανοτήτων και Στατιστική), Φιλοσοφία της Επιστήμης, Οικονομικά, (Cover and Thomas, 1991), Οικολογία, Βιολογία, Νομικά κ.α. (Papaioannou, 1970).

### 2.1.1 Σύντομη ιστορική αναδρομή της Θεωρίας Πληροφοριών

#### Ανάπτυξη μετάδοσης της πληροφορίας

Στην ενότητα αυτή, θα αναφερθούμε σύντομα στην ιστορία της Θεωρίας Πληροφοριών (Arndt, 2001). Όλα μπορούμε να πούμε ότι ξεκίνησαν από την ανάπτυξη της μετάδοσης της πληροφορίας (information transmission). Ο Samuel Morse το 1837 ανέπτυξε το αλφάβητο Morse για τη μετάδοση μηνυμάτων με έναν βέλτιστο τρόπο. Η αντιστοίχιση των χαρακτήρων του αλφαβήτου στα σύμβολα δεν βασίστηκε στη σχετική συχνότητα με την οποία τα γράμματα εμφανίζονται σε ένα δοθέν κείμενο αλλά στην καταμέτρηση των τύπων των γραμμάτων ενός στοιχειοθέτη σε ένα γραφείο.

Ο Thomas Edison το 1874 εισήγαγε την τετραπλή (quadruplex) κωδικοποίηση χρησιμοποιώντας τα σύμβολα +3, +1, -1 και -3 για να αυξήσει περαιτέρω το ρυθμό μετάδοσης της πληροφορίας. Η ποσότητα της πληροφορίας που μπορεί να αποσταλεί μέσω ενός καναλιού εξαρτάται από τους τρεις παρακάτω παράγοντες: (i) πόσο γρήγορα μπορούμε να στείλουμε δυο συνεχόμενους χαρακτήρες χωρίς να οδηγηθούμε στη συνένωσή τους, (ii) πόσους διαφορετικούς χαρακτήρες μπορούμε να στείλουμε και (iii) πόσο χρόνο έχουμε για τη μετάδοση.

Τέλος, το 1924, ο Nyquist σε ένα άρθρο του αναφέρει ότι «Όταν στέλνουμε σύμβολα με σταθερό ρυθμό, ο ρυθμός της μετάδοσης  $W$  (σύμβολα ανά δευτερόλεπτο) συνδέεται με τον αριθμό των πιθανών διαφορετικών συμβόλων που μπορούν να αποσταλούν, μέσω της σχέσης:

$$W = B \ln(m),$$

όπου  $B$  είναι μια σταθερά που εξαρτάται από τον αριθμό των διαφορετικών χαρακτήρων που μπορούμε να αποστείλουμε ανά δευτερόλεπτο και  $m$  ο αριθμός των διαφορετικών συμβόλων που έχουμε στη διάθεσή μας προς αποστολή.»

#### Ανάπτυξη συναρτήσεων πληροφορίας

Ο Hartley το 1928, όρισε μια μονάδα και ένα μέτρο πληροφορίας διατυπώνοντας την παρακάτω πρόβλεψη: «Η απάντηση σε μια ερώτηση που μπορεί να υποθέσει τις δυο τιμές ΝΑΙ και ΟΧΙ (χωρίς να λαμβάνει υπόψη την έννοια της ερώτησης) περιέχει μια μονάδα πληροφορίας.»

Η μονάδα πληροφορίας καλείται “bit” γιατί μπορούμε να θεωρήσουμε τις δυο απαντήσεις σε ένα δυαδικό σύστημα ( $ΝΑΙ = 1$  και  $ΟΧΙ = 2$ ). Ο χαρακτηρισμός ενός συνόλου  $E_N$ , το

οποίο αποτελείται από  $N = 2^n$  δυαδικά στοιχεία, όπου η εμφάνιση των στοιχείων μπορεί να απαντηθεί με ΝΑΙ ή ΟΧΙ, απαιτεί πληροφορία

$$I(E_N) = n = \log_2 N = \frac{\ln(N)}{\ln(2)} \text{bit.}$$

Ο παραπάνω ορισμός της πληροφορίας είναι κυρίως γνωστός ως τύπος του Hartley και ισχύει υπό την υπόθεση ότι όλα τα  $N$  στοιχεία του συνόλου  $E_N$  έχουν την ίδια πιθανότητα εμφάνισης, δηλαδή  $p_{N_i} = \frac{1}{N}$  για  $i = 1, \dots, N$ .

Μια άλλη προσέγγιση της έννοιας της πληροφορίας βασίζεται στη σχέση αβεβαιότητας του μετασχηματισμού Fourier και προσπαθεί να βρει σήματα με ένα παράθυρο ελάχιστου χρόνου - συχνότητας (minimum time - frequency window). Πρόκειται για το μετασχηματισμό Gabor που πρότεινε ο Dennis Gabor το 1946. Η προσέγγιση αυτή αναλύει κάθε σήμα σε παράθυρα χρόνου - συχνότητας ενώ ο αριθμός τέτοιων παραθύρων αποτελεί την πληροφορία που περιέχεται στο παρατηρούμενο σήμα. Βέβαια, η προσέγγιση αυτή δεν έχει καθιερωθεί ως συνάρτηση ή μέτρο πληροφορίας καθώς ενδιαφέρεται για ντετερμινιστικές δομές και δεν λαμβάνει υπόψη της την αβεβαιότητα.

Ο Claude E. Shannon το 1948 επέκτεινε την έννοια της πληροφορίας του Hartley, θεωρώντας σύνολα που δεν απαιτείται να έχουν ίσες πιθανότητες εμφάνισης: Έστω  $E_1, \dots, E_n$ ,  $n$  αμοιβαίως ξένα μεταξύ (mutually exclusive) τους σύνολα με  $E = E_1 + \dots + E_n$ . Ο αριθμός των στοιχείων που περιλαμβάνονται στο σύνολο  $E_k$  είναι  $N_k$ , το οποίο σημαίνει ότι το σύνολο  $E$  αποτελείται από  $N = \sum_{k=1}^n N_k$  στοιχεία. Συνεπώς το σύνολο  $E_k$  εμφανίζεται με σχετική συχνότητα

$$p_k = \frac{N_k}{N}.$$

Όλα τα στοιχεία του συνόλου  $E_k$  έχουν την ίδια πιθανότητα εμφάνισης. Άρα, εάν γνωρίζουμε ότι ένα συγκεκριμένο στοιχείο είναι μέλος του συνόλου  $E_k$ , τότε ο ακριβής χαρακτηρισμός αυτού του στοιχείου στο σύνολο  $E_k$  χρειάζεται σύμφωνα με τον Hartley την πληροφορία

$$I(N_k) = \log_2 N.$$

Ο χαρακτηρισμός ενός στοιχείου του συνόλου  $E$  ξεχωριστά χρειάζεται τη μέση πληροφορία  $I_{binary}$ , όταν ήδη γνωρίζουμε ότι το στοιχείο είναι μέλος του συνόλου  $E_k$ , όπου

$$I_{binary} = \log_2 N + \sum_{k=1}^n p_k \log_2 p_k.$$

Από την παραπάνω σχέση παίρνουμε ότι

$$\log_2 N = I_{binary} - \sum_{k=1}^n p_k \log_2 p_k.$$

Η πληροφορία που απαιτείται για να χαρακτηρίσουμε ένα στοιχείο του συνόλου  $E$  είναι η συγχώνευση δυο κομματιών πληροφορίας, δηλαδή

$$\log_2 N = I_{binary} + I_{Shannon},$$

όπου το  $I_{Shannon}$  ορίζει το σύνολο  $E_k$  που περιέχει το στοιχείο και το  $I_{binary}$  ορίζει το συγκεκριμένο στοιχείο του γνωστού συνόλου  $E_k$ .

Συνεπώς, η εντροπία του Shannon ή εντροπία πληροφορίας δίνεται από τη σχέση

$$I_{Shannon} = \sum_{k=1}^n p_k \log_2 \left( \frac{1}{p_k} \right) = - \sum_{k=1}^n p_k \log_2 p_k.$$

Η πληροφορία του Shannon είναι η αναμενόμενη τιμή του τύπου του Hartley, όπου οι πιθανότητες δεν είναι απαραίτητο να είναι ίσες για όλα τα πιθανά ενδεχόμενα. Στην περίπτωση όμως που οι πιθανότητες είναι ίσες τότε η πληροφορία του Shannon ισούται με την πληροφορία του Hartley.

Συνήθως στον τύπο της πληροφορίας του Shannon χρησιμοποιείται ο νεπέριος λογάριθμος και συμβολίζεται με  $H(X)$ , δηλαδή

$$H(X) = - \sum_{k=1}^n p_k \ln(p_k).$$

Η μέγιστη εντροπία επιτυγχάνεται όταν όλα τα  $n$  σύμβολα εμφανίζονται με την ίδια πιθανότητα και ισούται με

$$H = H(X) = - \sum_{k=1}^n \frac{1}{n} \ln \left( \frac{1}{n} \right) = - \ln \left( \frac{1}{n} \right) = \ln(n).$$

Η πληροφορία του Shannon περιγράφει ακριβώς την αβεβαιότητα που χάνεται από μια μέτρηση. Χωρίς καμιά παρατήρηση μπορούμε να δώσουμε μια ασαφή, αόριστη περιγραφή της πιθανής απεικόνισης αποδίδοντας πιθανότητες στα πιθανά ενδεχόμενα. Λαμβάνοντας μια παρατήρηση, η εκ των προτέρων πιθανότητα εξαλείφεται καθώς μπορούμε να προσδιορίσουμε ένα ενδεχόμενο ως αποτέλεσμα της μέτρησης. Συνεπώς, η πληροφορία του Shannon μετρά την πληροφορία που περιέχεται σε μια παρατήρηση ή την αβεβαιότητα που χάνεται.

Η αρχική ονομασία της ποσότητας αυτής, όπως ο John von Neumann συμβούλευσε τον Shannon, ήταν εντροπία. Το μέτρο του Shannon περιγράφει την αρχική αβεβαιότητα που έχουμε πριν κάνουμε την πρώτη παρατήρηση. Έτσι γνωρίζουμε μόνο την κατανομή πυκνότητας του πειράματος καθώς δεν έχουμε κάποια συμπαγή απεικόνιση. Όμως η κατανομή πυκνότητας μας επιτρέπει να υπολογίσουμε την αβεβαιότητα (εντροπία) που έχουμε όσον αφορά την επιθυμητή απεικόνιση. Συνεπώς το μέτρο του Shannon είναι ένα μέτρο της

εντροπίας, όταν κοιτάμε το πείραμα στη βάση της αρχικής μας γνώσης.

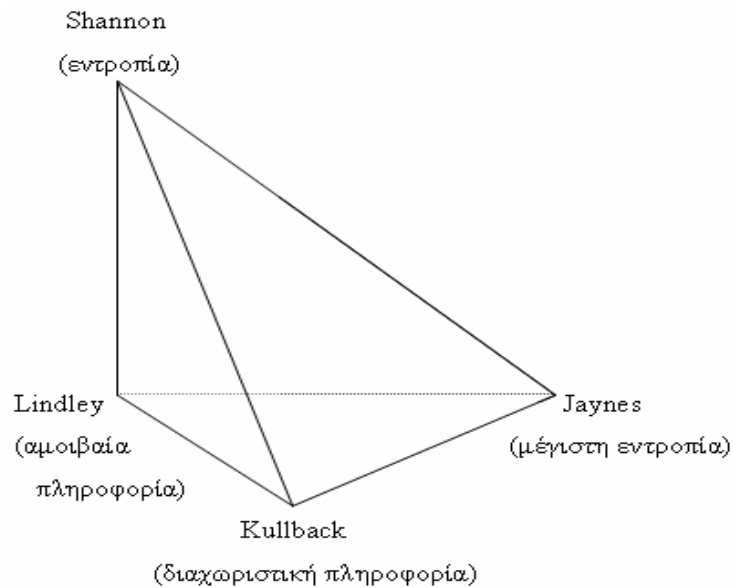
Το μέτρο του Shannon είναι επίσης ένα μέτρο πληροφορίας γιατί εκτός από την περιγραφή της αρχικής αβεβαιότητας, περιγράφει επιπλέον την πληροφορία που μπορούμε να πάρουμε από την επόμενη παρατήρηση. Όταν εξετάζουμε το μέτρο του Shannon υπό αυτό το πρίσμα, την τάση δηλαδή για μελλοντικό κέρδος πληροφορίας, μπορούμε να μιλάμε για πληροφορία του Shannon, καθώς περιγράφει την πληροφορία που περιέχεται στις ακόλουθες παρατηρήσεις.

Εδώ θα πρέπει να κάνουμε ένα διαχωρισμό μεταξύ διακριτών συναρτήσεων πιθανότητας και συνεχών συναρτήσεων πυκνότητας πιθανότητας, καθώς οι διακριτές πιθανότητες αποδίδονται σε αριθμήσιμα ενδεχόμενα, τα οποία μπορεί να εμφανιστούν ή όχι. Είναι λοιπόν πιθανόν να βρούμε έναν ακριβή προσδιορισμό της εμφάνισης ενός ενδεχομένου μέσω μιας απλής απόφασης της μορφής ΝΑΙ / ΟΧΙ εξαλείφοντας όλη την αβεβαιότητα που αφορά την απεικόνιση της παρατηρούμενης διακριτής τυχαίας μεταβλητής. Η περιγραφή ενός τυχαίου πειράματος με συνεχείς τυχαίες μεταβλητές δεν οδηγεί σε ακριβείς απεικονίσεις καθώς δεν μπορούμε να κάνουμε ακριβείς μετρήσεις. Στην περίπτωση αυτή έχουμε αβεβαιότητα ακόμα και μετά την παρατήρηση και το κέρδος σε πληροφορία που έχουμε από τη μέτρηση δεν εξαρτάται μόνο από την αρχική αβεβαιότητα αλλά και από την αβεβαιότητα που παραμένει μετά τη μέτρηση (η οποία στη διακριτή περίπτωση ισούται με μηδέν).

### 2.1.2 Μέτρα Πληροφορίας

Από τη στιγμή που υπάρχει πληροφορία, θα πρέπει με κάποιο τρόπο να την μετρήσουμε. Για το λόγο αυτό έχουν προταθεί και ακόμη προτείνονται μέτρα ή οικογένειες μέτρων πληροφορίας. Τα πιο γνωστά και θεμελιώδη μέτρα πληροφορίας, υπό την έννοια ότι σε αυτά στηρίζεται η κατασκευή νέων μέτρων, είναι του Fisher, των Kullback και Leibler και του Shannon. Τα μέτρα αυτά, αλλά και άλλα, αλληλοσυσχετίζονται, καθώς το ένα προέρχεται από το άλλο ή αποτελεί ειδική περίπτωση του. Οι στατιστικοί διαφωνούν στο γεωμετρικό σχήμα που αναπαριστά τα μέτρα πληροφορίας και τις αλληλοσυσχετίσεις τους. Ο Soofi (1994) συνοπτικά αναπαριστά τα μέτρα αυτά με την πυραμίδα Shannon - Kullback - Lindley - Jaynes (Σχήμα 2.1). Στην κορυφή της πυραμίδας τοποθετείται το μέτρο του Shannon, γνωστό ως εντροπία (entropy) ενώ στη βάση τοποθετούνται τρεις γενικεύσεις της εντροπίας. Το μέτρο του Kullback, η μέγιστη εντροπία του Jaynes και η αμοιβαία πληροφορία του Lindley. Η πλάγια πλευρά μεταξύ των κορυφών Shannon, Kullback και Jaynes είναι το επίπεδο της ελάχιστης διαχωριστικής πληροφορίας. Η πλάγια πλευρά μεταξύ των κορυφών Shannon, Lindley και Kullback είναι το επίπεδο της αμοιβαίας πληροφορίας, ενώ η τρίτη πλάγια πλευρά μεταξύ των κορυφών Shannon, Lindley και Jaynes είναι το επίπεδο της μεψυζιανής Θεωρίας Πληροφοριών. Τα υπόλοιπα μέτρα μπορούν να τοποθετηθούν, σύμφωνα με τον Soofi (1994), στο εσωτερικό της πυραμίδας.

Ο Papaioannou (2001) διαφωνεί με την παραπάνω αναπαράσταση, αφού ο Jaynes δεν



Σχήμα 2.1: Πυραμίδα Shannon - Kullback - Lindley - Jaynes

ανακάλυψε ένα καινούριο μέτρο πληροφορίας αλλά εισήγαγε την αρχή της μέγιστης εντροπίας (maximum entropy principle). Συνεπώς, ο Παραϊοαννου (2001) προτείνει ότι η Στατιστική Θεωρία Πληροφοριών πρέπει να αναπαριστάται από ένα τρίγωνο (Σχήμα 2.2), στην κορυφή του οποίου θα βρίσκεται το μέτρο του Fisher ενώ στη βάση τα μέτρα των Shannon και Kullback.

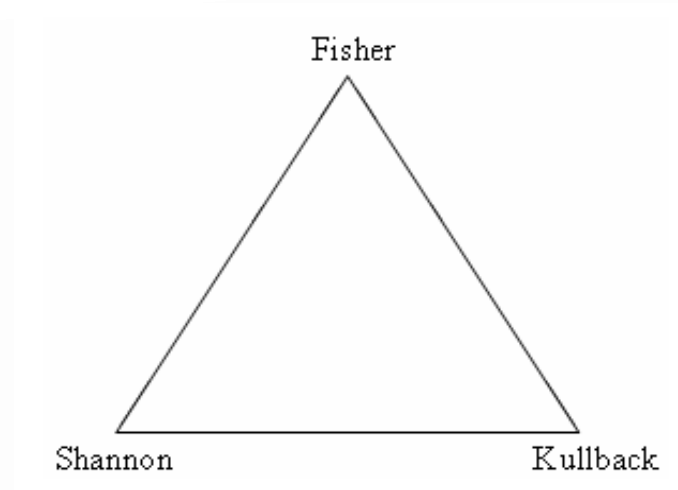
Τα μέτρα πληροφορίας μπορούν να χωριστούν σύμφωνα με τον Παραϊοαννου (1985 και 2001) σε τρεις κυρίως κατηγορίες: σε παραμετρικά ή τύπου Fisher, σε μη παραμετρικά ή τύπου απόστασης και σε τύπου εντροπίας μέτρα πληροφορίας. Τελευταία έχουν εμφανισθεί και μπεϋζιανά μέτρα πληροφορίας.

### Παραμετρικά μέτρα ή μέτρα τύπου Fisher

Τα παραμετρικά μέτρα, όπως εύκολα μπορεί κανείς να συμπεράνει από το όνομά τους, αναφέρονται σε παραμετρικές οικογένειες κατανομών  $\{f(X, \theta), \theta \in \Theta\}$  και μετρούν το ποσό της πληροφορίας που παρέχουν τα δεδομένα για την άγνωστη παράμετρο  $\theta$ . Τα μέτρα αυτά αποτελούν συναρτήσεις του  $\theta$ .

Ο κυριότερος εκπρόσωπος της κατηγορίας αυτής είναι το γνωστό μέτρο πληροφορίας του Fisher, το οποίο ορίζεται ως

$$I_X^F(\theta) = E \left( \frac{\partial \ln f(X, \theta)}{\partial \theta} \right)^2,$$



Σχήμα 2.2: Τρίγωνο Fisher - Shannon - Kullback

αν η  $\theta$  είναι μονοδιάστατη παράμετρος ή

$$I_X^F(\boldsymbol{\theta}) = \left( E \left\{ \left( \frac{\partial \ln f(X, \theta)}{\partial \theta_i} \right) \left( \frac{\partial \ln f(X, \theta)}{\partial \theta_j} \right) \right\} \right)_{p \times p},$$

αν η παράμετρος  $\boldsymbol{\theta}$  είναι  $p$  - διάστατο διάνυσμα. Ο πίνακας του Fisher είναι συμμετρικός και θετικά ημιορισμένος.

Ο Rao (1973) αναφέρει ότι το μέτρο πληροφορίας του Fisher μετρά το όριο μέχρι το οποίο η αβεβαιότητα μειώνεται από την παρατήρηση ενώ ο Lehmann (1983) την ευκολία με την οποία μια παράμετρος μπορεί να εκτιμηθεί.

Το μέτρο πληροφορίας του Fisher συνδέεται με το μέτρο πληροφορίας των Kullback - Leibler, το οποίο θα δούμε στην επόμενη παράγραφο, ως εξής:

Έστω  $f_1 = f_\theta$  και  $f_2 = f_{\theta+\Delta\theta}$ , όπου  $\theta$  και  $\theta + \Delta\theta$  είναι μονομεταβλητά γειτονικά σημεία στον παραμετρικό χώρο. Τότε ισχύει (Kullback, 1959, σελ. 28)

$$I_X^{KL}(f_\theta, f_{\theta+\Delta\theta}) = \frac{1}{2}(\Delta\theta)^2 I_X^F(\theta).$$

Υπό αυτή την έννοια, το μέτρο του Fisher μπορεί να θεωρηθεί ως μέτρο απόστασης μεταξύ γειτονικών σημείων στον παραμετρικό χώρο.

Οι Casella and Berger (2001) αναφέρουν ότι εάν ισχύει η σχέση

$$\frac{\partial}{\partial \theta} E \left( \frac{\partial}{\partial \theta} \ln f(X, \theta) \right) = \int \frac{\partial}{\partial \theta} \left( \frac{\partial}{\partial \theta} f(x, \theta) \ln f(x, \theta) \right) dx,$$

τότε το μέτρο πληροφορίας του Fisher παίρνει τη μορφή

$$I_X^F(\theta) = - \int f(x, \theta) \frac{\partial^2}{\partial \theta^2} \ln f(x, \theta) dx.$$

Στην περίπτωση αυτή, η πληροφορία μπορεί να θεωρηθεί ως μέτρο της αιχμηρότητας (sharpness) της καμπύλης στηρίγματος (support curve) κοντά στον εκτιμητή μέγιστης πιθανοφάνειας της παραμέτρου  $\theta$ .

Το μέτρο πληροφορίας του Fisher είναι ιδιαίτερης σημασίας στη Στατιστική κυρίως επειδή εμφανίζεται στην ανισότητα Cramer - Rao. Η ανισότητα αυτή αναφέρει ότι ο αντίστροφος της πληροφορίας του Fisher είναι ένα ασυμπτωτικό κάτω όριο της διακύμανσης οποιουδήποτε αμερόληπτου εκτιμητή της παραμέτρου  $\theta$ , δηλαδή:

Εάν  $E(T(X_1, X_2, \dots, X_n)) = \theta + b(\theta)$ , όπου  $T$  μια στατιστική συνάρτηση, και ισχύουν κάποιες συνθήκες ομαλότητας (βλέπε Rao (1953, σελ. 329) και Rohatgi (2001, σελ. 391 - 393)), τότε

$$\text{var}(T) \geq \frac{(1 + b'(\theta))^2}{nI_X^F(\theta)}.$$

Εάν επιπλέον ο  $T$  είναι αμερόληπτος εκτιμητής, τότε  $E(T(X_1, X_2, \dots, X_n)) = \theta$  και

$$\text{var}(T) \geq \frac{1}{nI_X^F(\theta)},$$

το οποίο είναι η μικρότερη δυνατή διακύμανση για όλους τους αμερόληπτους εκτιμητές της παραμέτρου  $\theta$  (βλέπε Παπαϊωάννου και Φερεντίνος, 2000).

Άλλα μέτρα πληροφορίας για την μονοδιάστατη παράμετρο  $\theta$  είναι τα μέτρα των Vajda, Mathai και Boeke, οι τύποι των οποίων είναι

$$I_X^V(\theta) = E \left| \frac{\partial \ln f(X, \theta)}{\partial \theta} \right|^a, a \geq 1$$

$$I_X^{Mat}(\theta) = \left( E \left| \frac{\partial \ln f(X, \theta)}{\partial \theta} \right|^a \right)^{\frac{1}{a}}, a \geq 1$$

και

$$I_X^{Bo}(\theta) = \left( E \left| \frac{\partial \ln f(X, \theta)}{\partial \theta} \right|^{\frac{s}{s-1}} \right)^{s-1}, 1 < s < \infty,$$

αντίστοιχα. Το μέτρο του Vajda ονομάζεται επίσης πληροφορία του Fisher τάξης  $a \geq 1$  ενώ είναι εύκολα αντιληπτό ότι τα μέτρα των Mathai και Boeke σχετίζονται με αυτό του Vajda μέσω των τύπων

$$I_X^{Mat}(\theta) = (I_X^V(\theta))^{\frac{1}{a}}$$

και

$$I_X^{Bo}(\theta) = (I_X^V(\theta))^{\frac{1}{a-1}},$$



αντίστοιχα.

Για  $p$  - διάστατο  $\theta$  το μόνο μέτρο πληροφορίας που υπάρχει είναι ο πίνακας πληροφορίας του Fisher. Οι Papaioannou and Kempthorne (1971) και Ferentinos and Papaioannou (1981) πρότειναν ως μονομεταβλητά μέτρα τα παρακάτω, τα οποία προέρχονται από τον πίνακα πληροφορίας του Fisher:

$$I_X^*(\theta) = \text{tr} (I_X^F(\theta)),$$

$$D_X(\theta) = \det (I_X^F(\theta)),$$

$$\lambda_X^*(\theta) = \lambda_i (I_X^F(\theta)),$$

όπου  $\lambda_i$  είναι η  $i$  - οστή μεγαλύτερη ιδιοτιμή του  $I_X^F(\theta)$ .

Τελευταία εμφανίστηκε σαν μέτρο πληροφορίας, η λεγόμενη παρατηρούμενη πληροφορία του Fisher ή αριθμός πληροφορίας του Fisher που δίνεται από τον τύπο

$$I_X^F(\theta) = E \left( \frac{\partial \ln f(x, \theta)}{\partial x} \right)^2.$$

Παρατηρούμε ότι το μέτρο αυτό είναι ίδιο με το μέτρο πληροφορίας του Fisher, με τη μόνη διαφορά πως παραγωγίζουμε ως προς  $x$  αντί ως προς  $\theta$ . Οι Papaioannou and Ferentinos (2002) προτείνουν να μην χρησιμοποιείται ο αριθμός πληροφορίας του Fisher ως μέτρο πληροφορίας αφού δεν ικανοποιεί δυο από τις βασικές ιδιότητες του μέτρου του Fisher: την μονοτονία και την αμεταβλητότητα (invariance). Τις ιδιότητες των μέτρων πληροφορίας θα παρουσιάσουμε σε επόμενη παράγραφο.

### Μη παραμετρικά μέτρα ή μέτρα τύπου απόκλισης

Το αντιπροσωπευτικό μέτρο της μη παραμετρικής κατηγορίας είναι το μέτρο των Kullback - Leibler (1951), το οποίο δίνεται από τον τύπο

$$I_X^{KL}(f_1, f_2) = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx,$$

και μετρά την απόσταση μεταξύ δυο κατανομών  $f_1$  και  $f_2$  (Kullback, 1959). Με άλλα λόγια δίνει την πληροφορία που κερδίζουμε αν αντικαταστήσουμε την κατανομή  $f_2$  με την  $f_1$ . Εναλλακτικά αναφέρεται ως σχετική ή διασταυρούμενη εντροπία (relative ή cross - entropy) και πληροφορία διαχωρισμού (discrimination information). Οι Alwan et al. (1998) παρουσιάζουν αναλυτικές εκφράσεις για τον υπολογισμό της τιμής του μέτρου των Kullback - Leibler για αρκετές μονομεταβλητές κατανομές, οι οποίες διαφέρουν ως προς τις παραμέτρους τους. Στο σημείο αυτό θα πρέπει να αναφέρουμε ότι το μέτρο αυτό δεν ικανοποιεί όλες τις ιδιότητες της απόστασης ή της μετρικής. Στην πραγματικότητα, δεν είναι συμμετρι-

κό μέτρο και δεν ικανοποιεί την ιδιότητα της τριγωνικής ανισότητας. Συνεπώς δεν μπορεί να θεωρηθεί ως μια καθαρή απόσταση και για το λόγο αυτό είναι μέτρο κατευθυνόμενης απόκλισης.

**Θεώρημα 2.1.** (Ανισότητα Πληροφορίας) Έστω  $f_1(x), f_2(x), x \in H$ , δυο συναρτήσεις πιθανότητας. Τότε  $I_X^{KL}(f_1, f_2) \geq 0$  με ισότητα αν και μόνο αν  $f_1 = f_2$  για κάθε  $x$ .

Απόδειξη. Έστω  $A = \{x : f_1(x) > 0\}$  το στήριγμα της  $f_1(x)$ . Τότε

$$\begin{aligned} -I_X^{KL}(f_1, f_2) &= - \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx \\ &= \int f_1(x) \ln \frac{f_2(x)}{f_1(x)} dx \leq \ln \int f_1(x) \frac{f_2(x)}{f_1(x)} dx \\ &= \ln \int f_2(x) dx \leq \ln \int f_2(x) dx \\ &= \ln 1 = 0, \end{aligned}$$

όπου χρησιμοποιήσαμε την ανισότητα του Jensen για τη συνάρτηση  $t \ln t$ . Επειδή η  $\ln t$  είναι αυστηρά κοίλη συνάρτηση του  $t$ , η  $t \ln t$  είναι αυστηρά κυρτή και συνεπώς η ισότητα επιτυγχάνεται αν και μόνο αν  $\frac{f_2(x)}{f_1(x)} = 1$  για όλα τα  $x$ . Δηλαδή αν  $f_1(x) = f_2(x)$  (Cover and Thomas, 1991, σελ. 26). ■

Το παραπάνω θεώρημα μας λέει ότι το μέτρο των Kullback - Leibler, δηλαδή η πληροφορία για την απόσταση μεταξύ δυο κατανομών είναι πάντοτε θετική. Το θεώρημα αυτό είναι αρκετά σημαντικό και αποτελεί τη βάση για τα περισσότερα αποτελέσματα που θα παρουσιαστούν στη συνέχεια της διατριβής.

Η ανισότητα του Jensen που χρησιμοποιείται στην παραπάνω απόδειξη, έχει ως εξής: Αν η  $f$  είναι κυρτή συνάρτηση και  $X$  μια τυχαία μεταβλητή, τότε

$$E[f(X)] \geq f[E(X)].$$

Επιπλέον, αν η  $f$  είναι αυστηρώς κυρτή, τότε η ισότητα υπονοεί ότι  $X = E(X)$  με πιθανότητα 1, δηλαδή η  $X$  είναι σταθερή (Cover and Thomas, 1991, σελ. 482).

Εάν έχουμε δυο τυχαίες μεταβλητές  $X$  και  $Y$  με από κοινού συνάρτηση πυκνότητας πιθανότητας  $p(x, y)$  και περιθώριες συναρτήσεις πυκνότητας πιθανότητας  $f(x)$  και  $g(y)$ , τότε η σχετική εντροπία μεταξύ της από κοινού κατανομής και του γινομένου των περιθώριων κατανομών, δηλαδή η

$$I(X, Y) = \int \int p(x, y) \ln \frac{p(x, y)}{f(x)g(y)} dx dy = I^{KL}(p(x, y), f(x)g(y))$$

ονομάζεται αμοιβαία πληροφορία (mutual information). Το παραπάνω μέτρο, ερμηνεύεται ως η πληροφορία που παρέχει η μια μεταβλητή για την πρόβλεψη της άλλης (Soofi, 1994).

Η αμοιβαία πληροφορία είναι μια τυπική αυστηρά ορισμένη έννοια της Θεωρίας Πληροφοριών που ορίζει ένα λογαριθμικό μέτρο εξάρτησης μεταξύ δυο ή περισσότερων τυχαίων μεταβλητών με αυθαίρετους μετρήσιμους δειγματικούς χώρους (Zografos, 2008b). Αρχικά χρησιμοποιήθηκε από τον Lindley (1956) για τη μπεύζιανή σύγκριση πειραμάτων. Η έννοια αυτή πρωτοεισήχθη από τον Shannon (1948) και στη συνέχεια γενικεύθηκε από διάφορους συγγραφείς μεταξύ των οποίων, οι Vajda (1989), Joe (1989) και Micheas and Zografos (2006).

Ο Jeffreys (1946) πρότεινε τη χρήση της συμμετρικής εκδοχής του μέτρου των Kullback - Leibler,

$$J_X(f_1, f_2) = I_X^{KL}(f_1, f_2) + I_X^{KL}(f_2, f_1),$$

ως μέτρο απόκλισης μεταξύ δυο κατανομών πιθανότητας. Το μέτρο αυτό ονομάζεται  $J$  - απόκλιση.

Άλλα μη παραμετρικά μέτρα είναι το μέτρο του Bhattacharyya (1943),

$$I_X^B(f_1, f_2) = -\ln \int \sqrt{f_1(x)}\sqrt{f_2(x)} dx,$$

καθώς και η γενίκευσή του, γνωστή ως μέτρο του Rényi (1961) που δίνεται από τον τύπο

$$I_X^R(f_1, f_2) = (a - 1)^{-1} \ln \int f_1^a(x) f_2^{1-a}(x) dx, \quad a > 0, \quad a \neq 1.$$

Το μέτρο αυτό είναι επίσης γνωστό ως πληροφορία τάξης  $a$  όταν η κατανομή  $f_2$  αντικατασταθεί από την  $f_1$ .

Αργότερα, οι Liese and Vajda (1987) επέκτειναν το μέτρο του Rényi για όλα τα  $a \neq 0, 1$ , ως

$$I_X^{RLV}(f_1, f_2) = [a(a - 1)]^{-1} \ln \int f_1^a(x) f_2^{1-a}(x) dx, \quad a \neq 0, 1.$$

Μερικές φορές στη βιβλιογραφία, το μέτρο των Liese and Vajda αναφέρεται ως μέτρο του Rényi. Για τις τιμές  $a = 0$  και  $a = 1$  το μέτρο ορίζεται ως όριο. Είναι εύκολο να δειχθεί ότι  $\lim_{a \rightarrow 0} I_X^R(f_1, f_2) = I_X^{KL}(f_2, f_1)$  και  $\lim_{a \rightarrow 1} I_X^R(f_1, f_2) = I_X^{KL}(f_1, f_2)$ .

Επίσης το μέτρο του Kagan (1963),

$$I_X^{Ka}(f_1, f_2) = \int \left[ 1 - \frac{f_1(x)}{f_2(x)} \right]^2 f_2(x) dx,$$

γνωστό ως  $\chi^2$  - απόκλιση, η γενίκευσή του

$$I_X^V(f_1, f_2) = \int \left| 1 - \frac{f_1(x)}{f_2(x)} \right|^a f_2(x) dx,$$

γνωστό ως μέτρο του Vajda ή  $\chi - a$  - απόκλιση και το μέτρο του Matusita (1951),

$$I_X^M(f_1, f_2) = \left[ \int \left( \sqrt{f_1(x)} - \sqrt{f_2(x)} \right)^2 dx \right]^{\frac{1}{2}} = [2(1 - \rho)]^{\frac{1}{2}},$$

όπου  $\rho = \int \sqrt{f_1(x)} \sqrt{f_2(x)} dx$  είναι η σχέση μεταξύ των κατανομών  $f_1(x)$  και  $f_2(x)$ .

Ένα σημαντικό μέτρο πληροφορίας είναι αυτό του Csiszar,

$$I_X^C(f_1, f_2) = I_X^\phi(f_1, f_2) = \int \phi \left( \frac{f_1(x)}{f_2(x)} \right) f_2(x) dx,$$

γνωστό και ως  $\phi$  - απόκλιση. Η  $\phi$  είναι πραγματική κυρτή συνάρτηση στο  $[0, \infty)$  με  $0\phi \left( \frac{0}{0} \right), a\phi \left( \frac{0}{0} \right) = a\phi_\infty$  και  $\phi_\infty = \lim_{u \rightarrow \infty} \left[ \frac{\phi(u)}{u} \right]$ . Το  $I_X^C(f_1, f_2)$  είναι μια ευρύτερη οικογένεια μέτρων πληροφορίας αφού για ειδικές επιλογές της συνάρτησης  $\phi$ , καταλήγουμε σε άλλα μη παραμετρικά μέτρα. Σημειώνουμε ότι το μέτρο αυτό προτάθηκε ανεξάρτητα από τον Csiszar (1963) και τους Ali and Silvey (1966). Παραδείγματα της συνάρτησης  $\phi$  για ορισμένα γνωστά μέτρα πληροφορίας, παρουσιάζονται στον Πίνακα 2.1.

$\phi(u)$	Μέτρο Πληροφορίας
$u \ln u$	$I_X^{KL}(f_1, f_2)$
$(1 - u)^2$	$I_X^{Ka}(f_1, f_2)$
$(1 - \sqrt{u})^2$	$I_X^M(f_1, f_2)$
$ 1 - u ^a, a \geq 1$	$I_X^V(f_1, f_2)$

Πίνακας 2.1: Ειδικές επιλογές της συνάρτησης  $\phi$

Δυο γνωστές παραμετρικές γενικεύσεις του μέτρου των Kullback - Leibler είναι το λεγόμενο  $r$  - τάξης και  $s$  - βαθμού μέτρο απόκλισης και το  $1$  - τάξης και  $s$  - βαθμού μέτρο απόκλισης που πρότειναν οι Sharma and Mittal (1977). Τα μέτρα αυτά δίνονται από τις σχέσεις

$$I_X^{r,s}(f_1, f_2) = \frac{1}{s-1} \left[ \left( \int f_1^r(x) f_2^{1-r}(x) dx \right)^{\frac{s-1}{r-1}} - 1 \right], \quad r, s \neq 1$$

και

$$I_X^{1,s}(f_1, f_2) = \frac{1}{s-1} \left[ \exp \left\{ (s-1) \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx \right\} - 1 \right], \quad s \neq 1,$$

αντίστοιχα. Και πάλι είναι εύκολο να δούμε ότι  $\lim_{s \rightarrow 1} I_X^{r,s}(f_1, f_2) = r I_X^R(f_1, f_2)$ ,  $\lim_{r \rightarrow 1} I_X^{r,s}(f_1, f_2) = I_X^{1,s}(f_1, f_2)$  και  $\lim_{s \rightarrow 1} I_X^{1,s}(f_1, f_2) = I_X^{KL}(f_1, f_2)$ .

Ένα μέτρο απόκλισης, το οποίο έχει μια ειδική απήχηση καθώς προέρχεται από την εντροπία του Shannon και την κοίλη ιδιότητα αυτής είναι η διαφορά του Jensen όπως την ονόμασαν οι Burbea and Rao (1982). Η διαφορά του Jensen μεταξύ δυο συναρτήσεων

πυκνότητας πιθανότητας  $f_1$  και  $f_2$ , δίνεται από τη σχέση

$$J(f_1, f_2) \equiv H\left(\frac{1}{2}(f_1 + f_2)\right) - \frac{1}{2}[H(f_1) + H(f_2)],$$

όπου  $H(f) = -\int f(x) \ln f(x) dx$  είναι η εντροπία του Shannon.

Μια άλλη γνωστή γενίκευση των μέτρων απόκλισης είναι η οικογένεια των αποκλίσεων δύναμης (power divergences) που πρότειναν οι Cressie and Read (1984) και δίνεται από τη σχέση

$$I_X^{CR}(f_1, f_2) = \frac{1}{\lambda(\lambda + 1)} \int f_1(x) \left[ \left( \frac{f_1(x)}{f_2(x)} \right)^\lambda - 1 \right] dx$$

για  $\lambda \in R$  ενώ για  $\lambda = 0$  και  $\lambda = -1$  ορίζεται ως όριο. Είναι όμως εύκολο να δειχθεί ότι για  $\lambda \rightarrow 0$  έχουμε

$$\lim_{\lambda \rightarrow 0} I_X^{CR}(f_1, f_2) = \int f_1(x) \ln \frac{f_1(x)}{f_2(x)} dx = I_X^{KL}(f_1, f_2),$$

το οποίο είναι το μέτρο των Kullback - Leibler ενώ για  $\lambda \rightarrow -1$  έχουμε

$$\lim_{\lambda \rightarrow -1} I_X^{CR}(f_1, f_2) = \int f_2(x) \ln \frac{f_2(x)}{f_1(x)} dx = I_X^{KL}(f_2, f_1).$$

Να σημειώσουμε ότι το μέτρο των Cressie and Read ανήκει στην οικογένεια των  $\phi$  - αποκλίσεων για

$$\phi(x) = [\lambda(\lambda + 1)]^{-1}(x^{\lambda+1} - x).$$

Με τη διακριτή περίπτωση του μέτρου των Cressie and Read θα ασχοληθούμε εκτενέστερα σε επόμενο κεφάλαιο.

Αν και το μέτρο του Csiszar είναι ένα γενικό μέτρο, παρόλα αυτά δεν περιέχει όλα τα γνωστά μέτρα. Για παράδειγμα, τα μέτρα των Rényi, Sharma and Mittal και Bhattacharyya δεν μπορούν να γραφούν ως  $\phi$  - αποκλίσεις. Για το λόγο αυτό οι Menendez et al. (1995) εισήγαγαν και μελέτησαν την οικογένεια των  $(h, \phi)$  - αποκλίσεων, η οποία δίνεται από τη σχέση

$$I_X^{h,\phi}(f_1, f_2) = h\left(I_X^\phi(f_1, f_2)\right),$$

όπου η  $h$  είναι μια διαφορίσιμη αύξουσα πραγματική συνάρτηση με  $h(0) = 0$  και  $h'(0) > 0$  ενώ η  $\phi$  όπως στην οικογένεια των  $\phi$  - αποκλίσεων.

Ο Lin (1991) όρισε την Jensen - Shannon difference

$$JS(f_1, f_2) = H(af_1 + (1 - a)f_2) - aH(f_1) - (1 - a)H(f_2), \quad 0 < a < 1.$$

Το μέτρο αυτό είναι επέκταση της διαφοράς του Jensen, που αναφέραμε παραπάνω, καθώς

επίσης αποτελεί μια ειδική περίπτωση των  $\phi$  - αποκλίσεων για

$$\phi(x) = ax \ln x - (ax + 1 - a) \ln(ax + 1 - a),$$

όπως αναφέρουν οι Menendez et al. (1997).

Μια από τις τελευταία προτεινόμενες οικογένειες μέτρων απόκλισης είναι η ΒΗΗJ απόκλιση δύναμης, η οποία προτάθηκε από τους Basu et al. (1998) και ορίζεται ως εξής

$$I_X^a(f_1, f_2) = \int \left[ f_2^{1+a}(x) - \left(1 + \frac{1}{a}\right) f_1(x) f_2^a(x) + \frac{1}{a} f_1^{1+a}(x) \right] dx, \quad a > 0.$$

Είναι εύκολο να δούμε ότι ισχύει  $\lim_{a \rightarrow 0} I_X^a(f_1, f_2) = I_X^{KL}(f_2, f_1)$ .

Η παραπάνω βέβαια οικογένεια μέτρων μπορεί να θεωρηθεί ως ειδική περίπτωση του μέτρου Bregman (Jones and Byrne, 1990; Csiszar, 1991), το οποίο έχει τη γενική μορφή

$$I_X^{Br}(f_1, f_2) = \int [H(f_1(x)) - H(f_2(x)) - (f_1(x) - f_2(x))H'(f_2(x))] dx,$$

όπου  $H$  είναι μια κυρτή συνάρτηση.

Τέλος, ο Mattheou (2007) πρότεινε τη γενικευμένη οικογένεια των ΒΗΗJ αποκλίσεων δύναμης. Η οικογένεια αυτή ορίζεται ως

$$I_X^a(f_1, f_2) = \int f_1^{1+a}(x) \Phi\left(\frac{f_2(x)}{f_1(x)}\right) dx,$$

όπου  $\Phi$  είναι μια κυρτή συνάρτηση στο  $[0, \infty)$  τέτοια ώστε  $\Phi(1) = 0$ ,  $\Phi'(1) = 0$  και  $\Phi''(1) \neq 0$ . Επιπλέον υποθέτουμε ότι ισχύουν οι σχέσεις  $0\Phi(0/0) = 0$  και  $0\Phi(u/0) = \lim_{u \rightarrow \infty} \frac{\Phi(u)}{u}$ , για  $u > 0$ .

Η οικογένεια των μέτρων ΒΗΗJ μπορεί να θεωρηθεί ως μέλος της παραπάνω γενικής οικογένειας παίρνοντας

$$\Phi(u) = u^{1+a} - \left(1 + \frac{1}{a}\right) u^a + \frac{1}{a}.$$

Επιπλέον εάν πάρουμε  $\Phi(u) = \phi(u)$  και  $a = 0$  τότε η γενικευμένη οικογένεια ΒΗΗJ καταλήγει στο μέτρο του Csiszar.

Στο σημείο αυτό κρίνουμε σκόπιμο να παρουσιάσουμε τη σχέση που συνδέει τα παραμετρικά με τα μη παραμετρικά μέτρα πληροφορίας. Έστω

$$I_X(f(x|\theta), f(x|\theta + \Delta\theta))$$

ένα μη παραμετρικό μέτρο πληροφορίας για τις πυκνότητες  $f(x|\theta)$  και  $f(x|\theta + \Delta\theta)$ . Ένας τρόπος για να κατασκευάσουμε παραμετρικά μέτρα πληροφορίας από μη παραμετρικά μέτρα

δίνεται από τη σχέση:

$$I_X(\theta) = \lim_{\Delta\theta \rightarrow 0} \inf \frac{1}{(\Delta\theta)^2} I_X(f(x|\theta), f(x|\theta + \Delta\theta))$$

Η μέθοδος αυτή έχει χρησιμοποιηθεί από πολλούς συγγραφείς μεταξύ των οποίων οι Kagan (1963), Vajda (1973), Boekee (1979), Ferentinos and Papaionnou (1981) και Taneja (1987).

### Μέτρα Τύπου Εντροπίας

Το κυριότερο μέτρο της τρίτης κατηγορίας είναι η *εντροπία του Shannon* ή *διαφορική εντροπία*, όπως διαφορετικά ονομάζεται (Shannon, 1948). Η εντροπία δίνεται από τον τύπο

$$H(X) = H(p) = - \sum_x p(x) \ln p(x)$$

ή

$$H(X) = H(f) = - \int f(x) \ln f(x) dx$$

για τη διακριτή ή συνεχή περίπτωση αντίστοιχα. Το μέτρο αυτό ποσοτικοποιεί την αναμενόμενη αβεβαιότητα που σχετίζεται με το αποτέλεσμα ενός πειράματος τύχης. Με άλλα λόγια παρέχει πληροφορία για την προβλεψιμότητα του αποτελέσματος μιας τυχαίας μεταβλητής  $X$ . Όσο μεγαλύτερη η εντροπία τόσο λιγότερο συγκεντρωμένη είναι η κατανομή της  $X$  και συνεπώς μια παρατήρηση της  $X$  παρέχει λίγη πληροφορία. Γενικεύοντας την έννοια της εντροπίας σε δυο τυχαίες μεταβλητές, έχουμε την *από κοινού εντροπία* (mutual entropy)

$$H(X, Y) = - \sum_x \sum_y p(x, y) \ln p(x, y) = -E[\ln p(x, y)],$$

όπου  $p(x, y)$  είναι η από κοινού κατανομή των μεταβλητών  $X, Y$ . Η υπό συνθήκη εντροπία ορίζεται ως

$$H(Y|X) = - \sum_x \sum_y p(x, y) \ln p(y|x) = -E[\ln p(y|x)].$$

Από τον κανόνα της αλυσίδας ισχύει  $H(X, Y) = H(X) + H(Y|X)$ .

Για τη σχέση της εντροπίας με την αμοιβαία πληροφορία  $I(X, Y)$  και κατ' επέκταση με το μέτρο πληροφορίας των Kullback - Leibler, δοθέντος ότι  $I(X, Y) = I^{KL}(p(x, y), f(x)g(y))$  ισχύει το παρακάτω θεώρημα (Cover and Thomas, 1991, σελ. 20):

**Θεώρημα 2.2.** (*Mutual Information and entropy*) *Ισχύουν οι παρακάτω σχέσεις:*

1.  $I(X, Y) = H(X) - H(X|Y)$
2.  $I(X, Y) = H(Y) - H(Y|X)$

$$3. I(X, Y) = H(X) + H(Y) - H(X, Y)$$

$$4. I(X, Y) = I^{KL}(Y, X)$$

$$5. I(X, X) = H(X)$$

Για ονομαστικές (nominal) αποκρίσεις, ο Theil (1970) πρότεινε έναν δείκτη χρησιμοποιώντας την εντροπία του Shannon. Για πίνακες συνάφειας, η αναλογική μείωση (proportional reduction) της εντροπίας ισούται με

$$U = -\frac{\sum_i \sum_j \pi_{ij} \ln(\pi_{ij}/\pi_{i+}\pi_{+j})}{\sum_j \pi_{+j} \ln \pi_{+j}}$$

και ονομάζεται συντελεστής αβεβαιότητας (uncertainty coefficient). Το μέτρο αυτό ορίζεται σωστά όταν όλα τα  $\pi_{+j} > 0$ . Παίρνει τιμή μεταξύ του 0 και 1. Το  $U = 0$  υποδηλώνει την ανεξαρτησία των μεταβλητών  $X$  και  $Y$ . Το  $U = 1$  ισοδυναμεί με έλλειψη υπο συνθήκης διασποράς (conditional variation), με την έννοια ότι για κάθε  $i$ ,  $\pi_{j|i} = 1$  για κάποια  $j$  (Agresti, 2002).

Ένα δεύτερο μέτρο της κατηγορίας αυτής είναι ο δείκτης Gini - Simpson (Gini, 1912; Simpson, 1949). Έστω  $\mathbf{p} = (p_1, p_2, \dots, p_n)$  μια διακριτή κατανομή πιθανότητας. Τότε ο δείκτης Gini - Simpson ορίζεται ως

$$I^{GS}(\mathbf{p}) = 1 - \sum_{i=1}^n p_i^2.$$

Αργότερα εμφανίστηκαν άλλες γενικεύσεις της εντροπίας του Shannon. Πιο συγκεκριμένα προτάθηκε η εντροπία του Rényi (1961), η οποία δίνεται από τη σχέση

$$H^{R,a}(f) = \frac{1}{a-1} \ln \int f^a(x) dx, \quad a > 0, \quad a \neq 1$$

και η εκτεταμένη εντροπία του Rényi για όλα τα  $a \in \mathbb{R} - \{0, 1\}$ , η οποία προτάθηκε από τους Liese and Vajda (1987) και δίνεται από τη σχέση

$$H^{RLV,a}(f) = \frac{1}{a(a-1)} \ln \int f^a(x) dx, \quad a \neq 0, 1.$$

Η εκτεταμένη εντροπία του Rényi δεν ορίζεται για  $a = 1$  και  $a = 0$ . Είναι εύκολο όμως να δειχθεί ότι για  $a \rightarrow 1$  η εντροπία του Rényi ισούται με την εντροπία του Shannon ενώ για  $a \rightarrow 0$  ισχύει

$$\lim_{a \rightarrow 0} H^{RLV,a}(f) = \int \ln f(x) dx.$$

Κατά αντιστοιχία με την οικογένεια των  $\phi$  - αποκλίσεων, οι Burbea and Rao (1982a,



1982b, 1982c) εισήγαγαν την οικογένεια των  $\phi$  - εντροπιών, η οποία περιλαμβάνει αρκετές από τις εντροπίες που έχουν προταθεί (Pardo, 2006). Η οικογένεια των  $\phi$  - εντροπιών ορίζεται από τη σχέση

$$H^\phi(f) = \int \phi(f(x)) dx,$$

όπου η  $\phi$  είναι μια συνεχής κυρτή συνάρτηση που ορίζεται στο  $(0, \infty)$  και ισχύει  $\phi(0) = \lim_{u \rightarrow 0} \phi(u) \in (-\infty, \infty]$ .

Όμως, όπως συμβαίνει και με την οικογένεια των  $\phi$  - αποκλίσεων, έτσι και ορισμένες σημαντικές εντροπίες δεν μπορούν να γραφούν ως  $\phi$  - εντροπίες. Το πρόβλημα αυτό έρχεται να λύσει μια ακόμη πιο γενική οικογένεια, την οποία πρότειναν οι Salicru et al. (1993) και ονομάζεται  $(h, \phi)$  - εντροπία. Η οικογένεια αυτή ορίζεται ως

$$H^{h,\phi}(f) = h \left( \int \phi(f(x)) dx \right),$$

όπου είτε η συνάρτηση  $\phi : (0, \infty) \rightarrow R$  είναι κοίλη και η συνάρτηση  $h : R \rightarrow R$  διαφορίσιμη και αύξουσα είτε η συνάρτηση  $\phi$  είναι κυρτή και η  $h : R \rightarrow R$  διαφορίσιμη και φθίνουσα. Τα μέτρα που ανήκουν στις οικογένειες των  $\phi$  και  $(h, \phi)$  - εντροπιών, δίνονται στον Πίνακα 2.2 (Pardo, 2006).

$\phi(x)$	$h(x)$	Εντροπία
$-x \ln x$	$x$	Shannon (1948)
$x^{r-m+1}$	$(m-r)^{-1} \ln x$	Rényi (1961) ( $r \neq 0, r \neq 1$ )
$(1-s)^{-1}(x^s - x)$	$x$	Havrda and Charvat (1967) ( $s \neq 1, s > 0$ )
$x^{\frac{1}{t}}$	$(t-1)^{-1}(x^t - 1)$	Arimoto (1971) ( $t \neq 1, t > 0$ )
$x \ln x$	$\frac{\exp[(s-1)x]-1}{1-2}$	Sharma and Mittal (1975) ( $s \neq 1, s > 0$ )
$(1+\lambda x) \ln(1+\lambda x)$	$(1+\frac{1}{\lambda}) \ln(1+\lambda) - \frac{x}{\lambda}$	Ferreri (1980) ( $\lambda > 0$ )
$\frac{x^s+(1-x)^s-1}{1-s}$	$x$	Kapur (1972) ( $s \neq 1$ )
$\frac{x^s-(1+x)^s+1+(s-1)^{-1}(2^s-2)x}{s-2}$	$x$	Burbea (1984)

Πίνακας 2.2:  $\phi$  - εντροπίες και  $(h, \phi)$  - εντροπίες

Οι Burbea and Rao (1982a, 1982b, 1982c) όρισαν επίσης την οικογένεια των  $R_\phi$  - αποκλίσεων, η οποία δίνεται από τη σχέση

$$R_\phi(f_1, f_2) = H_\phi \left( \frac{f_1 + f_2}{2} \right) - \frac{H_\phi(f_1) + H_\phi(f_2)}{2}$$

ενώ οι Pardo et al. (1993) γενίκευσαν την παραπάνω οικογένεια, με τη βοήθεια της  $(h, \phi)$

- εντροπίας, ορίζοντας την

$$R_{\phi}^h(f_1, f_2) = H_{\phi}^h\left(\frac{f_1 + f_2}{2}\right) - \frac{H_{\phi}^h(f_1) + H_{\phi}^h(f_2)}{2}.$$

Χρησιμοποιώντας

$$\phi(x) = \begin{cases} (1-a)^{-1}(x^a - x), & a \neq 1, a > 0 \\ -x \ln x, & a = 1 \end{cases},$$

η οικογένεια των  $R_{\phi}$  - αποκλίσεων οδηγεί σε μια άλλη σημαντική οικογένεια αποκλίσεων, την  $R_{\phi}^a$  - απόκλιση που πρότειναν οι Havrda and Charvat (1967).

### Μέτρα κατά Bayes

Μια τέταρτη κατηγορία μπορεί να συμπεριλάβει τα μέτρα που προέρχονται από την μπεϋζιανή προοπτική. Ο κύριος εκπρόσωπος της κατηγορίας αυτής είναι το μέτρο του Lindley (1956). Δίνεται από τον τύπο

$$\begin{aligned} I_X^L(\theta) &= H(\xi) - E[H(\xi|x)] \\ &= - \int \xi(\theta) \ln \xi(\theta) d\theta - \left\{ \left( - \int \xi(\theta|x) \ln \xi(\theta|x) d\theta \right) p(x) dx \right\}, \end{aligned}$$

όπου  $\xi(\theta)$ ,  $p(x)$  και  $\xi(\theta|x)$  είναι η εκ των προτέρων πιθανότητα, η περιθωριακή πιθανότητα και η εκ των υστέρων πιθανότητα, αντίστοιχα. Για  $f_1(x) = p(x, \theta)$  και  $f_2(x) = p(x)\xi(\theta)$ , όπου  $p(x, \theta)$  είναι η από κοινού πυκνότητα των  $X$  και  $\theta$ , παίρνουμε το μέτρο των Kullback - Leibler.

Ένα άλλο μέτρο της κατηγορίας αυτής είναι το μέτρο του Mallows (1959),

$$I_X^{Ma}(\theta) = \int p(x|\theta) \{W(\theta, \xi(\theta|x)) - W(\theta, \xi(\theta))\} dx,$$

όπου  $W(\theta, \xi(\cdot))$  είναι ένα αριθμητικό μέτρο της γνώσης του ερευνητή για την παράμετρο  $\theta$  και  $p(x|\theta)$  η πυκνότητα της τυχαίας μεταβλητής  $X$  δοθείσης της παραμέτρου  $\theta$ .

Για περισσότερες πληροφορίες για τα παραπάνω μέτρα πληροφορίας παραπέμπουμε στους Soofi (1994), Papaioannou (1985 και 2001) και Papaioannou and Ferentinos (2002).

Η πληθώρα των μέτρων πληροφορίας τα οποία χρησιμοποιούνται για τόσα χρόνια και το γεγονός ότι νέα μέτρα (που σχετίζονται ή όχι με τα κλασικά μέτρα) συνεχίζεται να εισάγονται δείχνει όχι μόνο τις διαφορετικές προοπτικές αλλά επίσης την αδυναμία να οριστεί μονοσήμαντα (uniquely) η έννοια της στατιστικής πληροφορίας.

### 2.1.3 Ιδιότητες Μέτρων Πληροφορίας

Τα μέτρα πληροφορίας έχουν κάποιες ιδιότητες τις οποίες πρέπει να ικανοποιούν τα νεοεμφανιζόμενα μέτρα. Αναφέρθηκε προηγουμένως ότι ο αριθμός πληροφορίας του Fisher δεν πρέπει να χαρακτηρίζεται ως μέτρο πληροφορίας καθώς δεν ικανοποιεί ορισμένες βασικές ιδιότητες του μέτρου του Fisher. Φυσικά, θα ήταν παράλογο και μη ρεαλιστικό όλα τα μέτρα που εμφανίζονται να ικανοποιούν όλες ανεξαιρέτως τις ιδιότητες. Οι στατιστικοί ακόμη δεν έχουν συμφωνήσει σε ένα σύνολο βασικών και λογικών ιδιοτήτων, τις οποίες τα νεοεμφανιζόμενα μέτρα θα πρέπει να ικανοποιούν.

Ας δούμε τώρα τις πιο βασικές και επιθυμητές ιδιότητες, όπως τις παρουσιάζει ο Paraiouannou (1985 και 2001).

Έστω  $I_X$  οποιοδήποτε μέτρο της πληροφορίας που περιέχεται στα δεδομένα  $\mathbf{X}$  και  $T(\mathbf{X})$  μια οποιοδήποτε στατιστική συνάρτηση του  $\mathbf{X}$ . Σημειώνουμε ότι για τα παραμετρικά μέτρα, απαιτείται επιπλέον η εκπλήρωση ορισμένων συνθηκών ομαλότητας.

#### 1. Μη αρνητικότητα (Nonnegativity)

Ισχύει  $I_X \geq 0$  με ισότητα αν και μόνο αν η κατανομή είναι ανεξάρτητη του  $\theta$  ή η οικογένεια κατανομών αποτελείται από μια μόνο κατανομή. Η ιδιότητα αυτή δεν ισχύει μόνο για το μέτρο του Csiszar και για τις συνεχείς περιπτώσεις των εντροπιών του Shannon και του Rényi. Το μέτρο του Csiszar, ωστόσο ικανοποιεί την πιο γενική ανισότητα  $I_X^C \geq \phi(1)$ .

#### 2. Προσθετικότητα - Υποπροσθετικότητα (Additivity - Subadditivity)

Υπάρχουν δυο είδη προσθετικότητας: η αδύναμη και η δυνατή. Η αδύναμη προσθετικότητα δίνεται από τη σχέση  $I_{X,Y} = I_X + I_Y$  αν οι  $X, Y$  είναι ανεξάρτητες. Η δυνατή, δίνεται από τη σχέση  $I_{X,Y} = I_X + I_{Y|X}$ , όπου  $I_{Y|X} = E_X[I_{Y|X=x}]$  ενώ η υποπροσθετικότητα από τη σχέση  $I_{X,Y} \leq I_X + I_Y$  με ισότητα αν η  $X$  είναι ανεξάρτητη της  $Y$ .

Η δυνατή προσθετικότητα υπονοεί την αδύναμη. Η αδύναμη προσθετικότητα ισχύει για όλα τα μέτρα εκτός από αυτά των Matusita, Kagan, Vajda και συνεπώς των Mathai και Boeke. Η δυνατή προσθετικότητα ισχύει μόνο για τα μέτρα των Fisher, Kullback - Leibler και Shannon. Η υποπροσθετικότητα ισχύει μόνο για την εντροπία του Shannon, ενώ η ισχύς της για το μέτρο του Csiszar εξαρτάται από την επιλογή της συνάρτησης  $\phi$ .

#### 3. Υπό Συνθήκη Ανισότητα (Conditional Inequality)

Ισχύει  $I_{Y|X} = I_X$ . Αν ισχύουν οποιοδήποτε δυο από τις ιδιότητες: δυνατή προσθετικότητα, υποπροσθετικότητα και υπό συνθήκη ανισότητα, τότε ισχύει και η τρίτη. Η υπό συνθήκη ανισότητα ικανοποιείται από την εντροπία του Shannon αλλά όχι από τα μέτρα των

Fisher και Kullback - Leibler.

#### 4. Μέγιστη Δυνατή Πληροφορία (Maximal Information)

Η ιδιότητα αυτή δίνεται από τη σχέση  $I_{T(X)} \leq I_X$  και δεν ικανοποιείται μόνο από τη συνεχή περίπτωση των εντροπιών του Shannon και του Rényi.

#### 5. Αμεταβλητότητα (Invariance) υπό επαρκείς μετασχηματισμούς

Ισχύει  $I_X = I_{T(X)}$  αν και μόνο αν η στατιστική συνάρτηση  $T(X)$  είναι επαρκής. Ισχύει για όλα τα μέτρα εκτός από τις εντροπίες των Shannon και Rényi.

#### 6. Κυρτότητα (Convexity)

Έστω  $a_1 \geq 0, a_2 \geq 0$  με  $a_1 + a_2 = 1$  και  $f_1, f_2$  δυο πυκνότητες. Τότε ισχύει  $I_X(a_1 f_1 + a_2 f_2) \leq a_1 I_X(f_1) + a_2 I_X(f_2)$ . Η κυρτότητα δεν ισχύει για τα μέτρα των Rényi με  $a > 1$ , Matusita και τις εντροπίες των Shannon και Rényi.

#### 7. Απώλεια Πληροφορίας (Loss of Information)

Έστω  $G$  το σύνολο όλων των διαμερίσεων  $g$  του  $R^k$  και  $I_g$  το μέτρο πληροφορίας για το  $g$ . Τότε  $\sup_{g \in G} I_g = I_X$  ή  $I_g \rightarrow I_X$  καθώς η λεπτότητα (refinement)  $\lambda(g)$  του  $g$  τείνει στο μηδέν. Αυτό σημαίνει ότι η απώλεια πληροφορίας που οφείλεται σε ομαδοποίηση των παρατηρήσεων, μπορεί να γίνει αυθαίρετα μικρή επιλέγοντας κατάλληλη διαμέριση  $g$ . Σημειώνεται ότι αν η στατιστική συνάρτηση είναι επαρκής, δεν υπάρχει απώλεια πληροφορίας. Η ιδιότητα αυτή ισχύει για όλα τα μέτρα εκτός από τη συνεχή περίπτωση των εντροπιών των Shannon και Rényi.

#### 8. Επάρκεια στα Πειράματα (Sufficiency in Experiments)

Αν τα πειράματα  $E_X$  και  $E_Y$  έχουν τον ίδιο παραμετρικό χώρο  $\Theta$  και το  $E_X$  είναι επαρκές για το  $E_Y$  σύμφωνα με τον ορισμό του Blackwell, τότε  $I_X \geq I_Y$ . Η ιδιότητα αυτή ισχύει για όλα τα μέτρα πληροφορίας με διαφορετικούς περιορισμούς το καθένα.

#### 9. Εμφάνιση στις Ανισότητες Cramer - Rao.

Όλα τα παραμετρικά μέτρα εμφανίζονται στις ανισότητες τύπου Cramer - Rao με την προϋπόθεση ότι ισχύουν ορισμένες συνθήκες ομαλότητας (Παπαϊωάννου και Φερεντίνος, 2000). Για παράδειγμα, αν  $g(\theta)$  είναι μια παραμετρική συνάρτηση του  $\theta$ , τότε για το μέτρο

πληροφορίας του Fisher ισχύει  $E[T(X) - g(\theta)]^2 \geq \frac{[g'(\theta)]^2}{I_X^F(\theta)}$ .

10. Αμεταβλητότητα (Invariance) υπό παραμετρικούς μετασχηματισμούς

Η ιδιότητα αυτή δεν ισχύει για τα μέτρα των Fisher, Vajda, Matusita και Boekee.

11. Ανισότητα της Ενοχλητικής Παραμέτρου (Nuisance Parameter Inequality)

Αν  $\theta_2$  είναι ενοχλητική παράμετρος, τότε  $I_X(\theta_1, \theta_2) \leq I_X(\theta_1)$ . Αυτό σημαίνει ότι καθώς ο αριθμός των παραμέτρων αυξάνει η πληροφορία μειώνεται.

12. Ιδιότητα Διατήρησης Σειράς (Order Preserving Property)

Έστω  $I$  ένα μέτρο το οποίο αποδεχόμαστε ως τυπικό και  $J$  ένα οποιοδήποτε άλλο μέτρο. Τότε  $I_{X_1} \leq I_{X_2} \Rightarrow J_{X_1} \leq J_{X_2}$ . Η ιδιότητα αυτή είναι φυσική με την έννοια ότι ένα μέτρο πληροφορίας οποιουδήποτε τύπου μετρά την ποσότητα της πληροφορίας που υπάρχει και συνεπώς εάν μια τυχαία μεταβλητή περιέχει περισσότερη πληροφορία από μια άλλη τυχαία μεταβλητή για ένα συγκεκριμένο μέτρο τότε είναι λογικό να αναμένουμε να περιέχει περισσότερη πληροφορία για οποιοδήποτε μέτρο. Αυτό ισχύει αν το  $I$  είναι η εντροπία του Shannon και το  $J$  η εντροπία του Rényi με  $a > 1$ .

13. Ασυμπτωτική Συμπεριφορά

Η ακολουθία των τυχαίων μεταβλητών  $X_n$  συγκλίνει, όταν  $n \rightarrow \infty$ , στην τυχαία μεταβλητή  $X$  αν και μόνο αν  $I_{X_n} \rightarrow I_X$ . Υπό ορισμένες συνθήκες η ιδιότητα αυτή ισχύει για τα μέτρα των Kullback - Leibler και Rényi.

Για περισσότερες πληροφορίες, ιδιότητες και παραδείγματα ο αναγνώστης παραπέμπεται στους Papaiοannου (2001) και Kullback (1959).

Από τις παραπάνω ιδιότητες δικαιολογείται η άποψη των στατιστικών επιστημόνων ότι περισσότερα δεδομένα σημαίνουν περισσότερη πληροφορία. Συγκεκριμένα, σε αυτό συμβάλλουν οι ιδιότητες της προσθετικότητας, μη αρνητικότητας και μέγιστης δυνατής πληροφορίας.

## 2.2 Θεωρία Πληροφοριών και Στατιστική

Η Στατιστική, η επιστήμη της εξαγωγής πληροφορίας από δεδομένα, εμφανίζεται ως το φυσικό πεδίο εφαρμογής της Θεωρίας Πληροφοριών, πέρα από τη Θεωρία Επικοινωνιών. Ιστορικά, ένα μέτρο πληροφορίας είχε χρησιμοποιηθεί από τους στατιστικούς επιστήμονες πριν από τον Shannon. Πρόκειται για το μέτρο πληροφορίας του Fisher (Fisher, 1925). Οι στατιστικοί ερευνητές έχουν παίξει κυρίαρχο ρόλο στην ανάπτυξη της Θεωρίας Πληροφοριών και έχουν αποδείξει ότι αποτελεί ένα πλαίσιο για ενασχόληση με πολλά και διάφορα

προβλήματα στατιστικής φύσεως. Αρκετά αποτελέσματα εάν εκ των υστέρων κοιταχθούν ως εφαρμογές της Θεωρίας Πληροφοριών στη Στατιστική, στην πραγματικότητα είχαν θεμελιωθεί από στατιστικούς ανεξάρτητα από τη Θεωρία Πληροφοριών. Όπως αναφέρει ο Kullback (1959, σελ. 2): «Αν και ο Wald δεν ανέφερε ξεκάθαρα τον όρο πληροφορία στην αντιμετώπιση της ακολουθιακής ανάλυσης (sequential analysis), θα πρέπει να αναφερθεί ότι η δουλειά του πρέπει να θεωρηθεί ως σημαντική συνεισφορά στις στατιστικές εφαρμογές της Θεωρίας Πληροφοριών».

Ο κύριος στόχος της Θεωρίας Πληροφοριών είναι να μετρήσει την ποσότητα της πληροφορίας που εμπλέκεται σε στατιστικά προβλήματα. Συγκεκριμένα στα προβλήματα που εμπλέκεται δειγματοληψία, αναζητούμε την ποσοτικοποίηση της πληροφορίας που φέρουν τα δεδομένα για τις άγνωστες παραμέτρους που θέλουμε να εκτιμήσουμε (Kullback, 1959). Σημαντικό ρόλο παίζει η Θεωρία Πληροφοριών στη σημειακή εκτιμητική, αφού το μέτρο πληροφορίας του Fisher εμφανίζεται στις ανισότητες Cramer - Rao. Επίσης τα μέτρα πληροφορίας χρησιμοποιούνται στον έλεγχο στατιστικών υποθέσεων, κυρίως στα τεστ σημαντικότητας, καλής προσαρμογής και ανεξαρτησίας, καθώς και στην μεϋζιανή στατιστική, όπου τα μέτρα τύπου απόκλισης χρησιμοποιούνται για τον προσδιορισμό των εκ των προτέρων πιθανοτήτων. Άλλη εφαρμογή έχει η Στατιστική Θεωρία Πληροφοριών στα διακριτά δεδομένα και συγκεκριμένα στην ανάλυση πινάκων συνάφειας και αφορά την κατασκευή μοντέλων που είναι όσο το δυνατόν πιο κοντά, από την πλευρά του μέτρου απόκλισης των Kullback - Leibler και της  $\phi$ -απόκλισης, στο μοντέλο συμμετρίας (Kateri and Papaioannou; 1994, 1997). Επίσης στην επιλογή του καταλληλότερου μοντέλου στην παλινδρόμηση. Η ανάπτυξη της Θεωρίας Πληροφοριών θέτει τις βάσεις για την χρησιμοποίηση των μέτρων πληροφορίας και σε άλλους τομείς όπως είναι για παράδειγμα η ανάπτυξη σε logit μοντέλα μεθόδων παρόμοιων με αυτές που χρησιμοποιούνται στην ανάλυση κυρίων συνιστωσών στην παλινδρόμηση (βλέπε Soofi, 1994; Papaioannou, 1970 και 1985, και αντίστοιχες αναφορές). Πληροφοριακά αναφέρουμε ότι η Θεωρία Πληροφοριών μπορεί επίσης να χρησιμοποιηθεί στην πολυμεταβλητή ανάλυση, στα τεστ ομοιομορφίας, στους ελέγχους εκθετικότητας, στη σύγκριση πειραμάτων, την μεϋζιανή ανάλυση αξιοπιστίας καθώς και στον στατιστικό έλεγχο ποιότητας.

Στη θεωρία της σημειοεκτιμητικής, το μέτρο πληροφορίας του Fisher παίζει κυρίαρχο ρόλο καθώς εμφανίζεται στις ανισότητες τύπου Cramer - Rao. Εδώ βέβαια υπάρχουν κάποιοι περιορισμοί καθώς το κάτω φράγμα των Cramer - Rao εφαρμόζεται σε κανονικές (regular) οικογένειες κατανομών. Οι Ferentinos and Papaioannou (1981), προσπάθησαν να μετρήσουν την πληροφορία σχετικά με την παράμετρο  $\theta$  που περιέχεται στα δεδομένα για ευρύτερες οικογένειες κατανομών συμπεριλαμβανομένων και μη κανονικών (non regular) αλλά τα αποτελέσματα δεν ήταν ικανοποιητικά καθώς τα νέα παραμετρικά μέτρα ήταν περίπλοκα και περιοριστικές περιπτώσεις των μέτρων τύπου απόκλισης (divergence type). Το γεγονός αυτό οδήγησε τους συγγραφείς να εικάσουν ότι το μέτρο πληροφορίας του Fisher είναι μοναδικό up to a constant.

Οι ίδιοι συγγραφείς σκέφτηκαν να χρησιμοποιήσουν τα μέτρα πληροφορίας ως συναρτήσεις απώλειας (loss functions) και να παράγουν με αυτό τον τρόπο τους καλύτερους από πλευράς πληροφορίας (information optimum) εκτιμητές. Να παράγουν δηλαδή εκτιμητές ελαχιστοποιώντας τον κίνδυνο  $E(I(f_{\hat{\theta}}, f_{\theta}))$ , για όλους τους εκτιμητές  $\hat{\theta}$  που ανήκουν σε μια συγκεκριμένη ομάδα ομοιόμορφα στο  $\theta$ . Κατέληξαν σε αποτελέσματα ανάλογα με αυτά της θεωρίας αποφάσεων στη σημειοεκτιμητική, αλλά τα αποτελέσματα αυτά βασίζονταν κατά πολύ στην κυρτότητα του  $I(f_{\theta}, f_{\theta^*})$  ως προς το  $\theta^*$  για όλα τα  $\theta$  και γι' αυτό δεν συνέχισαν περαιτέρω τη διερεύνηση του θέματος. Είναι γνωστό ότι για την εκθετική οικογένεια κατανομών, οι εκτιμητές μέγιστης πιθανοφάνειας  $\hat{\theta}$  του  $\theta$  συγκλίνει σχεδόν βέβαια στο  $\theta^*$ , την τιμή της παραμέτρου που ελαχιστοποιεί το μέτρο απόκλισης των Kullback - Leibler  $I^{KL}(f, f(x, \theta))$  μεταξύ της πραγματικής κατανομής των δεδομένων  $f$  και της  $f(x, \theta)$  (McCulloch, 1989). Τα μέτρα πληροφορίας τύπου απόκλισης έχουν χρησιμοποιηθεί στη γενική θεωρία της εκτιμητικής σαν συναρτήσεις απώλειας για να εισαχθούν οι εκτιμητές ελάχιστης απόστασης / απόκλισης και να παραχθούν γενικεύσεις των κλασικών αποτελεσμάτων όπως είναι η συνέπεια, η ασυμπτωτική κανονικότητα κ.α. (Papaioannou, 1970).

Μια πολλά υποσχόμενη εφαρμογή των μέτρων πληροφορίας στη Μπεϋζιανή εκτιμητική είναι να χρησιμοποιηθούν μέτρα τύπου απόκλισης για τον προσδιορισμό των εκ των προτέρων κατανομών που χρειάζονται στη Μπεϋζιανή ανάλυση, έτσι ώστε να ικανοποιούνται διάφορα κριτήρια, όπως να ελαχιστοποιείται η απόκλιση μεταξύ των εκ των προτέρων και εκ των υστέρων κατανομών ή οι κατά Bayes εκτιμητές των παραμέτρων να είναι όσο το δυνατόν πιο κοντά προς τους κλασικούς εκτιμητές.

Στον έλεγχο στατιστικών υποθέσεων, τα μέτρα πληροφορίας έχουν χρησιμοποιηθεί για την ανάπτυξη ελέγχων σημαντικότητας. Η βασική ιδέα είναι να θεωρήσουμε έναν δειγματικό εκτιμητή του μέτρου πληροφορίας και να βρούμε την ασυμπτωτική του κατανομή, η οποία συνήθως είναι η κανονική ή η  $\chi^2$ .

Οι εφαρμογές των μέτρων πληροφορίας είναι πολυάριθμες. Ορισμένα μόνο από τα θέματα στα οποία εμπλέκονται τα μέτρα πληροφορίας είναι η Μπεϋζιανή στατιστική θεωρία πληροφοριών, η διαφορική γεωμετρία και πληροφορία, η παρατηρούμενη πληροφορία του Fisher και ο αλγόριθμος EM, η αρχή της μέγιστης εντροπίας, τα μέτρα εξάρτησης κλπ. Επιπλέον, τα μέτρα πληροφορίας χρησιμοποιήθηκαν για τη μέτρηση της πληροφορίας και την εισαγωγή και μελέτη νέων στατιστικών μοντέλων από τους Tsairidis et al. (1996) στην περίπτωση της τυχαίας λογοκρισίας, quantal τυχαίας λογοκρισίας με τυχαία ή μη καταγραφή μη λογοκριμένων παρατηρήσεων. Επιπλέον οι Kateri and Papaioannou (1994, 1997) κατασκεύασαν μοντέλα ψευδο-συμμετρίας (quasi-symmetry models) για πίνακες συνάφειας βασισμένα στο μέτρο απόκλισης των Kullback - Leibler και στην  $\phi$  - απόκλιση, τα οποία είναι όσο το δυνατόν πιο κοντά στο μοντέλο συμμετρίας.

## 2.3 Θεωρία Πληροφοριών και Βιοστατιστική

Όπως έχουμε ήδη αναφέρει, η σχετική εντροπία είναι ένα μέτρο της απόστασης μεταξύ δυο κατανομών πιθανότητας. Ο Benish (1999) πρότεινε ότι η ποσότητα της πληροφορίας που κερδίζουμε διενεργώντας έναν διαγνωστικό έλεγχο μπορεί να ποσοτικοποιηθεί υπολογίζοντας τη σχετική εντροπία μεταξύ των μετά του ελέγχου και πριν του ελέγχου κατανομές πιθανότητας. Η στατιστική αυτή συνάρτηση, ποσοτικοποιεί το βαθμό στον οποίο τα αποτελέσματα ενός διαγνωστικού ελέγχου είναι πιθανόν να μειώσουν την έκπληξή μας σχετικά με την τελική μας γνώση για τη διάγνωση του ασθενούς.

Για την επιλογή του κατάλληλου διαγνωστικού ελέγχου, ισχύει το παρακάτω: Ο πιο ειδικός (specific) έλεγχος χρησιμοποιείται για να επιβεβαιώσει τη διάγνωση και ο πιο ευαίσθητος (sensitive) χρησιμοποιείται για να επιβεβαιώσει ότι η ασθένεια δεν είναι πιθανή. Οι παραπάνω έννοιες μπορούν να χαρακτηρισθούν από το λόγο πιθανοφάνειας (likelihood ratio). Ο Lee (1999) πρότεινε τη χρήση της απόκλισης των Kullback - Leibler ως ένα νέο μέτρο μεταξύ της επιβεβαίωσης ή μη της ασθένειας. Το μέτρο αυτό ενοποιεί με κατάλληλο τρόπο δυο πηγές πληροφορίας - την κατανομή των αποτελεσμάτων του ελέγχου και τη συνάρτηση του λόγου πιθανοφάνειας και μετρά την ικανότητα του διαγνωστικού ελέγχου στην αναθεώρηση της πιθανότητας της ασθένειας (odds). Επίσης το μέτρο αυτό προβλέπει τη μοίρα (fate) του μέσου ατόμου πριν τον έλεγχο.

Το μέγεθος της διαγνωστικής αβεβαιότητας μετά από οποιοδήποτε αποτέλεσμα ενός ελέγχου μπορεί να ποσοτικοποιηθεί μέσω της Στατιστικής Θεωρίας Πληροφοριών. Η ποσοτικοποίηση της πληροφορίας έχει μεγάλη σχέση με την επιλογή και τη χρήση διαγνωστικών ελέγχων, καθώς κάποιος μπορεί να αναλύσει υποκειμενικά την αξία διαφορετικών κριτηρίων ερμηνείας, να συγκρίνει μεταξύ τους ελέγχους και να αξιολογήσει την αποτελεσματικότητα κόστους (cost-effectiveness) τόσο ενός απλού ελέγχου όσο και πιθανών συνδυασμών ελέγχων. Η εφαρμογή της Στατιστικής Θεωρίας Πληροφοριών στους διαγνωστικούς ελέγχους στην Ιατρική προκύπτει από την αναγνώριση ότι όλοι οι κλινικοί έλεγχοι είναι ελλιπείς. Η ατέλεια αυτή εισάγει αβεβαιότητα (ή «θόρυβο») στην ερμηνεία του ελέγχου. Η πληροφορία (και, αντιστρόφως, η αβεβαιότητα που υπάρχει πριν και μετά από έναν διαγνωστικό έλεγχο) μπορεί να ποσοτικοποιηθεί εάν κάποιος γνωρίζει τα τρία συστατικά τους. Αυτά είναι η προ του ελέγχου πιθανότητα ασθένειας (επιπολασμός - prevalence), η υπό συνθήκη πιθανότητα της εμφάνισης της νόσου στους ασθενείς (true-positive rate) και η υπό συνθήκη πιθανότητα της εμφάνισης της νόσου στους μη ασθενείς (false-positive rate). Η αλλαγή στη διαγνωστική αβεβαιότητα που εμφανίζεται ως αποτέλεσμα του ελέγχου, αποτελεί ένα μέτρο της αποτελεσματικότητας του ελέγχου.

Η Ανάλυση Επιβίωσης είναι ένα σύνολο στατιστικών μεθόδων για την ανάλυση δεδομένων, τα οποία αποτελούν μετρήσεις μιας μεταβλητής που δηλώνει το χρόνο μέχρι να συμβεί ένα ενδεχόμενο. Συνήθως το ενδεχόμενο είναι ο θάνατος, αν αναφερόμαστε σε ανθρώπους. Είναι ένας τελευταία ανεπτυγμένος κλάδος που ασχολείται με μοντέλα επιβίωσης,



δηλαδή παραμετρικά και μη παραμετρικά μοντέλα που περιγράφουν το χρόνο ζωής ατόμων ή αντικειμένων. Ασχολείται επίσης με την εκτίμηση της συνάρτησης επιβίωσης, τη σύγκριση καμπυλών επιβίωσης καθώς και με την επιλογή κατάλληλων μοντέλων. Η Ανάλυση Επιβίωσης παίζει σημαντικό ρόλο στη βιοστατιστική, τη μηχανική και την αναλογιστική επιστήμη. Η Στατιστική Θεωρία Πληροφοριών χρησιμοποιείται στην Ανάλυση Επιβίωσης, αφού το μέτρο πληροφορίας του Fisher, χρησιμοποιείται για την εκτίμηση της ασυμπτωτικής διακύμανσης της παραμέτρου  $b$  του αναλογικού μοντέλου του Cox καθώς επίσης και στην επιλογή του καλύτερου μοντέλου που περιγράφει την κατανομή του χρόνου ζωής (Johnson and Johnson, 1980). Επειδή βασικό στοιχείο της Ανάλυσης Επιβίωσης είναι τα λογοκριμένα δεδομένα, τα μέτρα πληροφορίας χρησιμοποιούνται για την ποσοτικοποίηση της αβεβαιότητας που περιέχουν τα λογοκριμένα δεδομένα. Περισσότερες πληροφορίες για τα μέτρα αυτά και τις ιδιότητές τους, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στους Tsairidis et al. (1996, 2001) και Papaioannou and Tsairidis (2001).

Στις κλινικές δοκιμές Φάσης I, τα σχέδια φαρμακοκινητικής δειγματοληψίας συχνά πρέπει να τροποποιηθούν καθώς οι δοκιμές βρίσκονται σε εξέλιξη. Οι τροποποιήσεις βασίζονται στις παρατηρήσεις από έναν περιορισμένο αριθμό υποκειμένων ή σε ορισμένες ακραίες περιπτώσεις από ένα μόνο άτομο. Οι Yafune and Ishiguro (2001) περιέγραψαν τη χρήση του μέτρου απόκλισης των Kullback - Leibler για την τροποποίηση των σχεδίων φαρμακοκινητικής δειγματοληψίας σε κλινικές δοκιμές Φάσης I. Η προσέγγιση αυτή προϋποθέτει ότι το φαρμακοκινητικό προφίλ για κάθε άτομο καθορίζεται από ένα κατάλληλο φαρμακοκινητικό μοντέλο. Η καλύτερη τροποποίηση επιλέγεται από αυθαίρετα επιλεγμένες υποψήφιας με βάση το μέτρο των Kullback - Leibler.

Οι translational κλινικές δοκιμές είναι μικρού μεγέθους μελέτες θεραπειών που εμφανίζονται από το εργαστήριο. Οι δοκιμές αυτές είναι ουσιώδεις και απαραίτητες για την παραγωγή πρώιμης ένδειξης που αφορούν επιπτώσεις της θεραπείας σε συγκεκριμένες εκφάνσεις της εξέλιξης μιας νόσου και για την καθοδήγηση μελλοντικών μελετών. Οι στατιστικές ιδιότητες τέτοιων μελετών έχουν παραμεληθεί, κυρίως, επειδή δεν προσαρμόζονται στο γνωστό εξελικτικό παράδειγμα των κλινικών μελετών (clinical trials developmental paradigm). Ο Piantadosi (2005) περιγράφει μια προσέγγιση, η οποία στηρίζεται στην πληροφορία και συγκεκριμένα στην εντροπία, για την κατανόηση των ιδιοτήτων και της χρήσης τέτοιων κλινικών δοκιμών. Ο συνδυασμός της βιολογικής γνώσης με έναν πειραματικό σχεδιασμό είναι ένα δυναμικό μέσο για την εξάλειψη σε σημαντικό βαθμό της αβεβαιότητας που περιβάλλει πρωτοεμφανιζόμενες θεραπευτικές μεθόδους.

Ο έλεγχος της βιοϊσοδυναμίας (bioequivalence) έχει παραδοσιακά επικεντρωθεί σε μεταβλητές όπως η AUC, η  $C_{max}$  και η  $t_{max}$ , οι οποίες φιλτράρουν την «εσωτερική» πληροφορία που μεταφέρεται από διακριτές ακολουθιακές παρατηρήσεις χρόνου συγκέντρωσης (sequential concentration-time observations). Συγκρίνοντας ολόκληρα προφίλ χρόνων συγκέντρωσης μεταξύ σκευασμάτων ελέγχου και αναφοράς για σκοπούς βιοϊσοδυναμίας παρέχει ισχυ-

ρότερη ένδειξη σχετικά είτε με την ομοιότητά τους είτε με την ανομοιότητά τους. Ο Pereira (2007) χρησιμοποίησε το κριτήριο πληροφορίας των Kullback - Leibler για κάθε χρόνο συγκέντρωσης για όλα τα υποκείμενα μεταξύ διαφορετικών μορφών του ίδιου φαρμάκου, σε τυπικές διασταυρούμενες κλινικές δοκιμές. Απέδειξε ότι υπό κατάλληλη κλίμακα ακολουθεί την  $\chi^2$  κατανομή και p-values μπορούν να υπολογιστούν έτσι ώστε να κατασκευαστούν κριτήρια βιοϊσοδυναμίας. Η προσέγγιση αυτή συγκρίθηκε με τις τρέχουσες τυπικές μεθόδους μέσω εκτεταμένων μελετών προσομοίωσης και πραγματικά δεδομένα.

Οι Martin and Durrani (2007) παρουσίασαν ένα νέο μέτρο απόκλισης για την καταχώριση (registration) ιατρικών απεικονίσεων, το οποίο αξιολογεί τις ιδιότητες της τροποποιημένης συνάρτησης Bessel δεύτερου είδους. Ανέλυσαν τις ιδιότητες του νέου μέτρου και τις συνέκριναν με αυτές των κλασικών μέτρων πληροφορίας και απόκλισης. Επιπλέον, προκειμένου να επιβεβαιώσουν την αποτελεσματικότητά του καθώς και την ευρεία εφαρμοσιμότητά του σε οποιοδήποτε τύπου δεδομένα, η απόδοση - συμπεριφορά του νέου μέτρου αναλύθηκε για διάφορες γνωστές συναρτήσεις κατανομής πιθανότητας. Τα αποτελέσματα επιβεβαίωσαν την ανθεκτικότητά του. Το νέο μέτρο χρησιμοποιήθηκε για την καταχώριση των CT σε MR ιατρικών απεικονίσεων για να τεκμηριώσει τη βελτίωση στην ακρίβεια της καταχώρισης.

## 2.4 Θεωρία Πληροφοριών και Αναλογιστική Επιστήμη

Η Αναλογιστική επιστήμη διακρίνεται σε δυο κυρίως κλάδους: τις ασφαλίσσεις που αφορούν τη ζωή ενός ατόμου (Life Insurance) και τις ασφαλίσσεις που δεν αφορούν τη ζωή του ατόμου (Non-Life Insurance). Βασικός σκοπός του πρώτου κλάδου είναι η εκτίμηση των πιθανοτήτων θνησιμότητας, επιβίωσης, αναπηρίας κλπ με βάση τις οποίες υπολογίζονται τα ασφάλιστρα που υποχρεούται να καταβάλλει ο κάθε ασφαλισμένος. Επίσης ασχολείται με την εύρεση του καταλληλότερου προτύπου θνησιμότητας. Η έννοια της λογοκρισίας υπεισέρχεται άμεσα στην αναλογιστική επιστήμη αφού τα δεδομένα που χρησιμοποιεί για την κατασκευή πινάκων θνησιμότητας είναι λογοκριμένα. Συνεπώς, υπό αυτή την έννοια στην αναλογιστική επιστήμη χρησιμοποιείται και η Θεωρία Πληροφοριών. Η πιο ενδιαφέρουσα όμως εφαρμογή της Θεωρίας Πληροφοριών στον αναλογισμό έγκειται στο γεγονός ότι δίνει εναλλακτικές προσεγγίσεις επίλυσης σε ήδη υπάρχοντα προβλήματα για την επίλυση των οποίων χρησιμοποιούνται κλασικές μέθοδοι. Πιο συγκεκριμένα, τα μέτρα πληροφορίας χρησιμοποιούνται για την επιλογή του καλύτερου - από πλευράς πληροφορίας - μοντέλου για ένα στοχαστικό φαινόμενο. Επίσης, στην προσπάθεια από μια τυπική κατανομή να προκύψει μια άλλη που να είναι πιο κοντά στα ιδιαίτερα χαρακτηριστικά ενός πελάτη. Επίσης η Θεωρία Πληροφοριών χρησιμοποιείται στην προσαρμογή πινάκων θνησιμότητας, έτσι ώστε αυτοί να συμφωνούν με τις ιδιαιτερότητες των πελατών αλλά ταυτόχρονα να είναι όσο το

δυνατόν γίνεται όμοιοι με τους τυπικούς πίνακες. Ακόμη η Θεωρία Πληροφοριών εφαρμόζεται για να επιτευχθεί εξομάλυνση (graduation) στα παρατηρούμενα αναλογιστικά δεδομένα για την εξαγωγή καλύτερων εκτιμητών. Με το θέμα αυτό θα ασχοληθούμε με περισσότερη λεπτομέρεια σε επόμενο κεφάλαιο.

Ο Brockett (1991) χρησιμοποιεί τις έννοιες της θεωρίας πληροφοριών στην προσπάθειά του να παράγει από μια τυπική κατανομή μια άλλη, η οποία θα είναι πιο κοντά στα ιδιαίτερα χαρακτηριστικά ενός μεμονωμένου ασφαλισμένου. Επίσης χρησιμοποιείται στην προσαρμογή πινάκων θνησιμότητας, έτσι ώστε αυτοί να αναπαριστούν γνωστή πληροφορία για τους ασφαλισμένους αλλά ταυτόχρονα να είναι όσο το δυνατόν γίνεται πιο όμοιοι με τους τυπικούς πίνακες (Brockett and Cox, 1984). Μια επιπλέον εφαρμογή της Στατιστικής Θεωρίας Πληροφοριών είναι στη χρήση δευτερευόντων δεδομένων. Συγκεκριμένα, οι Brockett et al. (1995) παρουσιάζουν τη χρήση των μέτρων πληροφορίας για την εκτίμηση πιθανοτήτων και άλλων στατιστικών συναρτήσεων βάσει πινακοποιημένων δεδομένων και για τη συγκέντρωση πληροφορίας από περισσότερες από μια πηγές καθώς και από διαγράμματα.

Τέλος, μέσω της Στατιστικής Θεωρίας Πληροφοριών μπορεί να επιτευχθεί εξομάλυνση (graduation) στα ασφαλιστικά δεδομένα (ποσοστά θνησιμότητας, συνταξιοδοτικά ποσά κ.α.). Στη διατριβή αυτή θα παρουσιάσουμε τη χρήση της στατιστικής θεωρίας πληροφοριών στην εξομάλυνση ποσοστών θνησιμότητας. Συγκεκριμένα θα εξετάσουμε κατά πόσο τα μέτρα πληροφορίας μπορούν να χρησιμοποιηθούν στην εξομάλυνση των ποσοστών θνησιμότητας.

Η εντροπία, ως μέτρο αβεβαιότητας και πληροφορίας, είναι χρήσιμη για τη μελέτη και την αξιολόγηση αναλογιστικών μοντέλων. Μια μέθοδος εκτίμησης, η οποία χρησιμοποιείται για την κατασκευή μοντέλων είναι η μέθοδος της μέγιστης εντροπίας (ME). Σύμφωνα με τη μέθοδο αυτή, ξεκινώντας από κάποιες ροπές, οι οποίες αποτελούν τη μόνη γνωστή πληροφορία για το μοντέλο, ως το καταλληλότερο μοντέλο επιλέγεται εκείνο που έχει τη μεγαλύτερη εντροπία. Η μέθοδος αυτή, χρησιμοποιείται ευρύτατα σε διάφορες επιστήμες, όπως τα οικονομικά, η λογιστική, η βιολογία, η ιατρική, η οικολογία, ο αναλογισμός κλπ (Kapur, 1989).

Οι Berliner and Lev (1978) κατέληξαν στις κατανομές Poisson και Pareto για ασφαλιστικά προβλήματα χρησιμοποιώντας τη μέθοδο της μέγιστης εντροπίας και έδειξαν με εφαρμογή του θεωρήματος του Bayes ότι η ME μέθοδος οδηγεί επίσης στις εκτιμήσεις των παραμέτρων της συνάρτησης πυκνότητας πιθανότητας. Συμπέρασμά τους ήταν ότι η αρχή της ME έχει μεγάλη πιθανότητα εφαρμογής στον αναλογισμό και αποτελεί ένα απέραντο και ενδιαφέρον πλαίσιο για τους αναλογιστές. Πολλές από τις διαδικασίες που έχουν κατασκευαστεί από τους αναλογιστές είτε με απλό είτε με σύνθετο εμπειρικό τρόπο για την αντιμετώπιση ειδικών προβλημάτων μπορούν να έχουν μια απλή ερμηνεία μέσω της έννοιας της ME.

Ο Haberman, στο σχολιασμό που περιλαμβάνεται στο άρθρο του Moore (1980), προτείνει τη χρήση της αρχής της ME ως μεθόδου επιλογής μιας κατανομής σε ένα ευρύ φάσμα περιπτώσεων όπου η κανονική κατανομή χρησιμοποιείται από τον Moore. Άλλες πιθανές

περιοχές του αναλογισμού που μπορεί να χρησιμοποιηθεί η μέθοδος ME, είναι σύμφωνα με τον Haberman, η εκτίμηση των μεγαλύτερων ή μικρότερων μελών ενός συνόλου (δηλαδή ακραίες τιμές (extreme values)), οι οποίες μπορεί να έχουν εφαρμογή σε πολλά και διαφορετικά πεδία όπως η εκτίμηση μεγάλων ασφαλιστικών απαιτήσεων και η πρόβλεψη κορυφών (peaks) (ή βάσεων (troughs)) για τον μακροπρόθεσμο ρυθμό τόκων (long-term interest rate) στο χρηματιστήριο.

Μια μέθοδος εξομάλυνσης ποσοστών θνησιμότητας χρησιμοποιώντας ιδέες της στατιστικής θεωρίας πληροφοριών προτάθηκε από τους Brockett and Zhang (1986). Πιο συγκεκριμένα, οι Zhang and Brockett (1987) προσπάθησαν να κατασκευάσουν μια ομαλή σειρά  $n$  πιθανοτήτων θανάτου  $\{u_x\}$ ,  $x = 1, 2, \dots, n$ , οι οποίες είναι όσο το δυνατόν πιο κοντά σε μια παρατηρούμενη σειρά  $\{u_x\}$ , υπό τρεις μαθηματικούς και δυο αναλογιστικούς περιορισμούς.

Ο Kapur (1989) παρουσιάζει ορισμένα παραδείγματα εφαρμογής της μεθόδου μέγιστης εντροπίας στην αναλογιστική επιστήμη. Πιο συγκεκριμένα, εφαρμόζει τη μέθοδο ME για την εύρεση της κατανομής πιθανότητας του αριθμού των απαιτήσεων μιας ασφαλιστικής εταιρείας σε ένα χρονικό διάστημα, για την εύρεση της κατανομής καταστροφικών γεγονότων κλπ.

Ο Brockett (1991) δίνει μια πολύ καλή περιγραφή της χρήσης της Στατιστικής Θεωρίας Πληροφοριών στην αναλογιστική επιστήμη. Περιγράφει τη χρήση της απόκλισης των Kullback-Leibler για την επιλογή μοντέλων και πώς η Στατιστική Θεωρία Πληροφοριών ενοποιεί και επεκτείνει συγκεκριμένες μεπυζιανές μεθόδους που χρησιμοποιούνται στην αναλογιστική επιστήμη. Επίσης περιγράφει το μοντέλο loglinear, και την ειδική περίπτωση του μοντέλου logit, τα οποία έχουν εφαρμογές σε διάφορες περιοχές του αναλογισμού ως συνέπεια της μοντελοποίησης μέσω Στατιστικής Θεωρίας Πληροφοριών. Επίσης περιγράφει τη χρήση της Στατιστικής Θεωρίας Πληροφοριών για τον προσδιορισμό της κατανομής απώλειας των πελατών και την προσαρμογή πινάκων θνησιμότητας.

Οι Xu et al. (1998) υπολόγισαν άνω και κάτω φράγματα για το stop-loss premium, δηλαδή την αναμενόμενη πληρωμή από τον αντασφαλιστή, όταν η κατανομή της απαίτησης είναι άγνωστη αλλά υποτίθεται ότι είναι όσο το δυνατόν πιο κοντά στην εμπειρική κατανομή παλαιότερων απαιτήσεων. Η «απόσταση» από την παρατηρούμενη εμπειρική κατανομή μετρείται μέσω της I-divergence, δηλαδή του αριθμού πληροφορίας των Kullback-Leibler. Στη συνέχεια, η απόσταση αυτή χρησιμοποιείται για τον προσδιορισμό των φραγμάτων.

Μέτρα πληροφορίας χρησιμοποιήθηκαν για την εύρεση της εκ των προτέρων κατανομής για την παράμετρο διασποράς  $\lambda$  των μοντέλων εκθετικής διασποράς (exponential dispersion model). Οι Landsman and Makov (1998) χρησιμοποίησαν την αρχή της μέγιστης εντροπίας με περιορισμό στην αναμενόμενη τιμή του  $\lambda$  και οι Landsman and Makov (1999) ελαχιστοποίησαν το μέτρο πληροφορίας του Fisher με περιορισμούς στη μέση τιμή και τη διακύμανση του  $\lambda$  για να βρουν το βέλτιστο συντελεστή αξιοπιστίας και φράγματα αυτού. Το δεύτερο κριτήριο χρησιμοποιήθηκε για την εύρεση της εκ των προτέρων κατανομής για το  $\lambda$  σε συνδυασμό με τη γνώση της πιθανότητας ότι μια απαίτηση ξεπερνά ένα συγκεκριμένο όριο

(threshold), εξασφαλίζοντας έτσι πληροφορία για τη συμπεριφορά της ουράς στην επιρροή στο ασφάλιστρο. Οι Promislow and Young (2000) υπολόγισαν δίκαια ασφάλιστρα αξιοπιστίας (equitable credibility premiums) χρησιμοποιώντας ως μέτρο της σχετικής διαφοράς μεταξύ των χρεωμένων ασφαλιστρών και των πραγματικών ασφαλιστρών μια συνάρτηση απώλειας τύπου εντροπίας (entropy loss function) αντί του συνηθισμένου τετραγωνικού σφάλματος.

Η σχετική εντροπία (relative entropy) χρησιμοποιείται συχνά για την τιμολόγηση οικονομικών κεφαλαίων μεγάλου κινδύνου (risky financial assets) σε ελλιπείς (incomplete) αγορές, ενώ η distortion χρησιμοποιείται ευρύτατα για την τιμολόγηση ασφαλιστικών κινδύνων και στη διοίκηση κινδύνου (Reesor and McLeish, 2002).

Η αρχή της ME, όταν εφαρμόζεται στον χρεωστικό κίνδυνο (credit risk), οδηγεί σε μοντέλα που περιέχουν τις ελάχιστες συναφείς υποθέσεις με την υπάρχουσα πληροφορία (Brunel, 2004). Τα μοντέλα αυτά, ονομάζονται ελάχιστα μοντέλα (minimal models) και αποτελούν ένα σημείο αναφοράς όταν κατασκευάζουμε ένα νέο μοντέλο. Εφάρμοσε την προσέγγιση αυτή στην επιλογή της κατανομής απωλειών ενός χρεωστικού χαρτοφυλακίου, των αξιογράφων (asset backed securities) και της κατανομής των ρυθμών ανάκαμψης (recovery rates), και έδειξε πώς μπορούμε να τη χρησιμοποιήσουμε για την επιβεβαίωση των υποθέσεων οποιουδήποτε μοντέλου σε αυτές τις περιοχές.

Ο Darooneh (2004) χρησιμοποιεί τη ME μέθοδο για την τιμολόγηση των ασφαλίσεων μη ζωής (non-life insurance). Πιο συγκεκριμένα, εφαρμόζει τη ME μέθοδο με σκοπό να υπολογίσει την πυκνότητα της τιμής (η ασφαλιστική τιμή ορίζεται με αναφορά σε μια συνάρτηση πιθανότητας), θεωρώντας ότι ο μέσος όρος του πλούτου της αγοράς (market's wealth) είναι σταθερός. Οι Luthi and Doege (2005) ασχολούνται με την οικογένεια μέτρων κινδύνου που βασίζονται στην εντροπία (entropy based risk measures).

Τέλος, οι Sachlas and Papaioannou (2008, 2009a) παρουσιάζουν τη χρήση των αποκλίσεων των Cressie and Read στο πρόβλημα της εξομάλυνσης πινάκων θνησιμότητας και προτείνουν την ελαχιστοποίηση των αποκλίσεων υπό έναν επιπλέον περιορισμό από αυτούς που πρότειναν οι Zhang and Brockett (1987). Οι ιδέες αυτές αποτελούν το μεγαλύτερο κομμάτι της παρούσας διατριβής.

## 2.5 Πρόσφατη έρευνα στο πεδίο της Στατιστικής Θεωρίας Πληροφοριών

Η Στατιστική Θεωρία Πληροφοριών είναι ένας τομέας ανοιχτός σε έρευνα. Αρκετοί συγγραφείς έχουν ασχοληθεί και συνεχίζουν να ασχολούνται με το θέμα. Μεταξύ αυτών είναι οι Nadarajah and Zografos (2003 και 2005), Zografos and Nadarajah (2005a, b), Karagrigoriou and Papaioannou (2006), Pardo (2006) και Mattheou (2007). Κυρίως ασχολούνται

με την εύρεση της έκφρασης υπολογισμού μέτρων πληροφορίας για διάφορες γνωστές και μη κατανομές. Επίσης νέα μέτρα πληροφορίας είτε απόκλισης είτε εντροπίας έχουν προταθεί καθώς επίσης και πεδία εφαρμογής τους.

Οι Ebrahimi et al. (2007) κατασκεύασαν μέτρα πληροφορίας για πολυμεταβλητές κατανομές όταν το στήριγμά (support) τους είναι σταδιακά αποκομμένο (progressively censored). Ιδιαίτερη βάση δίνουν στις από κοινού, περιθώριες και υπο συνθήκη εντροπίες και την από κοινού πληροφορία για τις κατανομές υπολειπόμενου χρόνου ζωής, των οποίων το στήριγμα είναι αποκομμένο στις τρέχουσες ηλικίες των στοιχείων του συστήματος. Οι τρέχουσες ηλικίες των στοιχείων εισάγουν (induce) μια από κοινού δυναμική στα μέτρα πληροφορίας του υπολειπόμενου χρόνου ζωής. Παρουσιάζουν εκφράσεις της εντροπίας για διάφορα διμεταβλητά και πολυμεταβλητά μοντέλα χρόνου ζωής.

Οι Matheou et al. (2008) μελέτησαν μέτρα απόκλισης και κριτήρια επιλογής μοντέλων. Επιπλέον πρότειναν ένα νέο κριτήριο επιλογής μοντέλου, το οποίο αποκαλούν divergence information criterion (DIC) και βασίζεται στο μέτρο BHHJ. Μέσω διάφορων μελετών προσομοίωσης που διενήργησαν, έδειξαν ότι το κριτήριο αυτό έχει αρκετά ικανοποιητική συμπεριφορά συγκρινόμενη με άλλα γνωστά κριτήρια. Προς αποφυγή παρεξηγήσεων, θα πρέπει να αναφέρουμε ότι στη βιβλιογραφία με τη σύντμηση DIC αναφέρεται και το μέτρο deviance information criterion, το οποίο είναι μια γενίκευση ιεραρχικής μοντελοποίησης των κριτηρίων AIC (Akaike information criterion) και BIC (Bayesian information criterion).

Ο Zografos (2008a) υπολόγισε και έκανε μια συγκριτική μελέτη δυο γενικών μέτρων πολυμεταβλητής κύρτωσης, το μέτρο  $\beta_{2,p}$  του Mardia και το μέτρο  $S(f)$  του Song. Τα μέτρα αυτά στηρίζονται σε μέτρα πληροφορίας όπως είναι η εντροπία του Rényi.

Ο Zografos (2008b) πρότεινε νέα περιγραφικά μέτρα που στηρίζονται στην εντροπία του Shannon για την περίπτωση των μεικτών, συνεχών και διακριτών πολυμεταβλητών δεδομένων.

## ΚΕΦΑΛΑΙΟ 3

# ΜΕΤΡΑ ΑΠΟΚΛΙΣΗΣ ΜΕ ΜΗ ΠΙΘΑΝΟΤΙΚΑ ΜΕΤΡΑ

### 3.1 Εισαγωγή

Υπάρχουν αρκετά πρακτικά προβλήματα στα οποία εμπλέκονται μη πιθανοτικά διανύσματα (nonprobability vectors). Με τον όρο μη πιθανοτικό διάνυσμα εννοούμε το διάνυσμα, του οποίου τα στοιχεία δεν αθροίζουν στη μονάδα. Ένα τέτοιο πρόβλημα που εμφανίζεται σε επιστήμες όπως ο Αναλογισμός, η Βιοστατιστική και η Δημογραφία είναι η εξομάλυνση ποσοτήτων όπως τα ποσοστά θνησιμότητας, ποσοστά γονιμότητας, ποσοστά νοσηρότητας κλπ. Αν και τα μέτρα πληροφορίας ορίζονται για διανύσματα πιθανότητας, στην πράξη, χρησιμοποιούνται εξ ίσου και στην περίπτωση των μη πιθανοτικών διανυσμάτων.

Σημειώνουμε ότι μια διδιάστατη συνάρτηση  $D(f, g)$  δυο πιθανοτικών συναρτήσεων ή διανυσμάτων  $f, g$  είναι μέτρο απόκλισης, εάν ισχύει  $D(f, g) \geq 0$  με ισότητα αν και μόνο αν  $f = g$  (βλέπε Basu et al., 1998). Αυτή είναι η ελάχιστη (minimal) απαίτηση για να θεωρηθεί ένα μέτρο  $D(f, g)$  ως είδος απόστασης μεταξύ των  $f$  και  $g$ . Ο Pardo (2006, σελ. 2) αναφέρει ότι ο συντελεστής με την ιδιότητα να αυξάνει καθώς οι δυο κατανομές που εμπλέκονται σε αυτόν απομακρύνονται η μια από την άλλη θα ονομάζεται μέτρο απόκλισης μεταξύ των δυο κατανομών. Για άλλες απαιτήσεις βλέπε Read and Cressie (1988) και Mathie and Rathie (1975).

Τα μέτρα απόκλισης, μέτρα πληροφορίας και οι ιδιότητές τους, είναι ακόμα ένα θέμα υπό διερεύνηση. Νέα μέτρα απόκλισης προτείνονται και οι ιδιότητές τους μελετώνται στον Mattheou (2007) ενώ οι Papaioannou and Ferentinos (2005) ερευνούν τον πληροφοριακό αριθμό του Fisher υπό το πρίσμα των ιδιοτήτων των κλασικών μέτρων πληροφορίας. Παρόλα αυτά, δεν υπάρχει καθολική συμφωνία μεταξύ των στατιστικών επιστημόνων σχετικά με το ποιές ιδιότητες συνιστούν ή ορίζουν ένα μέτρο στατιστικής πληροφορίας καθώς η προσέγγιση είναι περισσότερο λειτουργική (operational) παρά αξιωματική (Papaioannou and

Ferentinos, 2005).

Ο Παπαϊοαννου (1985, 2001) παρουσιάζει με λεπτομέρεια τις ιδιότητες των μέτρων πληροφορίας και απόκλισης με πιθανοτικά διανύσματα, όπως είδαμε στο προηγούμενο κεφάλαιο. Ο κύριος σκοπός του κεφαλαίου αυτού είναι η διερεύνηση των ιδιοτήτων των μέτρων απόκλισης όταν εμπλέκονται σε αυτά μη πιθανοτικά διανύσματα. Συγκεκριμένα μελετάμε τις εξής ιδιότητες: μη αρνητικότητα, ισχυρή και ασθενή προσθετικότητα, μέγιστη πληροφορία, επάρκεια και ασυμπτωτική ιδιότητα.

### 3.2 Το μέτρο απόκλισης των Kullback - Leibler με μη πιθανοτικά διανύσματα

Το πιο δημοφιλές μέτρο απόκλισης, όπως αναφέραμε στο προηγούμενο κεφάλαιο, είναι το μέτρο πληροφορίας των Kullback - Leibler, το οποίο βασίζεται σε δυο διανύσματα πιθανότητας  $\mathbf{p}^*$  και  $\mathbf{q}^*$ , όπου  $\mathbf{p}^* = (p_1^*, \dots, p_n^*)$  και  $\mathbf{q}^* = (q_1^*, \dots, q_n^*)$  με  $\sum_{i=1}^n p_i^* = \sum_{i=1}^n q_i^* = 1$  και δίνεται από τη σχέση

$$I^{KL}(\mathbf{p}^*, \mathbf{q}^*) = \sum_i p_i^* \ln \frac{p_i^*}{q_i^*}$$

[βλέπε Kullback (1959, σελ. 6 - 7)]. Είναι ένα μέτρο κατευθυνόμενης απόκλισης (directed divergence) με την έννοια ότι δεν ικανοποιεί όλες τις ιδιότητες της απόστασης ή της μετρικής, (στην πραγματικότητα δεν είναι συμμετρικό μέτρο και δεν ικανοποιεί την ιδιότητα της τριγωνικής ανισότητας) και συνεπώς δεν μπορεί να θεωρηθεί ως μια καθαρή απόσταση. Η κατευθυνόμενη απόκλιση των Kullback - Leibler ορίζεται για διανύσματα πιθανότητας και ικανοποιεί τις περισσότερες από τις ιδιότητες που ικανοποιούν τα μέτρα πληροφορίας.

Στόχος της παραγράφου αυτής είναι να εξετάσουμε εάν η κατευθυνόμενη απόκλιση των Kullback - Leibler μεταξύ δυο μη πιθανοτικών διανυσμάτων μπορεί να θεωρηθεί ως μέτρο πληροφορίας. Αυτό γίνεται εξετάζοντας τις ιδιοτήτες του υπό το πρίσμα των γενικών ιδιοτήτων των μέτρων πληροφορίας και απόκλισης. Ας ξεκινήσουμε με τον ορισμό της κατευθυνόμενης απόκλισης των Kullback - Leibler μεταξύ δυο μη πιθανοτικών διανυσμάτων.

**Ορισμός 3.1.** Η κατευθυνόμενη απόκλιση των Kullback - Leibler μεταξύ δυο  $n \times 1$  μη πιθανοτικών διανυσμάτων  $\mathbf{p}$  και  $\mathbf{q}$ , ορίζεται ως

$$D^{KL}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} \quad (3.1)$$

όπου  $\mathbf{p} = (p_1, p_2, \dots, p_n)^T > \mathbf{0}$ ,  $\mathbf{q} = (q_1, q_2, \dots, q_n)^T > \mathbf{0}$  με  $\sum_{i=1}^n p_i \neq 1$  και  $\sum_{i=1}^n q_i \neq 1$ .

Η σχέση που συνδέει την απόκλιση των Kullback - Leibler μεταξύ δυο μη πιθανοτικών διανυσμάτων και την απόκλιση των Kullback - Leibler μεταξύ διανυσμάτων πιθανότητας



δίνεται στο επόμενο λήμμα.

**Λήμμα 3.1.** Για την κατευθυνόμενη απόκλιση των Kullback - Leibler με μη πιθανοτικά διανύσματα, ισχύει ότι

$$D^{KL}(\mathbf{p}, \mathbf{q}) = \left( \sum_{i=1}^n p_i \right) [I^{KL}(\mathbf{p}^*, \mathbf{q}^*) + \ln k],$$

όπου  $k = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i}$ , και  $I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$  είναι το μέτρο των Kullback - Leibler που εμπλέκει πιθανοτικά διανύσματα  $\mathbf{p}^*$  και  $\mathbf{q}^*$ , όπου τα στοιχεία των  $\mathbf{p}^*$  και  $\mathbf{q}^*$  είναι τα τυποποιημένα στοιχεία των  $\mathbf{p}$  και  $\mathbf{q}$ , δηλαδή  $p_i^* = p_i / \sum_{i=1}^n p_i$  και  $q_i^* = q_i / \sum_{i=1}^n q_i$ ,  $i = 1, \dots, n$ .

Απόδειξη. Η Εξίσωση 3.1 γράφεται ως εξής:

$$\begin{aligned} D^{KL}(\mathbf{p}, \mathbf{q}) &= \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} = \sum_{i=1}^n \left( p_i^* \sum_{i=1}^n p_i \right) \ln \frac{p_i^* \sum_{i=1}^n p_i}{q_i^* \sum_{i=1}^n q_i} \\ &= \left( \sum_{i=1}^n p_i \right) \sum_{i=1}^n p_i^* \left( \ln \frac{p_i^*}{q_i^*} + \ln k \right) \\ &= \left( \sum_{i=1}^n p_i \right) \left[ \sum_{i=1}^n p_i^* \ln \frac{p_i^*}{q_i^*} + (\ln k) \sum_{i=1}^n p_i^* \right] \\ &= \left( \sum_{i=1}^n p_i \right) [I^{KL}(\mathbf{p}^*, \mathbf{q}^*) + \ln k], \end{aligned}$$

όπου  $k = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i}$ . ■

**Πρόταση 3.1.** (Μη αρνητικότητα) Ισχύει

$$D^{KL}(\mathbf{p}, \mathbf{q}) \geq 0, \quad (3.2)$$

εάν ισχύει μια από τις παρακάτω συνθήκες:

$$(i) \sum_{i=1}^n p_i \geq \sum_{i=1}^n q_i, \quad (ii) \sum_{i=1}^n p_i < \sum_{i=1}^n q_i \text{ και } \ln k > -I^{KL}(\mathbf{p}^*, \mathbf{q}^*) \text{ όπου } k = \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i}.$$

Η ισότητα στη Σχέση 3.2 ισχύει εάν  $\mathbf{p} = \mathbf{q}$  ή  $\ln k = -I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$ . Επιπλέον εάν  $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i$  τότε  $D^{KL}(\mathbf{p}, \mathbf{q}) \geq 0$  με ισότητα εάν και μόνο εάν  $\mathbf{p} = \mathbf{q}$ .

Απόδειξη. Η απόδειξη είναι προφανής, βασιζόμενη στην ιδιότητα της μη αρνητικότητας του μέτρου  $I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$ . ■

Σημειώνουμε ότι η ισότητα  $D^{KL}(\mathbf{p}, \mathbf{q}) = 0$  δεν υπονοεί απαραίτητα ότι  $\mathbf{p} = \mathbf{q}$  εκτός και αν ισχύει  $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i$ . Συνεπώς η ελάχιστη απαίτηση για τη χρήση του  $D^{KL}(\mathbf{p}, \mathbf{q})$  ως μέτρου απόκλισης είναι η ισότητα  $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i$ . Αριθμητικά αυτό μπορεί να ελεγχθεί χρησιμοποιώντας τα διανύσματα  $\mathbf{p} = (0.1584514, 0.2201928, 0.7247736)^T$  και  $\mathbf{q} = (0.4, 0.8, 0.4)^T$ ,  $\sum_i p_i \neq \sum_i q_i$ , για τα οποία το μέτρο  $D^{KL}(p, q)$  ισούται με  $4.78877 \times 10^{-7}$  δηλαδή είναι περίπου ίσο με το μηδέν, ενώ η Ευκλείδεια απόσταση  $\|\mathbf{p} - \mathbf{q}\|$  ισούται με 0.707107.

Ο Ορισμός 3.1 έχει προφανείς επεκτάσεις στη διμεταβλητή και πολυμεταβλητή περίπτωση. Στη συνέχεια παρουσιάζουμε τους σχετικούς ορισμούς για τη διμεταβλητή περίπτωση. Οι ορισμοί αυτοί απαιτούνται για την απόδειξη των ιδιοτήτων του μέτρου  $D^{KL}(\mathbf{p}, \mathbf{q})$ .

**Ορισμός 3.2.** (Διμεταβλητή απόκλιση) Έστω  $p_i(x, y)$ ,  $i = 1, 2$ , δυο διμεταβλητές (μη πιθανοτικές συναρτήσεις) που σχετίζονται με δυο διακριτές μεταβλητές  $(X, Y)$  στο  $R^2$  για τις οποίες ισχύει  $\sum_x \sum_y p_i(x, y) \neq 1$ . Ορίζουμε την κατευθυνόμενη απόκλιση των Kullback - Leibler μεταξύ δυο διμεταβλητών μη πιθανοτικών συναρτήσεων  $p_1, p_2$  ως

$$D_{X,Y}^{KL}(p_1, p_2) = \sum_x \sum_y p_1(x, y) \ln \frac{p_1(x, y)}{p_2(x, y)}.$$

**Ορισμός 3.3.** (Υπό συνθήκη απόκλιση) Για τις διακριτές μεταβλητές  $X, Y$  και τις διμεταβλητές μη πιθανοτικές συναρτήσεις  $p_i(x, y)$ ,  $i = 1, 2$ , όπως αυτές δόθηκαν παραπάνω, έστω  $f_i(x) = \sum_y p_i(x, y)$ ,  $h_i(y|x) = \frac{p_i(x, y)}{f_i(x)}$ ,  $g_i(y) = \sum_x p_i(x, y)$ , και  $r_i(x|y) = \frac{p_i(x, y)}{g_i(y)}$ ,  $i = 1, 2$ . Θέτουμε

$$D_{Y|X=x}^{KL}(h_1, h_2) = \sum_y h_1(y|x) \ln \frac{h_1(y|x)}{h_2(y|x)}, D_{X|Y=y}^{KL}(r_1, r_2) = \sum_x r_1(x|y) \ln \frac{r_1(x|y)}{r_2(x|y)}$$

και ορίζουμε τις

$$D_{Y|X}^{KL}(h_1, h_2) = E_X [D_{Y|X=x}^{KL}(h_1, h_2)] = \sum_x f_1(x) \sum_y h_1(y|x) \ln \frac{h_1(y|x)}{h_2(y|x)},$$

$$D_{X|Y}^{KL}(r_1, r_2) = E_Y [D_{X|Y=y}^{KL}(r_1, r_2)] = \sum_y g_1(y) \sum_x r_1(x|y) \ln \frac{r_1(x|y)}{r_2(x|y)}.$$

**Πρόταση 3.2.** (Ισχυρή προσθετικότητα) Έστω  $p_1, p_2$  δυο διμεταβλητές μη πιθανοτικές συναρτήσεις που σχετίζονται με δυο διακριτές μεταβλητές  $(X, Y)$  στο  $R^2$  όπως στον Ορισμό 3.2. Τότε

$$D_{X,Y}^{KL}(p_1, p_2) = D_X^{KL}(f_1, f_2) + D_{Y|X}^{KL}(h_1, h_2) = D_Y^{KL}(g_1, g_2) + D_{X|Y}^{KL}(r_1, r_2),$$

όπου οι συναρτήσεις  $f_i, h_i, g_i, r_i$ ,  $i = 1, 2$  είναι όπως στον Ορισμό 3.3.

Απόδειξη. Για τις μεταβλητές  $X, Y$  έχουμε ότι

$$\begin{aligned}
& D_X^{KL}(f_1, f_2) + D_{Y|X}^{KL}(h_1, h_2) \\
&= \sum_x f_1(x) \ln \frac{f_1(x)}{f_2(x)} + \sum_x f_1(x) \sum_y h_1(y|x) \ln \frac{h_1(y|x)}{h_2(y|x)} \\
&= \sum_x f_1(x) \ln \frac{f_1(x)}{f_2(x)} + \sum_x f_1(x) \sum_y \frac{p_1(x, y)}{f_1(x)} \ln \frac{p_1(x, y) f_2(x)}{f_1(x) p_2(x, y)} \\
&= \sum_x f_1(x) \ln \frac{f_1(x)}{f_2(x)} + \sum_x \sum_y p_1(x, y) \left\{ \ln \frac{p_1(x, y)}{p_2(x, y)} + \ln \frac{f_2(x)}{f_1(x)} \right\} \\
&= \sum_x f_1(x) \ln \frac{f_1(x)}{f_2(x)} + \sum_x \sum_y p_1(x, y) \ln \frac{p_1(x, y)}{p_2(x, y)} + \sum_x f_1(x) \ln \frac{f_2(x)}{f_1(x)} \\
&= D_{X,Y}^{KL}(p_1, p_2).
\end{aligned}$$

Με ανάλογο τρόπο, αποδεικνύουμε ότι

$$D_{X,Y}^{KL}(p_1, p_2) = D_Y^{KL}(g_1, g_2) + D_{X|Y}^{KL}(r_1, r_2).$$

■

**Πόρισμα 3.1.** (i)  $D_{X,Y}^{KL}(p_1, p_2) \geq D_X^{KL}(f_1, f_2)$  με ισότητα εάν και μόνο εάν  $D_{Y|X}^{KL}(h_1, h_2) = 0$ .

(ii)  $D_{X,Y}^{KL}(p_1, p_2) \geq D_Y^{KL}(g_1, g_2)$  με ισότητα εάν και μόνο εάν  $D_{X|Y}^{KL}(r_1, r_2) = 0$ .

(iii)  $D_{X,Y}^{KL}(p_1, p_2) \geq D_{Y|X}^{KL}(h_1, h_2)$  με ισότητα εάν και μόνο εάν  $D_X^{KL}(f_1, f_2) = 0$ .

(iv)  $D_{X,Y}^{KL}(p_1, p_2) \geq D_{X|Y}^{KL}(r_1, r_2)$  με ισότητα εάν και μόνο εάν  $D_Y^{KL}(g_1, g_2) = 0$ .

Σε όλες τις παραπάνω περιπτώσεις, η ισότητα ισχύει εάν και μόνο εάν οι κανονικοποιημένες τιμές των  $X, Y$  είναι ανεξάρτητες.

Οι κανονικοποιημένες τιμές των  $X, Y$  δημιουργούν δυο τυχαίες μεταβλητές  $X^*, Y^*$  με διακριτή από κοινού συνάρτηση μάζας πιθανότητας  $p_i^*(x, y) = p_i(x, y) / \sum_x \sum_y p_i(x, y)$  και περιθώριες και υπό συνθήκη συνάρτηση μάζας πιθανότητας ως εξής:  $X^* \sim f_i^*, Y^*|X^* \sim h_i^*, Y^* \sim g_i^*, X^*|Y^* \sim r_i^*$ . Για τις τυχαίες μεταβλητές  $X^*, Y^*$  έχουμε ότι

$$I_{X^*, Y^*}^{KL}(p_1^*, p_2^*) = \sum_x \sum_y p_1^*(x, y) \ln \frac{p_1^*(x, y)}{p_2^*(x, y)}.$$

**Πρόταση 3.3.** (Ασθενής προσθετικότητα) Εάν  $h_i(y|x) = g_i(y)$  και συνεπώς  $p_i(x, y) = f_i(x)g_i(y)$ ,  $i = 1, 2$ , έχουμε ότι οι τυχαίες μεταβλητές  $X^*, Y^*$ , που δημιουργούνται κανονικοποιώντας τις  $X, Y$  όπως αναφέρθηκε πριν, είναι ανεξάρτητες, και ισχύει ότι

$$D_{X,Y}^{KL}(p_1, p_2) = D_X^{KL}(f_1, f_2) + D_Y^{KL}(g_1, g_2) - \xi \ln \eta,$$

όπου  $\xi = \sum_y g_1(y) = \sum_x f_1(x)$  και  $\eta = \sum_y g_1(y) / \sum_y g_2(y) = \sum_x f_1(x) / \sum_x f_2(x)$ .

Απόδειξη. Από τους παραπάνω ορισμούς, είναι εύκολο να δει κανείς ότι

$$p_i^*(x, y) = f_i^*(x)g_i^*(y),$$

το οποίο σημαίνει ότι οι τυχαίες μεταβλητές  $X^*, Y^*$  είναι ανεξάρτητες. Γνωρίζουμε ότι

$$I_{X^*, Y^*}^{KL}(p_1^*, p_2^*) = I_{X^*}^{KL}(f_1^*, f_2^*) + I_{Y^*}^{KL}(g_1^*, g_2^*).$$

Έτσι για τις μεταβλητές  $X, Y$  έχουμε ότι

$$\begin{aligned} D_{X, Y}^{KL}(p_1, p_2) &= \sum_x \sum_y p_1(x, y) \ln \frac{p_1(x, y)}{p_2(x, y)} \\ &= \left( \sum_x \sum_y p_1(x, y) \right) [I_{X^*, Y^*}^{KL}(p_1^*, p_2^*) + \ln s] \\ &= \left( \sum_x \sum_y p_1(x, y) \right) [I_{X^*}^{KL}(f_1^*, f_2^*) + I_{Y^*}^{KL}(g_1^*, g_2^*) + \ln s], \end{aligned} \quad (3.3)$$

όπου  $s = \sum_x \sum_y p_1(x, y) / \sum_x \sum_y p_2(x, y)$ . Επειδή  $\sum_x f_i(x) = \sum_y g_i(y) = \sum_x \sum_y p_i(x, y)$ ,  $i = 1, 2$  έχουμε ότι  $\xi = \sum_x \sum_y p_1(x, y)$  και  $s = \eta$ . Συνεπώς από τη Σχέση 3.3 έχουμε ότι

$$\begin{aligned} D_{X, Y}^{KL}(p_1, p_2) &= \xi I_{X^*}^{KL}(f_1^*, f_2^*) + \xi I_{Y^*}^{KL}(g_1^*, g_2^*) + \xi \ln \eta \\ &= D_X^{KL}(f_1, f_2) - \xi \ln \eta + D_Y^{KL}(g_1, g_2) - \xi \ln \eta + \xi \ln \eta \\ &= D_X^{KL}(f_1, f_2) + D_Y^{KL}(g_1, g_2) - \xi \ln \eta. \end{aligned}$$

■

Είναι τώρα εύκολο να διαπιστώσει κανείς ότι η ασθενής προσθετικότητα ισχύει εάν  $\sum_x f_1(x) = \sum_x f_2(x)$  ή  $\sum_y g_1(y) = \sum_y g_2(y)$ .

**Πρόταση 3.4.** (Μέγιστη πληροφορία και επάρκεια) Έστω  $Y = T(X)$  ένας μετρήσιμος μετασχηματισμός της  $X$  και  $p_i = p_i(x)$ ,  $g_i = g_i(y)$ ,  $i = 1, 2$ . Τότε ισχύει

$$D_X^{KL}(p_1, p_2) \geq D_Y^{KL}(g_1, g_2),$$

με την ισότητα εάν και μόνο εάν η  $Y$  είναι επαρκής ως προς το ζεύγος κατανομών  $p_1^*$  και  $p_2^*$ . Οι  $Y^*$  και  $X^*$  είναι οι κανονικοποιημένες  $Y$  και  $X$  κατανομές, αντίστοιχα.

Απόδειξη. Έστω  $g_i(y)$  είναι ένα μέτρο που σχετίζεται με την  $Y$ . Τότε  $g_i(y) = \sum_{x:T(x)=y} p_i(x)$ . Θέτοντας  $a = \sum_x p_1(x)$  και  $b = \sum_y g_1(y)$  και βάσει του Λήμματος 3.1, οι ακόλουθες ανισότητες είναι ισοδύναμες

$$\begin{aligned} D_X^{KL}(p_1, p_2) &\geq D_Y^{KL}(g_1, g_2) \Leftrightarrow \\ \sum_x p_1(x) \ln \frac{p_1(x)}{p_2(x)} &\geq \sum_y g_1(y) \ln \frac{g_1(y)}{g_2(y)} \Leftrightarrow \\ a \left[ \sum_x p_1^*(x) \ln \frac{p_1^*(x)}{p_2^*(x)} + \ln c \right] &\geq b \left[ \sum_y g_1^*(y) \ln \frac{g_1^*(y)}{g_2^*(y)} + \ln d \right] \Leftrightarrow \\ a [I_{X^*}^{KL}(p_1^*, p_2^*) + \ln c] &\geq b [I_{Y^*}^{KL}(g_1^*, g_2^*) + \ln d], \end{aligned}$$

όπου  $p_i^*(x) = p_i(x)/\sum_x p_i(x)$ ,  $g_i^*(y) = g_i(y)/\sum_y g_i(y)$ ,  $c = \sum_x p_1(x)/\sum_x p_2(x)$  και  $d = \sum_y g_1(y)/\sum_y g_2(y)$ . Οι  $X^*$ ,  $Y^*$  είναι οι δυο τυχαίες μεταβλητές που προκύπτουν κανονικοποιώντας τις τιμές των  $X$ ,  $Y$ . Επειδή  $\sum_x p_i(x) = \sum_y g_i(y)$ ,  $i = 1, 2$  και συνεπώς  $a = b$  και  $c = d$ , η τελευταία ανισότητα ισοδυναμεί με

$$\begin{aligned} a I_{X^*}^{KL}(p_1^*, p_2^*) + a \ln c &\geq b I_{Y^*}^{KL}(g_1^*, g_2^*) + b \ln d \Leftrightarrow \\ b I_{X^*}^{KL}(p_1^*, p_2^*) &\geq b I_{Y^*}^{KL}(g_1^*, g_2^*) \Leftrightarrow \\ I_{X^*}^{KL}(p_1^*, p_2^*) &\geq I_{Y^*}^{KL}(g_1^*, g_2^*) \end{aligned}$$

το οποίο ισχύει αφού οι  $X^*$  και  $Y^*$  είναι τυχαίες μεταβλητές και η στατιστική συνάρτηση  $Y^* = T(X^*)$  είναι επαρκής [βλέπε Pardo (2006, σελ. 11 - 12); Papaioannou (1985); Kullback (1959, σελ. 21)]. ■

**Πρόταση 3.5.** Ισχύει  $D^{KL}(\mathbf{p}, \mathbf{q}) \geq I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$  όταν μια από τις παρακάτω συνθήκες ισχύει: (i)  $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i \geq 1$ , (ii)  $\sum_{i=1}^n p_i > \sum_{i=1}^n q_i$  και  $\sum_{i=1}^n p_i \geq 1$ , (iii)  $\sum_{i=1}^n p_i < \sum_{i=1}^n q_i$  και  $\sum_{i=1}^n p_i < 1$ .

Απόδειξη. Ακολουθεί εύκολα από το Λήμμα 3.1 μέσω απλών αλγεβρικών πράξεων. ■

Μια βασική ιδιότητα των μέτρων πληροφορίας και απόκλισης είναι η ασυμπτωτική ιδιότητα. Η ιδιότητα αυτή σημαίνει ότι οι ακολουθίες των τυχαίων μεταβλητών συγκλίνουν όταν  $n \rightarrow \infty$  εάν και μόνο εάν  $I_{X_n} \rightarrow I_X$ , όπου το  $I$  συμβολίζει το μέτρο πληροφορίας. Υπό ορισμένες συνθήκες (Kullback, 1959) η ασυμπτωτική ιδιότητα ισχύει για την απόκλιση των Kullback - Leibler. Βλέπε επίσης Zografos et al. (1989).

Στην επόμενη πρόταση διερευνούμε εάν η ασυμπτωτική ιδιότητα ισχύει στην περίπτωση της απόκλισης των Kullback - Leibler με μη πιθανοτικά διανύσματα.

**Πρόταση 3.6.** (Ασυμπτωτική ιδιότητα) Έστω  $\{\mathbf{p}_n\}$  είναι μια φραγμένη από επάνω ακολουθία μη πιθανοτικών διανυσμάτων. Τότε  $\mathbf{p}_n \rightarrow \mathbf{p}$  εάν και μόνο εάν  $D^{KL}(\mathbf{p}_n, \mathbf{p}) \rightarrow 0$ , δηλαδή η ασυμπτωτική ιδιότητα ισχύει για την απόκλιση των Kullback - Leibler με μη πιθανοτικά διανύσματα.

Απόδειξη. Γνωρίζουμε ότι  $D^{KL}(\mathbf{p}_n, \mathbf{p}) = \left( \sum_i p_n(i) \right) \left[ I^{KL}(\mathbf{p}_n^*, \mathbf{p}^*) + \ln \frac{\sum_i p_n(i)}{\sum_i p(i)} \right]$ . Έστω  $\mathbf{p}_n \rightarrow \mathbf{p}$ . Στη συνέχεια έχουμε ότι

$$\begin{aligned} \lim_{n \rightarrow \infty} D^{KL}(\mathbf{p}_n, \mathbf{p}) &= \lim_{n \rightarrow \infty} \left\{ \left( \sum_i p_n(i) \right) \left[ I^{KL}(\mathbf{p}_n^*, \mathbf{p}^*) + \ln \frac{\sum_i p_n(i)}{\sum_i p(i)} \right] \right\} \\ &= \lim_{n \rightarrow \infty} \left( \sum_i p_n(i) \right) \lim_{n \rightarrow \infty} \left[ I^{KL}(\mathbf{p}_n^*, \mathbf{p}^*) + \ln \frac{\sum_i p_n(i)}{\sum_i p(i)} \right] \\ &= 0, \end{aligned}$$

επειδή  $\lim_{n \rightarrow \infty} I^{KL}(\mathbf{p}_n^*, \mathbf{p}^*) = 0$  και  $\lim_{n \rightarrow \infty} \left( \ln \frac{\sum_i p_n(i)}{\sum_i p(i)} \right) = 0$ .

Από την άλλη, έστω  $D^{KL}(\mathbf{p}_n, \mathbf{p}) \rightarrow 0$ . Τότε από τους Zografos et al. (1989) έχουμε για την  $\phi$ -απόκλιση ότι

$$\lim_{n \rightarrow \infty} \sum_i p(i) \phi \left( \frac{p_n(i)}{p(i)} \right) = 0.$$

όπου  $\phi(x)$ ,  $x > 0$  είναι μια σταθερή συνάρτηση με  $\phi(1) = 0$ .

Υποθέτουμε ότι η  $\mathbf{p}_n \rightarrow \mathbf{p}$  δεν ισχύει. Τότε υπάρχει μια υποακολουθία  $n_1 < n_2 < \dots < n_s < \dots$  φυσικών αριθμών και ένα διάνυσμα  $\mathbf{q}$  τέτοια ώστε

$$\lim_{s \rightarrow \infty} \mathbf{p}_{n_s} = \mathbf{q} \text{ και } \mathbf{p} \neq \mathbf{q}. \quad (3.4)$$

Επειδή η  $\phi$  είναι συνεχής έχουμε ότι

$$\lim_{s \rightarrow \infty} \sum_i p(i) \phi \left( \frac{p_{n_s}(i)}{p(i)} \right) = \sum_i p(i) \phi \left( \frac{q(i)}{p(i)} \right).$$

Όμως  $\left\{ \sum_i p(i) \phi \left( \frac{p_{n_s}(i)}{p(i)} \right) \right\}$  είναι μια υποακολουθία των  $\left\{ \sum_i p(i) \phi \left( \frac{p_n(i)}{p(i)} \right) \right\}$ , η οποία συγκλίνει στο  $\phi(1) = 0$ . Συνεπώς

$$\sum_i p(i) \phi \left( \frac{q(i)}{p(i)} \right) = \phi(1) = 0,$$

το οποίο είναι πιθανόν μόνο εάν  $p(i) = q(i)$ , το οποίο αντικρούει τη Σχέση 3.4. Συνεπώς έχουμε ότι  $\mathbf{p}_n \rightarrow \mathbf{p}$ , οπότε στην περίπτωση όπου  $\phi(x) = x \ln x$  η ασυμπτωτική ιδιότητα

ισχύει για την κατευθυνόμενη απόκλιση των Kullback - Leibler. ■

Όπως ήταν αναμενόμενο, το μέτρο απόκλισης των Kullback - Leibler  $D^{KL}(\mathbf{p}, \mathbf{q})$  με μη πιθανοτικά διανύσματα  $\mathbf{p}, \mathbf{q}$ , δεν ικανοποιεί γενικά τις ιδιότητες της κατευθυνόμενης απόκλισης των Kullback - Leibler με διανύσματα πιθανότητας  $\mathbf{p}^*, \mathbf{q}^*$ . Υπό ορισμένες συνθήκες, κάποιες από αυτές ικανοποιούνται. Πιο συγκεκριμένα, το  $D^{KL}(\mathbf{p}, \mathbf{q})$  είναι μη αρνητικό, προσθετικό, αναλλοίωτο υπό επαρκείς μετασχηματισμούς και μεγαλύτερο από το  $I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$ . Επιπλέον ικανοποιεί την ιδιότητα της μέγιστης πληροφορίας καθώς επίσης και την ασυμπτωτική ιδιότητα. Συνεπώς, το  $D^{KL}(\mathbf{p}, \mathbf{q})$ , σε γενικές γραμμές, μπορεί να θεωρηθεί ως μέτρο απόκλισης και να χρησιμοποιηθεί στην περίπτωση που δεν έχουμε διανύσματα πιθανότητας, υπό την προϋπόθεση ότι  $\sum_i p_i = \sum_i q_i$ .

Πριν κλείσουμε την παράγραφο αυτή κρίνουμε σκόπιμο να αναφέρουμε τη δουλειά του Csiszar (1991), ο οποίος θεώρησε γραμμικά αντίστροφα προβλήματα (linear inverse problems), δηλαδή προγράμματα με γραμμικούς περιορισμούς, με  $n$ -διάστατα διανύσματα πραγματικών αριθμών ή διανύσματα με θετικά στοιχεία ή πιθανοτικά διανύσματα. Σκοπός του ήταν να προσδιορίσει συνεπείς κανόνες επιλογής τέτοιων διανυσμάτων. Οι κανόνες επιλογής - προέκτασης (selection - projection) ελαχιστοποιούν «αποστάσεις» μεταξύ τέτοιων διανυσμάτων ή συναρτήσεων αυτών υπό γραμμικούς περιορισμούς. Διάφορα αξιώματα χαρακτηρίζουν τους κανόνες προέκτασης. Ως πορίσματα, αξιωματικοί χαρακτηρισμοί της μεθόδου ελαχίστων τετραγώνων και των μεθόδων της ελάχιστης διαχωριστικής πληροφορίας και μέγιστης εντροπίας, παρέχονται. Υπό αυτό το πλαίσιο παρουσιάζει μια επέκταση της απόκλισης των Kullback - Leibler σε διανύσματα με θετικά στοιχεία, δηλαδή μη πιθανοτικά διανύσματα. Πιο συγκεκριμένα, ο Csiszar (1991) προσέθεσε την ποσότητα  $\sum_i q_i - \sum_i p_i$  στην έκφραση του τυπικού μέτρου απόκλισης των Kullback - Leibler, δηλαδή όρισε το μέτρο

$$I^{KL}(\mathbf{p}, \mathbf{q}) = \sum_i \left( p_i \ln \frac{p_i}{q_i} - p_i + q_i \right),$$

όπου  $\mathbf{p} = (p_1, \dots, p_n)^T$ ,  $\mathbf{q} = (q_1, \dots, q_n)^T$  είναι διανύσματα με στοιχεία  $n$  θετικούς πραγματικούς αριθμούς.

Η παραπάνω  $I^{KL}$ -απόκλιση είναι μη αρνητική και ισούται με το μηδέν εάν και μόνο εάν ισχύει η ισότητα  $\mathbf{p} = \mathbf{q}$ . Συνεπώς ικανοποιεί την ελάχιστη απαίτηση για μια συνάρτηση να θεωρείται μέτρο απόκλισης. Στην περίπτωση που τα στοιχεία των διανυσμάτων  $\mathbf{p}$  και  $\mathbf{q}$  αθροίζουν στη μονάδα, αντίστοιχα, η παραπάνω απόκλιση είναι το τυπικό μέτρο απόκλισης των Kullback - Leibler. Θα αναφερόμαστε σε αυτό το μέτρο ως *εκτεταμένη απόκλιση των Kullback - Leibler* και θα το συμβολίζουμε με  $I_{ext}^{KL}(\mathbf{p}, \mathbf{q})$ . Για περισσότερες πιο αξιωματικές παρά από την πλευρά της Στατιστικής Θεωρίας Πληροφοριών ιδιότητες του μέτρου  $I_{ext}^{KL}(\mathbf{p}, \mathbf{q})$  ο αναγνώστης παραπέμπεται στον Csiszar (1991). Σημειώνουμε ότι εναλλακτικά κάποιος μπορεί να προσθέσει την ποσότητα  $\sum_i p_i - \sum_i q_i$  στην έκφραση του τυπικού μέτρου απόκλισης

των Kullback - Leibler.

### 3.3 Αποκλίσεις δύναμης με μη πιθανοτικά διανύσματα

Οι Cressie and Read (1984) όρισαν την απόκλιση δύναμης μεταξύ δυο διανυσμάτων πιθανότητας  $\mathbf{p}^*$ ,  $\mathbf{q}^*$  για λόγους καλής προσαρμογής ως

$$I^{CR}(\mathbf{p}^*, \mathbf{q}^*) = \frac{1}{\lambda(\lambda + 1)} \sum_{i=1}^n p_i^* \left[ \left( \frac{p_i^*}{q_i^*} \right)^\lambda - 1 \right],$$

όπου  $\lambda$  είναι μια παράμετρος στο σύνολο των πραγματικών αριθμών. Οι τιμές στα σημεία  $\lambda = 0, -1$  ορίζονται ως όρια. Για  $\lambda \rightarrow 0$ , έχουμε  $I^{CR}(\mathbf{p}^*, \mathbf{q}^*) = \sum_{i=1}^n p_i^* \ln \frac{p_i^*}{q_i^*}$ , η οποία είναι η κατευθυνόμενη απόκλιση των Kullback - Leibler, ενώ για  $\lambda \rightarrow -1$ , έχουμε  $I^{CR}(\mathbf{p}^*, \mathbf{q}^*) = I^{KL}(\mathbf{q}^*, \mathbf{p}^*)$ . Η απόκλιση δύναμης έχει τις ιδιότητες των άλλων μέτρων απόκλισης όπως η μη μηδενικότητα, η συνέχεια, η μη προσθετικότητα και η ισχυρή μη προσθετικότητα. Σημειώνουμε ότι το  $I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  είναι μια *κατευθυνόμενη απόκλιση* (Cressie and Read, 1984).

Επιπλέον οι Cressie and Read (1984) όρισαν κυρίως για την ενοποίηση των υπαρχόντων τεστ καλής προσαρμογής την οικογένεια των στατιστικών συναρτήσεων αποκλίσεων δύναμης. Εάν  $\mathbf{x} = (x_1, x_2, \dots, x_k)^T$  είναι ένα τυχαίο διάνυσμα που ακολουθεί την πολυωνυμική κατανομή με παραμέτρους  $n, \mathbf{p}^*$ , όπου  $\mathbf{p}^* = (p_1^*, p_2^*, \dots, p_k^*)^T$  είναι οι πιθανότητες των κελιών και  $\sum_{i=1}^k x_i = n$  και  $\sum_{i=1}^k p_i^* = 1$  και  $\hat{\mathbf{p}}^* = (\hat{p}_1^*, \hat{p}_2^*, \dots, \hat{p}_k^*)^T$  είναι η εκτιμήτρια μέγιστης πιθανοφάνειας του  $\mathbf{p}^*$ , τότε η οικογένεια των στατιστικών συναρτήσεων αποκλίσεων δύναμης ορίζεται ως

$$2nI^{CR}(\lambda) = \frac{2}{\lambda(\lambda + 1)} \sum_{i=1}^k x_i \left[ \left( \frac{x_i}{n\hat{p}_i^*} \right)^\lambda - 1 \right], \quad (3.5)$$

όπου  $\lambda$  είναι μία παράμετρος πραγματικών αριθμών, η οποία επιλέγεται από το χρήστη. Για τις τιμές  $\lambda = 0$  και  $\lambda = -1$  οι στατιστικές συναρτήσεις ορίζονται ως το όριο του  $2nI^{CR}(\lambda)$  καθώς το  $\lambda \rightarrow 0$  και  $\lambda \rightarrow -1$ , αντίστοιχα.

Μπορεί εύκολα να διαπιστώσει κανείς (Read and Cressie, 1988) ότι το  $2nI^{CR}(\lambda)$  που δίνεται στη Σχέση 3.5 ισούται με (i) τη στατιστική συνάρτηση  $\chi^2$  για  $\lambda = 1$ , (ii) τη στατιστική συνάρτηση  $G^2$  για  $\lambda \rightarrow 0$ , (iii) την τροποποιημένη στατιστική συνάρτηση λόγου πιθανοφανειών για  $\lambda \rightarrow -1$ , (iv) τη στατιστική συνάρτηση των Freeman - Tukey  $F^2$  για  $\lambda = -(1/2)$  και (v) την τροποποιημένη στατιστική συνάρτηση του Neyman  $\chi^2$  για  $\lambda = -2$ . Ως μια εναλλακτική των στατιστικών συναρτήσεων  $\chi^2$  και  $G^2$ , οι Cressie and Read (1984) όρισαν τη στατιστική συνάρτηση απόκλισης δύναμης με  $\lambda = 2/3$ , η οποία βρίσκεται μεταξύ των δυο.



Στόχος της παραγράφου αυτής είναι να εξετάσουμε τί συμβαίνει, ως προς τις ιδιότητες του μέτρου, στην περίπτωση που δεν έχουμε πιθανοτικά διανύσματα πιθανότητας. Αρχικά θα ορίσουμε την κατευθυνόμενη απόκλιση τάξης  $\lambda$  για μη πιθανοτικά διανύσματα.

**Ορισμός 3.4.** Ορίζουμε ως

$$D^{CR}(\mathbf{p}, \mathbf{q}) = \frac{1}{\lambda(\lambda + 1)} \sum_i p_i \left[ \left( \frac{p_i}{q_i} \right)^\lambda - 1 \right], \lambda \in \mathbb{R} \quad (3.6)$$

την κατευθυνόμενη απόκλιση των Cressie - Read τάξης  $\lambda$  μεταξύ δυο μη πιθανοτικών διανυσμάτων  $\mathbf{p} = (p_1, p_2, \dots, p_n)^T > \mathbf{0}$  και  $\mathbf{q} = (q_1, q_2, \dots, q_n)^T > \mathbf{0}$ , όπου  $\sum_i p_i \neq 1$  και  $\sum_i q_i \neq 1$ .

Για τη συνέχεια θα υποθέσουμε ότι  $\lambda \neq 0$  και  $\lambda \neq -1$ . Η σχέση που συνδέει το μέτρο απόκλισης των Cressie - Read μεταξύ μη πιθανοτικών διανυσμάτων και το μέτρο απόκλισης των Cressie - Read μεταξύ πιθανοτικών διανυσμάτων δίνεται από το ακόλουθο λήμμα.

**Λήμμα 3.2.** Για την κατευθυνόμενη απόκλιση των Cressie - Read μεταξύ μη πιθανοτικών διανυσμάτων  $\mathbf{p}, \mathbf{q}$ , ισχύει ότι

$$D^{CR}(\mathbf{p}, \mathbf{q}) = \left( \sum_i p_i \right) k^\lambda \left[ I^{CR}(\mathbf{p}^*, \mathbf{q}^*) - \frac{1 - k^\lambda}{k^\lambda} \frac{1}{\lambda(\lambda + 1)} \right],$$

όπου  $I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  είναι η κατευθυνόμενη απόκλιση των Cressie - Read μεταξύ δυο πιθανοτικών διανυσμάτων  $\mathbf{p}^*, \mathbf{q}^*$  όπως αυτά ορίστηκαν στο Λήμμα 3.1 και  $k = \sum_i p_i / \sum_i q_i$ .

*Απόδειξη.* Λαμβάνοντας υπόψη ότι  $p_i = p_i^* \sum_i p_i$  και  $q_i = q_i^* \sum_i q_i$ , η απόκλιση 3.6 μπορεί να γραφεί ως

$$\begin{aligned} D^{CR}(\mathbf{p}, \mathbf{q}) &= \frac{1}{\lambda(\lambda + 1)} \sum_i \left( p_i^* \sum_i p_i \right) \left[ \left( \frac{p_i^* \sum_i p_i}{q_i^* \sum_i q_i} \right)^\lambda - 1 \right] \\ &= \left( \sum_i p_i \right) \frac{1}{\lambda(\lambda + 1)} \sum_i p_i^* \left[ \left( \frac{p_i^*}{q_i^*} \right)^\lambda k^\lambda - 1 \right] \\ &= \left( \sum_i p_i \right) \frac{k^\lambda}{\lambda(\lambda + 1)} \sum_i p_i^* \left[ \left( \frac{p_i^*}{q_i^*} \right)^\lambda - \frac{1}{k^\lambda} \right] \\ &= \left( \sum_i p_i \right) k^\lambda I^{CR, k^\lambda}(\mathbf{p}^*, \mathbf{q}^*), \end{aligned} \quad (3.7)$$

όπου  $I^{CR,k^\lambda}(\mathbf{p}^*, \mathbf{q}^*) = \frac{1}{\lambda(\lambda+1)} \sum_i p_i^* \left[ \left( \frac{p_i^*}{q_i^*} \right)^\lambda - \frac{1}{k^\lambda} \right]$ . Όμως,

$$I^{CR,k^\lambda}(\mathbf{p}^*, \mathbf{q}^*) = I^{CR}(\mathbf{p}^*, \mathbf{q}^*) - \frac{1 - k^\lambda}{k^\lambda} \frac{1}{\lambda(\lambda + 1)}. \quad (3.8)$$

Έτσι, το επιθυμητό αποτέλεσμα ακολουθεί εύκολα με αντικατάσταση της Σχέσης 3.8 στη Σχέση 3.7. ■

**Πρόταση 3.7.** (Μη αρνητικότητα) Έστω

$$m = \frac{1 - k^\lambda}{k^\lambda} \frac{1}{\lambda(\lambda + 1)}.$$

Τότε ισχύει

$$D^{CR}(\mathbf{p}, \mathbf{q}) \geq 0,$$

εάν μια από τις ακόλουθες συνθήκες ισχύει:

- (i)  $\sum_i p_i = \sum_i q_i$ ;
- (ii)  $\sum_i p_i > \sum_i q_i$  και  $\lambda \notin (-1, 0)$ ;
- (iii)  $\sum_i p_i < \sum_i q_i$  και  $\lambda \in (-1, 0)$ ;
- (iv)  $m < I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ .

Όσον αφορά την ισότητα έχουμε τα εξής:

- (a) Εάν  $\sum_i p_i = \sum_i q_i$  η ισότητα ισχύει εάν και μόνο εάν  $\mathbf{p} = \mathbf{q}$ ;
- (b) εάν  $\sum_i p_i > \sum_i q_i$  ή  $\sum_i p_i < \sum_i q_i$  η ισότητα ισχύει εάν  $m = I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ .

Συμπερασματικά, εάν  $\sum_i p_i = \sum_i q_i$  τότε ισχύει  $D^{CR}(\mathbf{p}, \mathbf{q}) \geq 0$  με ισότητα εάν και μόνο εάν  $\mathbf{p} = \mathbf{q}$ .

Απόδειξη. Από το παραπάνω λήμμα έχουμε  $D^{CR}(\mathbf{p}, \mathbf{q}) = \left( \sum_i p_i \right) k^\lambda [I^{CR}(\mathbf{p}^*, \mathbf{q}^*) - m]$ .

Έχουμε τα εξής:

Εάν  $\sum_i p_i = \sum_i q_i$ , τότε  $k^\lambda = 1$  και ανεξάρτητα από την τιμή  $\lambda$ , ισχύει ότι  $m = 0$  και

συνεπώς  $D^{CR}(\mathbf{p}, \mathbf{q}) = \left( \sum_i p_i \right) I^{CR}(\mathbf{p}^*, \mathbf{q}^*) \geq 0$ .

Εάν  $\sum_i p_i > \sum_i q_i$ , τότε  $k^\lambda > 1$  και  $m < 0$  εάν  $\lambda \notin (-1, 0)$ . Στην περίπτωση αυτή έχουμε ότι  $D^{CR}(\mathbf{p}, \mathbf{q}) \geq 0$  εάν  $m \leq I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ , ενώ όταν  $m \geq I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  έχουμε  $D^{CR}(\mathbf{p}, \mathbf{q}) \leq 0$ .

Εάν  $\sum_i p_i < \sum_i q_i$ , τότε  $k^\lambda < 1$  και  $m < 0$  εάν  $\lambda \in (-1, 0)$ . Έτσι, στην περίπτωση αυτή έχουμε τα ίδια με την παραπάνω περίπτωση.

Συνοψίζοντας, οι συνθήκες (i) - (iv) της πρότασης συνεπάγονται ότι  $D^{CR}(\mathbf{p}, \mathbf{q}) \geq 0$ .

Είναι εύκολο να δούμε ότι όταν  $\sum_i p_i = \sum_i q_i$  η ισότητα ισχύει εάν  $\mathbf{p} = \mathbf{q}$ . Επίσης όταν  $\sum_i p_i > \sum_i q_i$  ή  $\sum_i p_i < \sum_i q_i$  η ισότητα ισχύει εάν  $m = I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ . ■

Σημειώνουμε ότι η ισότητα  $D^{CR}(\mathbf{p}, \mathbf{q}) = 0$  δεν συνεπάγεται απαραίτητα ότι  $\mathbf{p} = \mathbf{q}$  εκτός και εάν ισχύει η ισότητα  $\sum_i p_i = \sum_i q_i$ . Όπως και στην περίπτωση της απόκλισης των Kullback - Leibler, και εδώ έχουμε ότι η ελάχιστη απαίτηση για τη χρήση του  $D^{CR}(\mathbf{p}, \mathbf{q})$  ως μέτρου απόκλισης είναι η ισχύς της ισότητας  $\sum_i p_i = \sum_i q_i$  ανεξάρτητα από την τιμή της παραμέτρου  $\lambda$ .

**Πρόταση 3.8.**  $D^{CR}(\mathbf{p}, \mathbf{q}) \geq I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  όταν μια από τις παρακάτω συνθήκες ισχύει:

(i)  $\sum_i p_i = \sum_i q_i$ , (ii)  $\sum_i p_i > \sum_i q_i$  και  $\lambda \notin (-1, 0)$ , (iii)  $\sum_i p_i < \sum_i q_i$  και  $\lambda \in (-1, 0)$ . Η ισότητα ισχύει εάν  $m = I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  ανεξάρτητα από την τιμή του  $\lambda$ , όπου  $m$  όπως στην Πρόταση 3.7.

*Απόδειξη.* Από το Λήμμα 3.2 γνωρίζουμε ότι  $D^{CR}(\mathbf{p}, \mathbf{q}) = \left( \sum_i p_i \right) k^\lambda [I^{CR}(\mathbf{p}^*, \mathbf{q}^*) - m]$ .

Έχουμε τις τρεις ακόλουθες περιπτώσεις:

Εάν  $\sum_i p_i = \sum_i q_i$ , τότε  $k^\lambda = 1$  και ανεξάρτητα από την τιμή του  $\lambda$ , ισχύει ότι  $m = 0$  και συνεπακόλουθα  $D^{CR}(\mathbf{p}, \mathbf{q}) > I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ .

Εάν  $\sum_i p_i > \sum_i q_i$  (ή ισοδύναμα  $k^\lambda > 1$ ) και  $m < 0$  (όταν  $\lambda \notin (-1, 0)$ ) έχουμε ότι ισχύει πάντοτε  $D^{CR}(\mathbf{p}, \mathbf{q}) > I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ . Στην περίπτωση όπου  $\sum_i p_i > \sum_i q_i$  και  $m > 0$  (όταν  $\lambda \in (-1, 0)$ ) έχουμε ότι εάν  $m > I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  τότε  $D^{CR}(\mathbf{p}, \mathbf{q}) < 0$ , το οποίο είναι αδύνατο ενώ εάν  $m < I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  τότε  $D^{CR}(\mathbf{p}, \mathbf{q}) < I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ .

Εάν  $\sum_i p_i < \sum_i q_i$  (ή ισοδύναμα  $k^\lambda < 1$ ) και  $m < 0$  (όταν  $\lambda \in (-1, 0)$ ) έχουμε ότι ισχύει πάντοτε  $D^{CR}(\mathbf{p}, \mathbf{q}) > I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ .

Συνοψίζοντας, έχουμε ότι οι περιπτώσεις (i) - (iii) συνεπάγονται ότι  $D^{CR}(\mathbf{p}, \mathbf{q}) > I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$ . ■

**Ορισμός 3.5.** (Διμεταβλητή απόκλιση) Στο πλαίσιο του Ορισμού 3.2 ορίζουμε την κατευθυνόμενη απόκλιση των Cressie - Read μεταξύ δυο διμεταβλητών μη πιθανοτικών συναρτήσεων  $p_1, p_2$  ως

$$D_{X,Y}^{CR}(p_1, p_2) = \frac{1}{\lambda(\lambda + 1)} \sum_x \sum_y p_1(x, y) \left[ \left( \frac{p_1(x, y)}{p_2(x, y)} \right)^\lambda - 1 \right].$$

**Ορισμός 3.6.** (Υπό συνθήκη απόκλιση) Στο πλαίσιο του Ορισμού 3.3 θέτουμε

$$D_{Y|X=x}^{CR}(h_1, h_2) = \frac{1}{\lambda(\lambda + 1)} \sum_y h_1(y|x) \left[ \left( \frac{h_1(y|x)}{h_2(y|x)} \right)^\lambda - 1 \right]$$

και

$$\begin{aligned} D_{Y|X}^{CR}(h_1, h_2) &= E_X [D_{Y|X=x}^{CR}(h_1, h_2)] \\ &= \frac{1}{\lambda(\lambda+1)} \sum_x f_1(x) \sum_y h_1(y|x) \left[ \left( \frac{h_1(y|x)}{h_2(y|x)} \right)^\lambda - 1 \right], \end{aligned}$$

για τη μεταβλητή  $X$  ενώ η  $D_{X|Y}^{CR}(r_1, r_2)$  ορίζεται με ανάλογο τρόπο.

Η ισχυρή προσθετικότητα δεν ικανοποιείται για την απόκλιση δύναμης με πιθανοτικά διανύσματα όπως μπορεί κανείς να δει με το ακόλουθο αριθμητικό παράδειγμα με δυο τριωνυμικές κατανομές. Εάν  $(X^*, Y^*, Z^*)$  είναι τριωνυμικές  $M(n, p_i^*)$  με  $p_i^* = (p_{i1}, p_{i2}, p_{i3})$ ,  $p_{i1} + p_{i2} + p_{i3} = 1$ ,  $i = 1, 2$ , χρησιμοποιώντας τυπικά αποτελέσματα, λίγη άλγεβρα και προφανή συμβολισμό, έχουμε ότι

$$I_{X^*, Y^*}^{CR}(p_1^*, p_2^*) = \frac{1}{\lambda(\lambda+1)} \sum_{x,y,z} \binom{n}{x,y,z} p_{11}^x p_{12}^y p_{13}^z \left[ \left( \frac{p_{11}}{p_{21}} \right)^{\lambda x} \left( \frac{p_{12}}{p_{22}} \right)^{\lambda y} \left( \frac{p_{13}}{p_{23}} \right)^{\lambda z} - 1 \right]$$

$$I_{X^*}^{CR}(f_1^*, f_2^*) = \frac{1}{\lambda(\lambda+1)} \sum_x \binom{n}{x} p_{11}^x q_{11}^{n-x} \left[ \left( \frac{p_{11}}{p_{21}} \right)^{\lambda x} \left( \frac{q_{11}}{q_{21}} \right)^{\lambda(n-x)} - 1 \right], \quad q_{i1} = 1 - p_{i1}$$

και

$$I_{Y^*|X^*}^{CR}(h_1^*, h_2^*) = \frac{1}{\lambda(\lambda+1)} \sum_{x,y,z} \binom{n}{x,y,z} p_{11}^x p_{12}^y p_{13}^z \left[ \left( \frac{p_{12}}{p_{22}} \right)^{\lambda y} \left( \frac{p_{13}}{p_{23}} \right)^{\lambda z} \left( \frac{q_{21}}{q_{11}} \right)^{\lambda(n-x)} - 1 \right],$$

όπου  $\binom{n}{x,y,z}$  είναι ο τριωνυμικός συντελεστής και  $x + y + z = n$ . Για  $n = 5$ ,  $p_{11} = 0.2$ ,  $p_{12} = 0.2$ ,  $p_{13} = 0.6$  και  $p_{21} = 0.3$ ,  $p_{22} = 0.4$  και  $p_{23} = 0.3$  και  $\lambda = 1.2$  παίρνουμε

$$I_{X^*}^{CR}(f_1^*, f_2^*) + I_{Y^*|X^*}^{CR}(h_1^*, h_2^*) = 0.133 + 2.037 = 2.17 < I_{X^*, Y^*}^{CR}(p_1^*, p_2^*) = 3.451.$$

Για  $n = 5$ , τα ίδια  $p_{ij}$  και  $\lambda = -0.4$  παίρνουμε

$$I_{X^*}^{CR}(f_1^*, f_2^*) + I_{Y^*|X^*}^{CR}(h_1^*, h_2^*) = 0.132 + 0.804 = 0.936 > I_{X^*, Y^*}^{CR}(p_1^*, p_2^*) = 0.877.$$

Περαιτέρω αριθμητική διερεύνηση αποκάλυψε ότι όταν η παράμετρος  $\lambda$  παίρνει θετικές τιμές ισχύει η ιδιότητα της υποπροσθετικότητας, ενώ όταν η παράμετρος  $\lambda$  παίρνει αρνητικές τιμές ισχύει η ιδιότητα της υπερπροσθετικότητας. Η ισότητα ισχύει μόνο όταν  $\lambda = 0$ , η οποία είναι η περίπτωση της Kullback - Leibler απόκλισης.

Δεν προέκυψε κατάλληλη έκφραση για την ισχυρή προσθετικότητα για την περίπτωση των μη πιθανοτικών διανυσμάτων. Για την ασθενή προσθετικότητα έχουμε την ακόλουθη

πρόταση.

**Πρόταση 3.9.** (Ασθενής προσθετικότητα) Εάν  $h_i(y|x) = g_i(y)$  και συνεπώς  $p_i(x, y) = f_i(x)g_i(y)$ ,  $i = 1, 2$ , έχουμε ότι οι τυχαίες μεταβλητές  $X^*$ ,  $Y^*$ , οι οποίες είναι οι «τυποποιημένες» τιμές των  $X$ ,  $Y$ , είναι ανεξάρτητες, τότε ισχύουν

(a)

$$D_{X,Y}^{CR}(p_1, p_2) = D_X^{CR}(f_1, f_2) + D_Y^{CR}(g_1, g_2) + p_{1..}\eta^\lambda \lambda(\lambda + 1) I_{X^*}^{CR}(f_1^*, f_2^*) I_{Y^*}^{CR}(g_1^*, g_2^*) + p_{1..} (1 - \eta^\lambda) \frac{1}{\lambda(\lambda + 1)}$$

όπου  $p_{i..} = \sum_x \sum_y p_i(x, y)$ ,  $i = 1, 2$  και  $\eta = p_{1..}/p_{2..}$ .

(b)  $D_{X,Y}^{CR}(p_1, p_2) = D_X^{CR}(f_1, f_2) + D_Y^{CR}(g_1, g_2)$  εάν  $\eta = 1$  και εάν ένα από τα περιθώρια ζεύγη  $(f_1^*, f_2^*)$ ,  $(g_1^*, g_2^*)$  είναι identical.

Απόδειξη. (a) Έχουμε ήδη δει στην Πρόταση 3.3 ότι οι τυχαίες μεταβλητές  $X^*$ ,  $Y^*$  είναι ανεξάρτητες. Γνωρίζουμε ότι [βλέπε Read and Cressie (1988)]

$$I_{X^*, Y^*}^{CR}(p_1^*, p_2^*) = I_{X^*}^{CR}(f_1^*, f_2^*) + I_{Y^*}^{CR}(g_1^*, g_2^*) + \lambda(\lambda + 1) I_{X^*}^{CR}(f_1^*, f_2^*) I_{Y^*}^{CR}(g_1^*, g_2^*).$$

Χρησιμοποιώντας το Λήμμα 3.2 έχουμε ότι

$$\begin{aligned} D_{X,Y}^{CR}(p_1, p_2) &= p_{1..}\eta^\lambda \left[ I_{X^*, Y^*}^{CR}(p_1^*, p_2^*) - \frac{1 - \eta^\lambda}{\eta^\lambda \lambda(\lambda + 1)} \right] \\ &= p_{1..}\eta^\lambda \left[ I_{X^*}^{CR}(f_1^*, f_2^*) + I_{Y^*}^{CR}(g_1^*, g_2^*) \right. \\ &\quad \left. + \lambda(\lambda + 1) I_{X^*}^{CR}(f_1^*, f_2^*) I_{Y^*}^{CR}(g_1^*, g_2^*) - \frac{1 - \eta^\lambda}{\eta^\lambda \lambda(\lambda + 1)} \right] \\ &= D_X^{CR}(f_1, f_2) + D_Y^{CR}(g_1, g_2) \\ &\quad + p_{1..}\eta^\lambda \left[ \lambda(\lambda + 1) I_{X^*}^{CR}(f_1^*, f_2^*) I_{Y^*}^{CR}(g_1^*, g_2^*) + \frac{1 - \eta^\lambda}{\eta^\lambda \lambda(\lambda + 1)} \right] \\ &= D_X^{CR}(f_1, f_2) + D_Y^{CR}(g_1, g_2) + p_{1..}\eta^\lambda \lambda(\lambda + 1) I_{X^*}^{CR}(f_1^*, f_2^*) I_{Y^*}^{CR}(g_1^*, g_2^*) \\ &\quad + p_{1..} (1 - \eta^\lambda) \frac{1}{\lambda(\lambda + 1)}. \end{aligned}$$

(b) Εάν  $p_{1..} = p_{2..}$ , τότε  $\eta = 1$ . Ανεξάρτητα λοιπόν από την τιμή του  $\lambda$ , ο τελευταίος όρος της παραπάνω ισότητας ισούται με το 0. Επιπρόσθετα εάν  $f_1^* = f_2^*$  ή  $g_1^* = g_2^*$  τότε  $I_{X^*}^{CR}(f_1^*, f_2^*) I_{Y^*}^{CR}(g_1^*, g_2^*) = 0$ . Συνεπώς  $D_{X,Y}^{CR}(p_1, p_2) = D_X^{CR}(f_1, f_2) + D_Y^{CR}(g_1, g_2)$ . ■

Από την παραπάνω πρόταση, συμπεραίνουμε ότι η ιδιότητα της ασθενούς προσθετικότητας, ισχύει στην περίπτωση όπου  $\eta = 1$  και ένα από τα περιθώρια ζεύγη  $(f_1^*, f_2^*)$ ,  $(g_1^*, g_2^*)$  είναι identical.

**Πρόταση 3.10.** (Μέγιστη πληροφορία και επάρκεια) Έστω  $Y = T(X)$  ένας μετρήσιμος μετασχηματισμός της  $X$ . Τότε ισχύει

$$D_X^{CR}(p_1, p_2) \geq D_Y^{CR}(g_1, g_2),$$

όταν  $c > 1$ , όπου  $c = \left( \frac{\sum_x p_1(x)}{\sum_x p_2(x)} \right)^\lambda$ , με ισότητα εάν και μόνο εάν η  $Y$  είναι «επαρκής», όπως αυτό δικαιολογείται στην Πρόταση 3.4 όπου  $p_i = p_i(x)$ ,  $g_i = g_i(y)$ ,  $i = 1, 2$ .

Απόδειξη. Έστω  $g_i(y)$  το μέτρο που σχετίζεται με την  $Y$ . Τότε  $g_i(y) = \sum_{x:T(x)=y} p_i(x)$ . Οι ακόλουθες ανισότητες είναι ισοδύναμες

$$D_X^{CR}(p_1, p_2) \geq D_Y^{CR}(g_1, g_2) \Leftrightarrow \left( \sum_x p_1(x) \right)^b [I_{X^*}^{CR}(p_1^*, p_2^*) - k] \geq \left( \sum_y g_1(y) \right)^c [I_{Y^*}^{CR}(g_1^*, g_2^*) - l],$$

όπου  $c = \left( \frac{\sum_x p_1(x)}{\sum_x p_2(x)} \right)^\lambda$ ,  $d = \left( \frac{\sum_y g_1(y)}{\sum_y g_2(y)} \right)^\lambda$ ,

$$k = \frac{1-c}{c} \frac{1}{\lambda(\lambda+1)}$$

και

$$l = \frac{1-d}{d} \frac{1}{\lambda(\lambda+1)}.$$

Εφόσον  $\sum_x p_i(x) = \sum_y g_i(y)$ ,  $i = 1, 2$  και συνεπώς  $c = d$  και  $k = l$ , η τελευταία ανισότητα είναι ισοδύναμη με την

$$\left( \sum_x p_1(x) \right)^b [I_{X^*}^{CR}(p_1^*, p_2^*) - k] \geq \left( \sum_x p_1(x) \right)^b [I_{Y^*}^{CR}(g_1^*, g_2^*) - k] \Leftrightarrow I_{X^*}^{CR}(p_1^*, p_2^*) \geq I_{Y^*}^{CR}(g_1^*, g_2^*),$$

η οποία ισχύει όταν  $c > 1$ . Η ισότητα ισχύει εάν και μόνο εάν η στατιστική συνάρτηση  $Y^* = T(X^*)$  είναι επαρκής [Pardo (2006, σελ. 11 - 12)]. ■

Οι Zografos et al. (1989) απέδειξαν ότι η ασυμπτωτική ιδιότητα ισχύει για το μέτρο του Csiszar ( $\phi$ -divergence) που ορίζεται ως

$$I^C(f_1, f_2) = \int f_2(x) \phi \left( \frac{f_1(x)}{f_2(x)} \right) dx,$$

όπου  $\phi$  είναι μια κυρτή συνάρτηση πραγματικών αριθμών που ικανοποιεί ορισμένες συνθήκες.

Η απόκλιση των Cressie and Read προέρχεται από το μέτρο του Csiszar παίρνοντας  $\phi(x) = [\lambda(\lambda + 1)]^{-1}(x^{\lambda+1} - x)$  στη διακριτή περίπτωση του μέτρου (Pardo, 2006). Συνεπώς η ασυμπτωτική ιδιότητα ισχύει επίσης και για το μέτρο των Cressie and Read. Στην επόμενη πρόταση, διερευνούμε εάν η ασυμπτωτική ιδιότητα ισχύει και στην περίπτωση που δεν έχουμε πιθανοτικά διανύσματα.

**Πρόταση 3.11.** (Ασυμπτωτική ιδιότητα) Έστω  $\mathbf{p}_n$  μια ακολουθία από μη πιθανοτικά διανύσματα. Τότε ισχύει  $\mathbf{p}_n \rightarrow \mathbf{p}$  εάν και μόνο εάν  $D^{CR}(\mathbf{p}_n, \mathbf{p}) \rightarrow 0$ , δηλαδή η ασυμπτωτική ιδιότητα ισχύει για το μέτρο των Cressie - Read με μη πιθανοτικά διανύσματα.

Απόδειξη. Χρησιμοποιώντας το Λήμμα 3.2 έχουμε ότι

$$\begin{aligned} \lim_{n \rightarrow \infty} D^{CR}(\mathbf{p}_n, \mathbf{p}) &= \lim_{n \rightarrow \infty} \left\{ \left( \sum_i p_n(i) \right) k^\lambda \left[ I^{CR}(\mathbf{p}_n^*, \mathbf{p}^*) - \frac{1 - k^\lambda}{k^\lambda} \frac{1}{\lambda(\lambda + 1)} \right] \right\} \\ &= \lim_{n \rightarrow \infty} \left( \sum_i p_n(i) \right) \lim_{n \rightarrow \infty} k^\lambda \lim_{n \rightarrow \infty} \left[ I^{CR}(\mathbf{p}_n^*, \mathbf{p}^*) - \frac{1 - k^\lambda}{k^\lambda} \frac{1}{\lambda(\lambda + 1)} \right] \\ &= 0, \end{aligned}$$

επειδή  $\lim_{n \rightarrow \infty} I^{CR}(\mathbf{p}_n^*, \mathbf{p}^*) = 0$  και  $\lim_{n \rightarrow \infty} k^\lambda = 1$ .

Από την άλλη, έστω  $D^{CR}(\mathbf{p}_n, \mathbf{p}) \rightarrow 0$ . Τότε, αγνοώντας τη σταθερά  $1/(\lambda(\lambda + 1))$  έχουμε,

$$\lim_{n \rightarrow \infty} \sum_i p_n(i) \left[ \left( \frac{p_n(i)}{p(i)} \right)^\lambda - 1 \right] = 0$$

ή

$$\lim_{n \rightarrow \infty} \sum_i p_n(i) \phi \left( \frac{p_n(i)}{p(i)} \right) = 0,$$

όπου η συνάρτηση  $\phi(x) = (x^{\lambda+1} - x)$ ,  $x > 0$ ,  $\lambda \neq 0, -1$  είναι μια θετική, συνεχής συνάρτηση για την οποία ισχύει  $\phi(1) = 0$ . Επαναλαμβάνοντας τη διαδικασία του δεύτερου μέρους της απόδειξης της Πρότασης 3.6 παίρνουμε ότι  $\mathbf{p}_n \rightarrow \mathbf{p}$ . Συνεπώς, η ασυμπτωτική ιδιότητα ισχύει για την κατευθυνόμενη απόκλιση των Cressie and Read. ■

Ανακεφαλαιώνοντας τα παραπάνω αποτελέσματα, η κατευθυνόμενη απόκλιση δύναμης  $D^{CR}(\mathbf{p}, \mathbf{q})$ , υπό ορισμένες συνθήκες είναι μη αρνητική, προσθετική, μεγαλύτερη από την  $I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  και αναλλοίωτη υπό επαρκείς μετασχηματισμούς. Επίσης διατηρεί την ιδιότητα της μέγιστης πληροφορίας και τη βασική ασυμπτωτική ιδιότητα. Συνεπώς, μπορούμε να θεωρήσουμε το μέτρο απόκλισης  $D^{CR}(\mathbf{p}, \mathbf{q})$  ως μέτρο απόκλισης, δεδομένου ότι ισχύει η ισότητα  $\sum_i p_i = \sum_i q_i$ .

Υπό το πλαίσιο της δουλειάς του Csiszar (1991), η οποία αναφέρθηκε στην προηγούμενη ενότητα, κάποιος μπορεί να θεωρήσει την εκτεταμένη απόκλιση δύναμης των Cressie and

Read για θετικά διανύσματα  $\mathbf{p}$  και  $\mathbf{q}$

$$I_{ext}^{CR}(\mathbf{p}, \mathbf{q}) = \frac{1}{\lambda(\lambda+1)} \sum_i \left\{ p_i \left[ \left( \frac{p_i}{q_i} \right)^\lambda - 1 \right] - p_i + q_i \right\}, \quad \lambda \in \mathbb{R}.$$

Οι τιμές για την παράμετρο  $\lambda = 0, -1$  ορίζονται ως όρια. Για  $\lambda \rightarrow 0$  το παραπάνω μέτρο είναι το ίδιο με το εκτεταμένο μέτρο απόκλισης των Kullback - Leibler  $I_{ext}^{KL}(\mathbf{p}, \mathbf{q})$ . Για  $\lambda \rightarrow -1$  το μέτρο γίνεται  $I_{ext}^{KL}(\mathbf{q}, \mathbf{p})$ . Για τις τιμές αυτές της παραμέτρου  $\lambda$  είναι γνωστό ότι η πλήρης ιδιότητα της μη αρνητικότητας ικανοποιείται. Θέτοντας  $g_i = p_i/q_i$  παίρνουμε

$$I_{ext}^{CR}(\mathbf{p}, \mathbf{q}) = \frac{1}{\lambda(\lambda+1)} \sum_i p_i (g_i^{\lambda+1} - 2g_i + 1).$$

Για  $\lambda = 1$ , έχουμε  $I_{ext}^{CR}(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \sum_i q_i (g_i - 1)^2 \geq 0$  με ισότητα εάν και μόνο εάν ισχύει η σχέση  $\mathbf{p} = \mathbf{q}$ . Για άλλες τιμές της παραμέτρου  $\lambda$  μια περαιτέρω διερεύνηση για το ελάχιστο της συνάρτησης  $h(y) = y^{\lambda+1} - 2y + 1$ ,  $y > 0$  αποκάλυψε ότι το μέτρο  $I_{ext}^{CR}$  μπορεί να είναι είτε αυστηρά θετικό είτε αυστηρά αρνητικό και συνεπώς δεν ικανοποιεί την ιδιότητα της πλήρους μη αρνητικότητας. Πραγματικό minima του μέτρου  $I_{ext}^{CR}$  εμφανίζεται μόνο εάν  $\lambda > 0$  και, για παράδειγμα, για  $\lambda = 2$  έχουμε ότι το μέτρο  $I_{ext}^{CR}(\mathbf{p}, \mathbf{q}) = 0$  εάν και μόνο εάν ισχύει  $\mathbf{p} = \mathbf{q}$  ή  $\mathbf{p} = (\frac{1}{2}\sqrt{5} - 1)\mathbf{q}$ . Και σε αυτή την περίπτωση, εάν τα  $\mathbf{p}$  και  $\mathbf{q}$  είναι διανύσματα με θετικά στοιχεία που αθροίζουν στη μονάδα, τότε το παραπάνω εκτεταμένο μέτρο απόκλισης είναι το ίδιο με το τυπικό μέτρο απόκλισης δύναμης των Cressie and Read. Επιπλέον, εάν ισχύει η ισότητα  $\sum_i p_i = \sum_i q_i$  το εκτεταμένο μέτρο είναι ταυτόσημο με το μέτρο απόκλισης τάξης  $\lambda$  των Cressie and Read μεταξύ δυο μη πιθανοτικών διανυσμάτων που ορίστηκε στον Ορισμό 3.4. Ο ορισμός του Csizsar στην ουσία ενσωματώνει τον περιορισμό  $\sum_i p_i = \sum_i q_i$  στο μέτρο απόκλισης.

### 3.4 Η διαφορά του Jensen με μη πιθανοτικά διανύσματα

Ένα μέτρο απόκλισης, το οποίο έχει μια ειδική στατιστική σημασία καθώς προέρχεται από την εντροπία του Shannon και την κοίλη ιδιότητα αυτής είναι η διαφορά του Jensen όπως την ονόμασαν οι Burbea and Rao (1982). Η διαφορά του Jensen μεταξύ δυο πιθανοτικών διανυσμάτων δίνεται από τη σχέση

$$J(\mathbf{p}^*, \mathbf{q}^*) \equiv H\left(\frac{1}{2}(\mathbf{p}^* + \mathbf{q}^*)\right) - \frac{1}{2}[H(\mathbf{p}^*) + H(\mathbf{q}^*)],$$

όπου  $H(\mathbf{p}^*) = -\sum_i p_i^* \ln p_i^*$  είναι η εντροπία του Shannon μεταξύ δυο διανυσμάτων πιθανότητας  $\mathbf{p}^* = (p_1^*, \dots, p_n^*)^T$  και  $\mathbf{q}^* = (q_1^*, \dots, q_n^*)^T$ .



Η διαφορά του Jensen είναι ένα φυσικό μέτρο της απόκλισης μεταξύ πιθανοτικών διανυσμάτων  $\mathbf{p}^*$  και  $\mathbf{q}^*$ , καθώς διατηρεί τις δυο βασικές ιδιότητες των μέτρων απόκλισης. Είναι μη αρνητικό και εξαλείφεται εάν και μόνο εάν ισχύει η ισότητα  $\mathbf{p}^* = \mathbf{q}^*$ . Μια επιπλέον ενδιαφέρουσα ιδιότητα του μέτρου  $J(\mathbf{p}^*, \mathbf{q}^*)$  είναι ότι θεωρούμενη ως συνάρτηση του ζεύγους  $(\mathbf{p}^*, \mathbf{q}^*)$  είναι κυρτή.

Στόχος της παραγράφου αυτής είναι να μελετήσουμε τις ιδιότητες της διαφοράς του Jensen όταν δεν έχουμε πιθανοτικά διανύσματα. Αρχικά όμως θα ορίσουμε τη διαφορά του Jensen για μη πιθανοτικά διανύσματα.

**Ορισμός 3.7.** Ορίζουμε ως

$$J(\mathbf{p}, \mathbf{q}) \equiv H\left(\frac{1}{2}(\mathbf{p} + \mathbf{q})\right) - \frac{1}{2}[H(\mathbf{p}) + H(\mathbf{q})], \quad (3.9)$$

τη διαφορά του Jensen μεταξύ των μη πιθανοτικών διανυσμάτων  $\mathbf{p} = (p_1, \dots, p_n)^T$  και  $\mathbf{q} = (q_1, \dots, q_n)^T$ , όπου  $\sum_i p_i \neq 1$  και  $\sum_i q_i \neq 1$ . Η  $H(\mathbf{p}) = -\sum_i p_i \ln p_i$  είναι η αντίστοιχη εντροπία του Shannon.

Το παρακάτω λήμμα μας δείχνει πώς η διαφορά του Jensen με μη πιθανοτικά διανύσματα μπορεί να γραφεί συναρτήσει των αντίστοιχων πιθανοτικών διανυσμάτων.

**Λήμμα 3.3.** Για τη διαφορά του Jensen με μη πιθανοτικά διανύσματα  $\mathbf{p}, \mathbf{q}$  ισχύει η σχέση

$$\begin{aligned} J(\mathbf{p}, \mathbf{q}) &= -\sum_{i=1}^n \frac{1}{2} \left( \frac{p_i^*}{\sum_i p_i} + \frac{q_i^*}{\sum_i q_i} \right) \ln \left[ \frac{1}{2} \left( \frac{p_i^*}{\sum_i p_i} + \frac{q_i^*}{\sum_i q_i} \right) \right] \\ &\quad - \frac{1}{2} \left\{ \frac{1}{\sum_i p_i} \left[ H(\mathbf{p}^*) + \ln \sum_i p_i \right] + \frac{1}{\sum_i q_i} \left[ H(\mathbf{q}^*) + \ln \sum_i q_i \right] \right\}, \end{aligned}$$

όπου  $H(\cdot)$  είναι η εντροπία του Shannon.

*Απόδειξη.* Για την εντροπία του Shannon μεταξύ μη πιθανοτικών διανυσμάτων έχουμε ότι

$$\begin{aligned} H(\mathbf{p}) &= -\sum_i p_i \ln p_i = -\sum_i \left( \frac{p_i^*}{\sum_i p_i} \right) \ln \left( \frac{p_i^*}{\sum_i p_i} \right) \\ &= -\frac{1}{\sum_i p_i} \left[ \sum_i p_i^* \ln p_i^* - \ln \sum_i p_i \right] \\ &= \frac{1}{\sum_i p_i} \left[ H(\mathbf{p}^*) + \ln \sum_i p_i \right], \end{aligned}$$

όπου  $H(\mathbf{p}^*) = -\sum_i p_i^* \ln p_i^*$  είναι η εντροπία του Shannon που σχετίζεται με το πιθανοτικό

διάνυσμα  $\mathbf{p}^*$ . Παρομοίως, έχουμε ότι

$$H(\mathbf{q}) = \frac{1}{\sum_i q_i} \left[ H(\mathbf{q}^*) + \ln \sum_i q_i \right]$$

και

$$H\left(\frac{1}{2}(\mathbf{p} + \mathbf{q})\right) = - \sum_{i=1}^n \frac{1}{2} \left( \frac{p_i^*}{\sum_i p_i} + \frac{q_i^*}{\sum_i q_i} \right) \ln \left[ \frac{1}{2} \left( \frac{p_i^*}{\sum_i p_i} + \frac{q_i^*}{\sum_i q_i} \right) \right].$$

Συνεπώς, αντικαθιστώντας τις παραπάνω σχέσεις στην Εξίσωση 3.9 παίρνουμε το επιθυμητό αποτέλεσμα. ■

Παρατηρούμε ότι η εντροπία  $H\left(\frac{1}{2}(\mathbf{p} + \mathbf{q})\right)$  δεν μπορεί να γραφεί συναρτήσει της  $H\left(\frac{1}{2}(\mathbf{p}^* + \mathbf{q}^*)\right)$ . Το γεγονός αυτό μας δυσκολεύει στο να βρούμε μια γενική και εύχρηστη έκφραση που να συνδέει τη διαφορά του Jensen μεταξύ μη πιθανοτικών διανυσμάτων με τη διαφορά του Jensen μεταξύ πιθανοτικών διανυσμάτων και στη συνέχεια να εξετάσουμε τις ιδιότητές του, όπως έγινε με τα μέτρα των Kullback - Leibler και Cressie and Read.

Για ευκολία στη συνέχεια θα θεωρήσουμε ότι ισχύει η ισότητα  $\sum_i p_i = \sum_i q_i$ . Άλλωστε όπως είδαμε στις δυο προηγούμενες ενότητες η ισότητα αυτή αποτελεί την ελάχιστη απαίτηση για ένα μέτρο απόκλισης με μη πιθανοτικά διανύσματα να θεωρηθεί ως κανονικό μέτρο απόκλισης. Η σχέση που συνδέει τη διαφορά του Jensen μεταξύ μη πιθανοτικών διανυσμάτων με τη διαφορά του Jensen μεταξύ πιθανοτικών διανυσμάτων, σε αυτή την περίπτωση, δίνεται από το παρακάτω λήμμα.

**Πόρισμα 3.2.** Για τη διαφορά του Jensen με μη πιθανοτικά διανύσματα  $\mathbf{p}, \mathbf{q}$ , με  $\sum_i p_i = \sum_i q_i$  ισχύει η σχέση

$$J(\mathbf{p}, \mathbf{q}) = \frac{1}{\sum_i p_i} J(\mathbf{p}^*, \mathbf{q}^*),$$

όπου  $J(\mathbf{p}^*, \mathbf{q}^*)$  είναι η διαφορά του Jensen μεταξύ των πιθανοτικών διανυσμάτων  $\mathbf{p}^*, \mathbf{q}^*$ .

*Απόδειξη.* Από το παραπάνω λήμμα γνωρίζουμε ότι ισχύουν οι σχέσεις

$$H(\mathbf{p}) = \frac{1}{\sum_i p_i} \left[ H(\mathbf{p}^*) + \ln \sum_i p_i \right]$$

και

$$H(\mathbf{q}) = \frac{1}{\sum_i q_i} \left[ H(\mathbf{q}^*) + \ln \sum_i q_i \right].$$

Δοθείσης της ισότητας  $\sum_i p_i = \sum_i q_i$ , έχουμε ότι

$$H(\frac{1}{2}(\mathbf{p} + \mathbf{q})) = \frac{1}{\sum_i p_i} \left[ H(\frac{1}{2}(\mathbf{p}^* + \mathbf{q}^*)) + \ln \sum_i p_i \right].$$

Συνεπώς, αντικαθιστώντας τις παραπάνω σχέσεις στην Εξίσωση 3.9 παίρνουμε το επιθυμητό αποτέλεσμα. ■

**Πρόταση 3.12.** Έστω  $\sum_i p_i = \sum_i q_i$ . Τότε ισχύει  $J(\mathbf{p}, \mathbf{q}) \geq 0$  με την ισότητα να ισχύει εάν και μόνο εάν  $\mathbf{p} = \mathbf{q}$ , όπου  $\mathbf{p}$  και  $\mathbf{q}$  είναι μη πιθανοτικά διανύσματα.

Απόδειξη. Η απόδειξη είναι προφανής δεδομένου ότι  $J(\mathbf{p}^*, \mathbf{q}^*) \geq 0$  και  $\sum_i p_i > 0$ . ■

**Ορισμός 3.8.** (Διμεταβλητή εντροπία του Shannon) Έστω  $p(x, y)$  ένα διμεταβλητό μέτρο (μη πιθανοτική συνάρτηση) που σχετίζεται με δυο διακριτές μεταβλητές  $X, Y$  στο  $R^2$  για το οποίο ισχύει  $\sum_x \sum_y p(x, y) \neq 1$ . Ορίζουμε τη διμεταβλητή εντροπία του Shannon για τη μη πιθανοτική συνάρτηση  $p$  ως

$$H_{X,Y}(p) = - \sum_x \sum_y p(x, y) \ln p(x, y).$$

**Ορισμός 3.9.** (Υπό συνθήκη εντροπία του Shannon) Για τις διακριτές μεταβλητές  $X, Y$  και τη διμεταβλητή μη πιθανοτική συνάρτηση  $p(x, y)$ , όπως αυτή δίνεται παραπάνω, έστω  $f(x) = \sum_y p(x, y)$ ,  $h(y|x) = \frac{p(x,y)}{f(x)}$ ,  $g(y) = \sum_x p(x, y)$ , και  $r(x|y) = \frac{p(x,y)}{g(y)}$ ,  $i = 1, 2$ . Θετούμε

$$H_{Y|X=x}(h) = \sum_y h(y|x) \ln h(y|x), H_{X|Y=y}(r) = \sum_x r(x|y) \ln r(x|y)$$

και ορίζουμε τις

$$H_{Y|X}(h) = E_X [H_{Y|X=x}(h)] = \sum_x f(x) \sum_y h(y|x) \ln h(y|x),$$

$$H_{X|Y}(r) = E_Y [H_{X|Y=y}(r)] = \sum_y g(y) \sum_x r(x|y) \ln r(x|y).$$

**Ορισμός 3.10.** (Διμεταβλητή διαφορά του Jensen) Έστω  $p_i(x, y)$ ,  $i = 1, 2$ , δυο διμεταβλητά μέτρα (μη πιθανοτικές συναρτήσεις) που σχετίζονται με δυο διακριτές μεταβλητές  $X, Y$  στο  $R^2$  για τις οποίες ισχύει  $\sum_x \sum_y p_i(x, y) \neq 1$ . Ορίζουμε τη διαφορά του Jensen μεταξύ δυο

διμεταβλητών μη πιθανοτικών συναρτήσεων  $p_1, p_2$  ως

$$\begin{aligned} J_{X,Y}(p_1, p_2) &= H\left(\frac{1}{2}(p_1 + p_2)\right) - \frac{1}{2}[H(p_1) + H(p_2)] \\ &= -\sum_x \sum_y \frac{1}{2}(p_1(x, y) + p_2(x, y)) \ln\left(\frac{1}{2}(p_1(x, y) + p_2(x, y))\right) \\ &\quad - \frac{1}{2} \left[ -\sum_x \sum_y p_1(x, y) \ln p_1(x, y) - \sum_x \sum_y p_2(x, y) \ln p_2(x, y) \right]. \end{aligned}$$

**Ορισμός 3.11.** (Υπό συνθήκη διαφορά του Jensen) Για τις διακριτές μεταβλητές  $X, Y$  και τις διμεταβλητές μη πιθανοτικές συναρτήσεις  $p_i(x, y)$ ,  $i = 1, 2$ , όπως αυτές δόθηκαν παραπάνω, έστω  $f_i(x) = \sum_y p_i(x, y)$ ,  $h_i(y|x) = \frac{p_i(x, y)}{f_i(x)}$ ,  $g_i(y) = \sum_x p_i(x, y)$ , και  $r_i(x|y) = \frac{p_i(x, y)}{g_i(y)}$ ,  $i = 1, 2$ . Θέτουμε

$$\begin{aligned} J_{Y|X=x}(h_1, h_2) &= H\left(\frac{1}{2}(h_1 + h_2)\right) - \frac{1}{2}[H(h_1) + H(h_2)] \\ &= -\sum_x \sum_y \frac{1}{2}(h_1(y|x) + h_2(y|x)) \ln\left(\frac{1}{2}(h_1(y|x) + h_2(y|x))\right) \\ &\quad - \frac{1}{2} \left[ -\sum_x \sum_y h_1(y|x) \ln h_1(y|x) - \sum_x \sum_y h_2(y|x) \ln h_2(y|x) \right], \end{aligned}$$

$$\begin{aligned} J_{X|Y=y}(r_1, r_2) &= H\left(\frac{1}{2}(r_1 + r_2)\right) - \frac{1}{2}[H(r_1) + H(r_2)] \\ &= -\sum_x \sum_y \frac{1}{2}(r_1(x|y) + r_2(x|y)) \ln\left(\frac{1}{2}(r_1(x|y) + r_2(x|y))\right) \\ &\quad - \frac{1}{2} \left[ -\sum_x \sum_y r_1(x|y) \ln r_1(x|y) - \sum_x \sum_y r_2(x|y) \ln r_2(x|y) \right] \end{aligned}$$

και ορίζουμε την

$$\begin{aligned} J_{Y|X}(h_1, h_2) &= E_X [J_{Y|X=x}(h_1, h_2)] \\ &= -\sum_x f_1(x) \sum_y \frac{1}{2}(h_1(y|x) + h_2(y|x)) \ln\left(\frac{1}{2}(h_1(y|x) + h_2(y|x))\right) \\ &\quad - \frac{1}{2} \left[ -\sum_x f_1(x) \sum_y h_1(y|x) \ln h_1(y|x) - \sum_x f_1(x) \sum_y h_2(y|x) \ln h_2(y|x) \right]. \end{aligned}$$

Με παρόμοιο τρόπο ορίζουμε την

$$\begin{aligned} J_{X|Y}(r_1, r_2) &= E_X [J_{X|Y=y}(r_1, r_2)] \\ &= - \sum_x g_1(x) \sum_y \frac{1}{2}(r_1(x|y) + r_2(x|y)) \ln(\frac{1}{2}(r_1(x|y) + r_2(x|y))) \\ &\quad - \frac{1}{2} \left[ - \sum_x g_1(x) \sum_y r_1(x|y) \ln r_1(x|y) - \sum_x g_1(x) \sum_y r_2(x|y) \ln r_2(x|y) \right]. \end{aligned}$$

**Πρόταση 3.13.** (Ισχυρή προσθετικότητα) Έστω  $p_1, p_2$  δυο διμεταβλητές μη πιθανοτικές μεταβλητές που σχετίζονται με δυο διακριτές μεταβλητές  $X, Y$  στο  $R^2$  όπως στον Ορισμό 3.11. Τότε

$$J_{X,Y}(p_1, p_2) = J_X(f_1, f_2) + J_{Y|X}(h_1, h_2) = J_Y(g_1, g_2) + J_{X|Y}(r_1, r_2),$$

όπου οι συναρτήσεις  $f_i, h_i, g_i, r_i, i = 1, 2$  είναι όπως στον Ορισμό 3.11.

Απόδειξη. Είναι γνωστό (Shannon, 1948) ότι

$$H_{X^*,Y^*}(p_i^*) = H_{X^*}(f_i^*) + H_{Y^*|X^*}(h_i^*) = H_{Y^*}(g_i^*) + H_{X^*|Y^*}(r_i^*), \quad i = 1, 2.$$

Συνεπώς, έχουμε ότι

$$\begin{aligned} J_{X^*,Y^*}(p_1^*, p_2^*) &= H_{X^*,Y^*}(\frac{1}{2}(p_1^* + p_2^*)) - \frac{1}{2} [H_{X^*,Y^*}(p_1^*) + H_{X^*,Y^*}(p_2^*)] \\ &= H_{X^*}(\frac{1}{2}(f_1^* + f_2^*)) + H_{Y^*|X^*}(\frac{1}{2}(h_1^* + h_2^*)) \\ &\quad - \frac{1}{2} [H_{X^*}(f_1^*) + H_{Y^*|X^*}(h_1^*) + H_{X^*}(f_2^*) + H_{Y^*|X^*}(h_2^*)] \\ &= H_{X^*}(\frac{1}{2}(f_1^* + f_2^*)) - \frac{1}{2} [H_{X^*}(f_1^*) + H_{X^*}(f_2^*)] \\ &\quad + H_{Y^*|X^*}(\frac{1}{2}(h_1^* + h_2^*)) - \frac{1}{2} [H_{Y^*|X^*}(h_1^*) + H_{Y^*|X^*}(h_2^*)] \\ &= J_{X^*}(f_1^*, f_2^*) + J_{Y^*|X^*}(h_1^*, h_2^*), \end{aligned}$$

το οποίο σημαίνει ότι η ισχυρή προσθετικότητα ισχύει για τη διαφορά του Jensen. Με παρόμοιο τρόπο αποδεικνύεται ότι ισχύει η σχέση

$$J_{X^*,Y^*}(p_1^*, p_2^*) = J_{Y^*}(g_1^*, g_2^*) + J_{X^*|Y^*}(r_1^*, r_2^*).$$

Για τις μεταβλητές  $X, Y$  έχουμε ότι

$$\begin{aligned}
 J_{X,Y}(p_1, p_2) &= \frac{1}{\sum_x \sum_y p_1(x, y)} J_{x^*, y^*}(p_1^*, p_2^*) \\
 &= \frac{1}{\sum_x \sum_y p_1(x, y)} [J_{X^*}(f_1^*, f_2^*) + J_{Y^*|X^*}(h_1^*, h_2^*)] \\
 &= \frac{1}{\sum_x \sum_y p_1(x, y)} J_{X^*}(f_1^*, f_2^*) + \frac{1}{\sum_x \sum_y p_1(x, y)} J_{Y^*|X^*}(h_1^*, h_2^*) \\
 &= J_X(f_1, f_2) + J_{Y|X}(h_1, h_2),
 \end{aligned}$$

καθώς  $\sum_x \sum_y p_1(x, y) = \sum_x f_1(x) = \sum_y g_1(y)$ . Με παρόμοιο τρόπο, αποδεικνύουμε ότι

$$J_{X,Y}(p_1, p_2) = J_Y(g_1, g_2) + J_{X|Y}(r_1, r_2).$$

■

Για την ασθενή προσθετικότητα έχουμε την παρακάτω πρόταση.

**Πρόταση 3.14.** (Ασθενής προσθετικότητα) Εάν  $h_i(y|x) = g_i(y)$  και συνεπώς  $p_i(x, y) = f_i(x)g_i(y)$ ,  $i = 1, 2$ , έχουμε ότι οι τυχαίες μεταβλητές  $X^*, Y^*$ , για τις οποίες οι “τυποποιημένες” τιμές των  $X, Y$ , είναι ανεξάρτητες. Τότε

$$J_{X,Y}(p_1, p_2) = J_X(f_1, f_2) + J_Y(g_1, g_2).$$

Απόδειξη. Είναι γνωστό (Shannon, 1948) ότι

$$H_{X^*, Y^*}(p_i^*) = H_{X^*}(f_i^*) + H_{Y^*}(g_i^*), \quad i = 1, 2.$$

Συνεπώς, έχουμε ότι

$$\begin{aligned}
 J_{X^*, Y^*}(p_1^*, p_2^*) &= H_{X^*, Y^*}(\frac{1}{2}(p_1^* + p_2^*)) - \frac{1}{2} [H_{X^*, Y^*}(p_1^*) + H_{X^*, Y^*}(p_2^*)] \\
 &= H_{X^*}(\frac{1}{2}(f_1^* + f_2^*)) + H_{Y^*}(\frac{1}{2}(g_1^* + g_2^*)) \\
 &\quad - \frac{1}{2} [H_{X^*}(f_1^*) + H_{Y^*}(g_1^*) + H_{X^*}(f_2^*) + H_{Y^*}(g_2^*)] \\
 &= H_{X^*}(\frac{1}{2}(f_1^* + f_2^*)) - \frac{1}{2} [H_{X^*}(f_1^*) + H_{X^*}(f_2^*)] \\
 &\quad + H_{Y^*}(\frac{1}{2}(g_1^* + g_2^*)) - \frac{1}{2} [H_{Y^*}(g_1^*) + H_{Y^*}(g_2^*)] \\
 &= J_{X^*}(f_1^*, f_2^*) + J_{Y^*}(g_1^*, g_2^*),
 \end{aligned}$$

το οποίο σημαίνει ότι η ιδιότητα της ασθενούς προσθετικότητας ισχύει για τη διαφορά του Jensen.

Στη συνέχεια, για τις μεταβλητές  $X, Y$  έχουμε ότι

$$\begin{aligned} J_{X,Y}(p_1, p_2) &= \frac{1}{\sum_x \sum_y p_1(x, y)} J_{X^*, Y^*}(p_1^*, p_2^*) \\ &= \frac{1}{\sum_x \sum_y p_1(x, y)} [J_{X^*}(f_1^*, f_2^*) + I_{Y^*}(g_1^*, g_2^*)]. \end{aligned}$$

Επειδή ισχύει ότι  $\sum_x \sum_y p_i(x, y) = \sum_x f_i(x) = \sum_y g_i(y)$ ,  $i = 1, 2$  τελικά έχουμε ότι

$$J_{X,Y}(p_1, p_2) = J_X(f_1, f_2) + J_Y(g_1, g_2).$$

■

**Πρόταση 3.15.** (Μέγιστη πληροφορία και επάρκεια) Έστω  $Y = T(X)$  ένας μετρήσιμος μετασχηματισμός της  $X$ . Τότε

$$J_X(p_1, p_2) \geq J_Y(g_1, g_2),$$

με την ισότητα να ισχύει εάν και μόνο εάν η  $Y$  είναι «επαρκής», όπου  $p_i = p_i(x)$ ,  $g_i = g_i(y)$ ,  $i = 1, 2$ .

Απόδειξη. Έστω  $g_i(y)$  είναι το μέτρο που σχετίζεται με την  $Y$ . Τότε  $g_i(y) = \sum_{x:T(x)=y} p_i(x)$ .

Οι ακόλουθες ανισότητες είναι ισοδύναμες

$$\begin{aligned} J_X(p_1, p_2) \geq J_Y(g_1, g_2) &\Leftrightarrow \\ \frac{1}{\sum_x p_1(x)} J_{X^*}(p_1^*, p_2^*) &\geq \frac{1}{\sum_y g_1(y)} J_{Y^*}(g_1^*, g_2^*). \end{aligned}$$

Επειδή ισχύει  $\sum_x p_i(x) = \sum_y g_i(y)$ ,  $i = 1, 2$ , η τελευταία ανισότητα είναι ισοδύναμη με την

$$J_{X^*}(p_1^*, p_2^*) \geq J_{Y^*}(g_1^*, g_2^*),$$

η οποία ισχύει πάντοτε. Η ισότητα ισχύει εάν και μόνο εάν η στατιστική συνάρτηση  $Y^* = T(X^*)$  είναι επαρκής. ■

Με βάση τα αποτελέσματα της παραπάνω ανάλυσης, συμπεραίνουμε ότι η διαφορά του Jensen  $J(\mathbf{p}, \mathbf{q})$  για μη πιθανοτικά διανύσματα, υπό ορισμένες συνθήκες είναι μη αρνητική, προσθετική και αναλλοίωτη υπό επαρκής μετασχηματισμούς. Επιπλέον διατηρεί την ιδιότητα της μέγιστης πληροφορίας. Συνεπώς, μπορούμε να θεωρήσουμε το μέτρο  $J(\mathbf{p}, \mathbf{q})$  ως μέτρο απόκλισης, με την προϋπόθεση ότι ισχύει η ισότητα  $\sum_i p_i = \sum_i q_i$ .

### 3.5 Συμπεράσματα Κεφαλαίου

Σκοπός του κεφαλαίου αυτού ήταν η διερεύνηση των ιδιοτήτων ορισμένων μέτρων απόκλισης όταν σε αυτά εμπλέκονται μη πιθανοτικά διανύσματα. Το ερέθισμα για τη διερεύνηση αυτή δόθηκε από τη χρήση του μέτρου των Kullback - Leibler στο πρόβλημα της εξομάλυνσης ποσοστών θνησιμότητας από τους Brockett and Zhang (1986), όπου τα ποσοστά θνησιμότητας, ως γνωστόν, δεν αθροίζουν στη μονάδα αλλά την ξεπερνούν. Συγκεκριμένα, τα μέτρα των οποίων τις ιδιότητες ελέγξαμε είναι το μέτρο των Kullback - Leibler, το μέτρο απόκλισης δύναμης των Cressie and Read και η διαφορά του Jensen.

Η θεωρητική διερεύνηση του μέτρου απόκλισης των Kullback - Leibler  $D^{KL}(\mathbf{p}, \mathbf{q})$  με μη πιθανοτικά διανύσματα, έδειξε ότι το μέτρο αυτό ικανοποιεί ορισμένες από τις βασικές ιδιότητες των μέτρων απόκλισης και στατιστικής πληροφορίας, τις οποίες ικανοποιεί και το μέτρο κατευθυνόμενης απόκλισης των Kullback - Leibler  $I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$ . Υπό ορισμένες συνθήκες, το μέτρο  $D^{KL}(\mathbf{p}, \mathbf{q})$  είναι μη αρνητικό, προσθετικό και αναλλοίωτο υπό επαρκείς περιορισμούς. Επιπλέον, το μέτρο αυτό ικανοποιεί την ιδιότητα της μέγιστης πληροφορίας. Συνεπώς, το μέτρο  $D^{KL}(\mathbf{p}, \mathbf{q})$  μπορεί να θεωρηθεί ως μέτρο απόκλισης και πληροφορίας.

Στη συνέχεια, αποδείξαμε ότι στην περίπτωση που δεν έχουμε πιθανοτικά διανύσματα, το μέτρο απόκλισης δύναμης των Cressie and Read ικανοποιεί ορισμένες από τις ιδιότητες που το μέτρο με πιθανοτικά διανύσματα ικανοποιεί. Πιο συγκεκριμένα, υπό ορισμένες συνθήκες είναι μη αρνητικό, προσθετικό, μεγαλύτερο από το μέτρο  $I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  και αναλλοίωτο υπό επαρκείς περιορισμούς. Επιπλέον, ικανοποιείται η ιδιότητα της μέγιστης πληροφορίας.

Μια επέκταση των μέτρων απόκλισης και ιδιαίτερα του μέτρου κατευθυνόμενης των Kullback - Leibler για μη αρνητικές συναρτήσεις ή διανύσματα έγινε από τον Csiszar (1991) προσθέτοντας την ποσότητα  $\sum_i q_i - \sum_i p_i$  στην έκφραση του τυπικού μέτρου. Η εκτεταμένη απόκλιση των Kullback - Leibler ικανοποιεί την ελάχιστη απαίτηση για μια συνάρτηση να θεωρείται μέτρο απόκλισης. Η ίδια τροποποίηση εφαρμόστηκε επίσης στο μέτρο απόκλισης δύναμης των Cressie - Read. Όμως αυτό το εκτεταμένο μέτρο ικανοποιεί την ελάχιστη απαίτηση μόνο για  $\lambda = 0$ ,  $\lambda = 1$  και  $\lambda = -1$ . Όταν ισχύει η ισότητα  $\sum_i p_i = \sum_i q_i$  τα δυο προαναφερθέντα εκτεταμένα μέτρα ταυτίζονται με τα μέτρα απόκλισης των Kullback - Leibler και Cressie - Read για μη πιθανοτικά διανύσματα, αντίστοιχα, όπως αυτά ορίστηκαν στους Ορισμούς 3.1 και 3.4, αντίστοιχα.

Παρόμοια ήταν τα αποτελέσματα της διερεύνησης και για τη διαφορά του Jensen που πρότειναν οι Burbea and Rao (1982). Και αυτό το μέτρο ικανοποιεί τις παραπάνω βασικές ιδιότητες των μέτρων απόκλισης και στατιστικής πληροφορίας.

Ένα σημαντικό, κατά τη γνώμη μας, αποτέλεσμα της διερεύνησης και ανάλυσης που διεξάγαμε ήταν η αποκάλυψη ενός περιορισμού ή συνθήκης για την ισχύ των ιδιοτήτων που αναφέραμε παραπάνω. Συγκεκριμένα, για να χαρακτηριστεί ένα μέτρο απόκλισης μεταξύ μη πιθανοτικών διανυσμάτων ως μέτρο απόκλισης και στατιστικής πληροφορίας, θα πρέπει



να ισχύει η ισότητα  $\sum_i p_i = \sum_i q_i$ . Η συνθήκη αυτή είναι μια γενική συνθήκη που περιλαμβάνει και την ισότητα  $\sum_i p_i^* = \sum_i q_i^* = 1$ , η οποία ισχύει στην περίπτωση των πιθανοτικών διανυσμάτων.

Στο επόμενο κεφάλαιο θα περιγράψουμε τη χρήση των μέτρων απόκλισης στο πρόβλημα της αναλογιστικής εξομάλυνσης ενώ σε κεφάλαιο που ακολουθεί θα γίνει εμφανής η χρησιμότητα της συνθήκης  $\sum_i p_i = \sum_i q_i$ , μέσω μιας αριθμητικής διερεύνησης.

РАНЕЕ НЕ ПЕРПА

## ΚΕΦΑΛΑΙΟ 4

# ΕΞΟΜΑΛΥΝΣΗ ΠΙΝΑΚΩΝ ΘΝΗΣΙΜΟΤΗΤΑΣ ΜΕΣΩ ΑΠΟΚΛΙΣΕΩΝ

### 4.1 Εισαγωγή

Η διερεύνηση των ιδιοτήτων των μέτρων απόκλισης όταν σε αυτά δεν εμπλέκονται πιθανοτικά διανύσματα, η οποία παρουσιάστηκε στο προηγούμενο κεφάλαιο, έδειξε ότι τα μέτρα αυτά μπορούν υπό ορισμένες συνθήκες να θεωρηθούν ως μέτρα απόκλισης με την έννοια που τα γνωρίζουμε. Συνεπώς τα μέτρα αυτά μπορούν να χρησιμοποιηθούν ως μέτρα απόκλισης, με την έννοια που αυτά έχουν στη Στατιστική Θεωρία Πληροφοριών, σε προβλήματα που εμπλέκονται μη πιθανοτικά διανύσματα. Ένα τέτοιο πρόβλημα, το οποίο εμφανίζεται στην Αναλογιστική Επιστήμη, τη Βιοστατιστική και τη Δημογραφία, είναι το πρόβλημα της εξομάλυνσης των ποσοστών θνησιμότητας ή άλλων ανάλογων ποσοτήτων.

Στο κεφάλαιο αυτό όπως και σε ολόκληρη τη διατριβή, θα περιοριστούμε στην περιγραφή της εξομάλυνσης ποσοστών θνησιμότητας. Η επέκταση στην περίπτωση άλλων ποσοτήτων νομίζουμε ότι είναι εύκολα κατανοητή. Θα ορίσουμε την έννοια της εξομάλυνσης, θα περιγράψουμε τα βασικά χαρακτηριστικά της, θα αναφερθούμε στους στατιστικούς ελέγχους που χρησιμοποιούνται για τον έλεγχο της εξομάλυνσης ενώ στο τέλος θα αναφερθούμε με συντομία στις μεθόδους εξομάλυνσης που έχουν παρουσιαστεί στη βιβλιογραφία και θα παρουσιάσουμε το πλαίσιο χρήσης των μέτρων απόκλισης.

### 4.2 Τι είναι εξομάλυνση

Για να μπορέσουμε να ορίσουμε την έννοια της εξομάλυνσης, θα πρέπει πρώτα να αναφερθούμε στον τρόπο υπολογισμού των ποσοστών θνησιμότητας. Στην ουσία δεν υπολογίζουμε τα πραγματικά ποσοστά θνησιμότητας, αλλά απλά τα εκτιμούμε. Γνωρίζοντας τον αριθμό

των θανάτων στην ηλικία  $x$ , έστω  $d_x$ , και τον αριθμό των ατόμων που βρίσκονται σε κίνδυνο στην ίδια ηλικία, έστω  $l_x$ , ο πιο απλός τρόπος εκτίμησης του ποσοστού θνησιμότητας γίνεται μέσω του τύπου

$$\overset{\circ}{q}_x = \frac{d_x}{l_x}, \quad x = 1, 2, \dots, n.$$

Το ποσοστό  $\overset{\circ}{q}_x$  αναφέρεται στη βιβλιογραφία ως *αδρός δείκτης θνησιμότητας* στην ηλικία  $x$ . Με άλλα λόγια μπορούμε να πούμε ότι ο αδρός δείκτης θνησιμότητας είναι η ανάλογη έννοια της έντασης θνησιμότητας (force of mortality), στην περίπτωση όπου η τυχαία μεταβλητή  $T$  που περιγράφει το χρόνο ζωής ενός ατόμου είναι διακριτή.

Επειδή συνήθως υποθέτουμε ότι το  $l_x$  είναι μεγάλο, συγκρινόμενο με το  $d_x$ , μπορούμε να υποθέσουμε ότι το  $d_x$  ακολουθεί τη διωνυμική κατανομή με παραμέτρους  $l_x$  και  $q_x$  και συνεπώς ο εκτιμητής  $\overset{\circ}{q}_x$  μπορεί να θεωρηθεί ως διωνυμικό ποσοστό. Έτσι μπορούμε να γράψουμε

$$E\left(\overset{\circ}{q}_x\right) = q_x$$

και

$$\text{var}\left(\overset{\circ}{q}_x\right) = \frac{q_x(1 - q_x)}{l_x},$$

όπου  $q_x$  είναι το πραγματικό ποσοστό θνησιμότητας. Βέβαια ο αριθμός των θανάτων δεν ακολουθεί ακριβώς διωνυμική κατανομή, καθώς το  $l_x$  δεν αντιστοιχεί ακριβώς σε αριθμό ανεξάρτητων διωνυμικών δοκιμών (London, 1985). Επιπλέον, αν παραστήσουμε γραφικά τα  $\overset{\circ}{q}_x$ ,  $x = 1, 2, \dots, n$ , τα οποία, υπό την υπόθεση της διωνυμικής κατανομής, είναι και εκτιμητές μέγιστης πιθανοφάνειας, θα πάρουμε ένα διάγραμμα με μεγάλες διακυμάνσεις, καθώς κάποια από αυτά θα είναι μεγαλύτερα και κάποια μικρότερα από τα πραγματικά ποσοστά θνησιμότητας του πληθυσμού στις αντίστοιχες ηλικίες. Το επόμενο λοιπόν βήμα της κατασκευής των μοντέλων θνησιμότητας, είναι η αναθεώρηση των αρχικών εκτιμήσεων  $\overset{\circ}{q}_x$  για να προκύψουν καλύτερες εκτιμήσεις των πραγματικών ποσοστών. Η αναθεώρηση αυτή ονομάζεται *εξομάλυνση (graduation)*.

Ο Miller (1949) όρισε την εξομάλυνση ως τη διαδικασία να πάρουμε, από μια μη ομαλή - μη κανονική (irregular) σειρά παρατηρούμενων τιμών μιας συνεχούς μεταβλητής, μια ομαλή σειρά τιμών οι οποίες να είναι «συνεπείς» με τις παρατηρούμενες τιμές. Ένα ερώτημα που γεννιέται είναι αν η εξομάλυνση μπορεί να γίνει σε οποιαδήποτε σειρά παρατηρούμενων τιμών. Σύμφωνα με τον London (1985), η απάντηση δίνεται από τα ίδια τα δεδομένα και όλες οι σειρές δεδομένων δεν ενδείκνυνται για εξομάλυνση. Συγκεκριμένα αναφέρει ότι «μόνο ορισμένοι τύποι δεδομένων είναι κατάλληλοι για εξομάλυνση, δηλαδή αυτοί για τους οποίους πιστεύεται ότι υπάρχει σχέση μεταξύ των στοιχείων της σειράς των δεδομένων». Ο Haberman (1998), για το ίδιο θέμα, αναφέρει ότι αν ο ερευνητής πιστεύει ότι τα δεδομένα είναι ανεξάρτητα μεταξύ τους, τότε οι αδρές τιμές θα είναι οι τελικές εκτιμήσεις των πραγματικών δεδομένων. Σε αντίθετη περίπτωση, πρέπει να γίνει εξομάλυνση.

Στην Αναλογιστική Επιστήμη, εξομάλυνση μπορεί να γίνει είτε σε εκτιμήσεις των ποσοστών θνησιμότητας  $q_x$  είτε σε εκτιμήσεις της έντασης θνησιμότητας  $\mu_x$ , χωρίς να αποκλείονται ποσότητες όπως ο αριθμός ατόμων σε κίνδυνο, συνταξιοδοτικά ποσά, ποσοστά αποχωρήσεων και άλλες ασφαλιστικές ποσότητες. Στη Βιοστατιστική εξομάλυνση μπορεί να εφαρμοσθεί σε ποσοστά νοσηρότητας (Haberman, 1983 και 1984). Ο Haberman (1998) αναφέρει ότι υπάρχει μια εκ των προτέρων άποψη, ότι κάθε πραγματική τιμή του  $q_x$  (ή του  $\mu_x$ ) είναι στενά συσχετισμένη με την επόμενη της και η σχέση αυτή εκφράζεται από την άποψη ότι οι τιμές των ποσοτήτων αυτών αυξάνονται ομαλά από ηλικία σε ηλικία. Συνεπώς χρειάζεται να γίνει εξομάλυνση των αρχικών εκτιμήσεων των ποσοστών θνησιμότητας ή της έντασης θνησιμότητας.

Ας δούμε τώρα πώς μπορούμε να εκτιμήσουμε την ένταση θνησιμότητας. Για να το επιτύχουμε αυτό θα στηριχθούμε στην υπόθεση ότι οι θάνατοι  $d_x$  είναι ανεξάρτητοι μεταξύ τους και κατανέμονται ως Poisson κατανομή με παράμετρο  $r_x^c \mu_x$ , όπου  $r_x^c$  είναι ο κεντρικός (central) χρόνος έκθεσης στον κίνδυνο του θανάτου. Η υπόθεση της κατανομής Poisson, δικαιολογείται με το ακόλουθο σκεπτικό: Υποθέτουμε ότι ένα άτομο μπορεί να πεθάνει ανά πάσα στιγμή μέσα στο χρονικό διάστημα  $(0, r_x^c)$ , το οποίο διαιρούμε σε  $n = \frac{r_x^c}{\Delta x}$  υποδιαστήματα μήκους  $\Delta x$ . Υποθέτουμε ότι η πιθανότητα  $p$  να πεθάνει ένα άτομο σε οποιοδήποτε υποδιάστημα  $(x, x + \Delta x)$ ,  $0 \leq x \leq r_x^c$  είναι  $\mu_x \Delta x$ , η οποία είναι σταθερή σε κάθε υποδιάστημα. Αν  $X$  είναι ο αριθμός των «επιτυχιών», δηλαδή των θανάτων στο  $(0, r_x^c)$  και υποθέσουμε ότι οι θάνατοι είναι ανεξάρτητοι από υποδιάστημα σε υποδιάστημα και ότι η πιθανότητα δυο ή περισσότερων θανάτων στο  $(x, x + \Delta x)$  είναι αμελητέα λόγω του μικρού  $\Delta x$ , τότε πρακτικά μπορούμε να θεωρήσουμε ότι η τυχαία μεταβλητή  $X$  ακολουθεί τη διωνυμική κατανομή με παραμέτρους  $n$  και  $p$ , όπου  $n = \frac{r_x^c}{\Delta x}$  και  $p = \mu_x r_x^c$ . Είναι γνωστό ότι το όριο της διωνυμικής κατανομής  $B(n, p)$  με  $n \rightarrow +\infty$  και  $np = \mu_x r_x^c$  σταθερό είναι η κατανομή Poisson με παράμετρο  $\mu_x r_x^c$  (βλέπε Παπαϊωάννου, 2000). Χρησιμοποιώντας τη μέθοδο της μέγιστης πιθανοφάνειας, προκύπτει ότι η εκτίμηση της έντασης θνησιμότητας δίνεται από τη σχέση

$$\mu_x = \frac{d_x}{r_x^c},$$

η οποία μπορεί να συμβολισθεί και με  $\overset{\circ}{\mu}_x$  κατά αντιστοιχία με το συμβολισμό του αδρού δείκτη θνησιμότητας  $\overset{\circ}{q}_x$ .

Παρατηρούμε ότι για την εκτίμηση της έντασης θνησιμότητας  $\mu_x$ , πρέπει να γνωρίζουμε το χρόνο έκθεσης στον κίνδυνο του θανάτου. Ο Hatzopoulos (1997) αναφέρει ότι ο χρόνος έκθεσης διακρίνεται σε κεντρικό (central) και αρχικό (initial) χρόνο. Επίσης παρατηρεί ότι κατά τη συλλογή των δεδομένων, αυτά θα πρέπει να προέρχονται από ομοιογενή πληθυσμό που σημαίνει ότι τα άτομα έχουν παρόμοια ένταση θνησιμότητας. Αυτό απαιτείται για την επίτευξη ακριβών αποτελεσμάτων.

Ένας τρόπος για τον υπολογισμό του κεντρικού χρόνου είναι να προσθέσουμε το χρόνο

παραμονής στην έρευνα κάθε ατόμου που συμμετέχει σε αυτή. Για παράδειγμα, αν το  $i$ -οστό άτομο αρχίσει να παρατηρείται στην ηλικία  $x + a_i$  και φύγει από την έρευνα είτε λόγω θανάτου είτε για οποιονδήποτε άλλο λόγο στην ηλικία  $x + b_i$ , τότε ο χρόνος παραμονής του στην έρευνα είναι  $b_i - a_i$ . Συνεπώς ο κεντρικός χρόνος έκθεσης στον κίνδυνο του θανάτου, ο οποίος πρόκειται για τον ακριβή χρόνο έκθεσης, είναι

$$r_x^c = \sum_i (b_i - a_i).$$

Η μέθοδος αυτή ονομάζεται ευθεία (direct) και χρησιμοποιείται όταν επιλέγουμε δείγμα από τον πληθυσμό.

Μια εναλλακτική μέθοδος υπολογισμού είναι αυτή της απογραφής (census). Σύμφωνα με αυτή ισχύει

$$r_x^c = \int_0^T P_x(t) dt,$$

όπου  $T$  είναι η διάρκεια της έρευνας και  $P_x(t)$  είναι ο πληθυσμός ατόμων ηλικίας  $x$  που βρίσκονται στη ζωή τη χρονική στιγμή  $t$ . Για τον υπολογισμό του παραπάνω ολοκληρώματος μπορεί να χρησιμοποιηθεί για παράδειγμα ο κανόνας του τραπεζίου οπότε έχουμε

$$r_x^c = \frac{T}{2} (P_x(0) + P_x(T)).$$

Διαιρώντας τον αριθμό των θανάτων  $d_x$  με το χρόνο έκθεσης στον κίνδυνο  $r_x^c$ , παίρνουμε την εκτίμηση της έντασης θνησιμότητας, όπως είδαμε παραπάνω. Σε περίπτωση που ο χρόνος έκθεσης στον κίνδυνο συνεχίζει να μετράται μέχρι τη στιγμή που το άτομο θα έφευγε κανονικά από την έρευνα, ακόμη και αν αυτό έχει πεθάνει, και ο χρόνος αυτός προστίθεται στο  $r_x^c$ , τότε μιλάμε για τον αρχικό χρόνο έκθεσης. Μια ικανοποιητική προσέγγιση αυτού δίνεται από τη σχέση

$$r_x^i = r_x^c + \frac{d_x}{2}.$$

Διαιρώντας τον αριθμό των θανάτων  $d_x$  με τον χρόνο έκθεσης στον κίνδυνο  $r_x^i$ , υπολογίζουμε τον αρχικό αδρό δείκτη θνησιμότητας  $\overset{\circ}{q}_x$ .

Γνωρίζουμε ότι ένας εκτιμητής, μπορεί να γραφεί ως το άθροισμα της πραγματικής τιμής και ενός θετικού ή αρνητικού σφάλματος. Χρησιμοποιώντας το συμβολισμό του London (1985), μπορούμε να γράψουμε ότι

$$u_x = t_x + e_x,$$

όπου  $u_x$  είναι οι αρχικές εκτιμήσεις των ποσοστών θνησιμότητας ή της έντασης θνησιμότητας,  $t_x$  είναι οι πραγματικές τιμές των παραπάνω ποσοτήτων,  $e_x$  είναι τα σφάλματα και

$x = 1, 2, \dots, n$  είναι η ηλικία. Σημειώνουμε ότι μπορούμε να εξομαλύνουμε έναν ολόκληρο πίνακα θνησιμότητας ή ένα μόνο κομμάτι του. Θεωρώντας τυχαίες μεταβλητές, έχουμε

$$U_x = t_x + E_x.$$

Σημειώνουμε ότι οι πραγματικές τιμές  $t_x$  δεν είναι τυχαίες μεταβλητές.

Σύμφωνα με τον Miller (1949), ο αναλογιστής επιθυμεί τα ποσοστά ή η ένταση θνησιμότητας να μεταβάλλονται ομαλά, γιατί έτσι εξασφαλίζει ότι θα μεταβάλλονται ομαλά και οι διάφορες ασφαλιστικές ποσότητες που υπολογίζονται με βάση τις τιμές αυτές.

### 4.3 Χαρακτηριστικά της εξομάλυνσης

Δυο είναι τα βασικά χαρακτηριστικά της εξομάλυνσης, η *ομαλότητα* (smoothness) και η *καλή προσαρμογή* (goodness of fit). Τα δυο αυτά στοιχεία είναι αντικρουόμενα και η επίτευξη του ενός απαιτεί τη θυσία του άλλου (Haberman, 1998 και Greville, 1983). Ας δούμε τώρα τί εννοούμε με τους δυο παραπάνω όρους.

#### 4.3.1 Ομαλότητα

Η ομαλότητα είναι μια μαθηματική έννοια, της οποίας όμως ο ορισμός δεν είναι ξεκάθαρος. Ο Weber (1976) ορίζει μια συνάρτηση  $f(x)$  ως *ομαλή*, αν τόσο η συνάρτηση όσο και η πρώτη της παράγωγος  $f'(x)$  είναι συνεχείς. Ο Mac Lane (1986) αναφέρει ως ομαλή, μια συνάρτηση με όσο το δυνατόν περισσότερες συνεχείς παραγώγους.

Η ομαλότητα μπορεί να ελεγχθεί γραφικά εξετάζοντας αν η γραφική παράσταση των αρχικών δεδομένων δεν έχει σημαντικές διακυμάνσεις. Εναλλακτικά, υπολογίζεται ένα αριθμητικό μέτρο της ομαλότητας το οποίο βασίζεται στις διαφορές τρίτης ή τέταρτης τάξης των εξομαλυμένων τιμών. Ως μέτρο της ομαλότητας μπορεί να χρησιμοποιηθεί το άθροισμα των τετραγώνων ή των απόλυτων τιμών αυτών των διαφορών (London, 1985 και Haberman, 1998). Συνήθως χρησιμοποιείται το

$$S = \sum_x (\Delta^z v_x)^2, \quad x = 1, 2, \dots, n - z,$$

όπου  $v_x$  είναι οι εξομαλυμένες τιμές,  $\Delta^z v_x$  είναι οι διαφορές των εξομαλυμένων τιμών και συνήθως  $z = 3$  ή  $4$ . Οι εξομαλυμένες τιμές προκύπτουν από μεθόδους που παρουσιάζονται σε επόμενη ενότητα του κεφαλαίου αυτού. Σημειώνουμε ότι οι διαφορές 1ης, 2ης, 3ης και 4ης τάξης ορίζονται ως

$$\Delta v_x = v_{x+1} - v_x, \quad \Delta^2 v_x = v_x - 2v_{x+1} + v_{x+2},$$

$$\Delta^3 v_x = -v_x + 3v_{x+1} - 3v_{x+2} + v_{x+3} \quad \text{και} \quad \Delta^4 v_x = v_x - 4v_{x+1} + 6v_{x+2} - 4v_{x+3} + v_{x+4},$$

αντίστοιχα.

Εάν η τιμή του μέτρου  $S$  είναι σχετικά μικρή (κοντά στο μηδέν), συμπεραίνουμε ότι η σειρά των νέων εκτιμήσεων είναι ομαλή και συνεπώς καλύτερη από τη σειρά των αρχικών εκτιμήσεων. Βέβαια το μέτρο  $S$  από μόνο του δεν έχει καμία πρακτική σημασία. Απλά μπορεί να χρησιμοποιηθεί ως μέτρο σύγκρισης της ομαλότητας που προκύπτει από διαφορετικές εξομαλύνσεις των ίδιων δεδομένων. Σημειώνεται ότι το μέτρο  $S$  μπορεί να υπολογισθεί και για τις αρχικές τιμές  $u_x$ .

### 4.3.2 Καλή προσαρμογή

Είναι λογικό να θέλουμε οι εξομαλυμένες τιμές να μην απέχουν πολύ από τις αρχικές εκτιμήσεις. Άλλωστε η απόκλιση από τις αρχικές τιμές υπάρχει για την επίτευξη του επιθυμητού βαθμού ομαλότητας. Ο London (1985) προτείνει ως μέτρα προσαρμογής τα παρακάτω:

$$F_1 = \sum_x w_x (u_x - v_x)$$

$$F_2 = \sum_x w_x (u_x - v_x)^2$$

$$F_3 = \sum_x x w_x (u_x - v_x),$$

όπου  $v_x$  είναι οι εξομαλυμένες τιμές,  $u_x$  είναι οι αρχικές εκτιμήσεις και  $w_x$  είναι βάρη για  $x = 1, 2, \dots, n$ .

Όσο μικρότερη είναι η τιμή του μέτρου τόσο καλύτερη προσαρμογή υπάρχει. Όσον αφορά το μέτρο  $F_1$ , αυτό είναι ανεπαρκές καθώς μεγάλες θετικές ή αρνητικές διαφορές μπορούν τυχαία να αναιρεθούν και έτσι να έχουμε μικρή τιμή για το  $F_1$  αλλά κακή προσαρμογή. Για την αποφυγή αυτού του προβλήματος, συνήθως χρησιμοποιείται το μέτρο  $F_2$ . Αν στο μέτρο  $F_3$  χρησιμοποιήσουμε ως βάρη  $w_x$  τον αριθμό των ατόμων σε κίνδυνο  $l_x$ , τότε μια μικρή τιμή του μέτρου αυτού, σημαίνει ότι ο αριθμός των θανάτων θα είναι περίπου ίδιος τόσο για τα παρατηρούμενα όσο και για τα εξομαλυμένα δεδομένα.

## 4.4 Στατιστικά τεστ για την εξομάλυνση

Μετά από κάθε εξομάλυνση ποσοστών θνησιμότητας, θα πρέπει να ελέγχεται κατά πόσο οι εξομαλυμένες τιμές βρίσκονται κοντά στις αρχικές εκτιμήσεις και ικανοποιούν τους επιπλέον περιορισμούς που θέτουμε όπως για παράδειγμα μονοτονία, κυρτότητα κ.α. ανάλογα με τη φύση του προβλήματος. Για παράδειγμα, ο αναλογιστής επιθυμεί τα ποσοστά ή η



ένταση θνησιμότητας να αναπαρίστανται από μια μονότονη και πιο συγκεκριμένα αύξουσα συνάρτηση, δηλαδή οι τιμές αυτές να αυξάνονται από ηλικία σε ηλικία. Επίσης προτιμά η συνάρτηση να είναι κυρτή, δηλαδή οι τιμές να αυξάνονται πιο απότομα στις μεγάλες ηλικίες.

Οι Benjamin and Pollard (1980), παρουσιάζουν ορισμένα τεστ για τον έλεγχο της υπόθεσης ότι τα παρατηρούμενα ποσοστά θνησιμότητας προέρχονται από έναν πληθυσμό με δεδομένα ποσοστά  $q_{x0}$ ,  $x = 1, 2, \dots, n$ , που για παράδειγμα παίρνουμε από κάποιον δημοσιευμένο πίνακα θνησιμότητας. Πέρα όμως από την προσαρμογή στις δεδομένες εκτιμήσεις θα πρέπει να ελέγχονται οι αποκλίσεις για την πιθανή ύπαρξη των παρακάτω προβλημάτων:

- α) αριθμός αρκετά μεγάλων αποκλίσεων οι οποίες μπορεί να αντισταθμίζονται από άλλες μικρές αποκλίσεις,
- β) μεγάλη αθροιστική απόκλιση σε μέρος ή ολόκληρο το εύρος των ηλικιών,
- γ) πολλές θετικές ή αρνητικές αποκλίσεις (δηλαδή συσσώρευση αποκλίσεων του ίδιου προσήμου) σε όλο το εύρος των ηλικιών.

Έτσι παίρνουμε ως μηδενική υπόθεση του ελέγχου, την υπόθεση ότι τα πραγματικά ποσοστά θνησιμότητας δεν διαφέρουν σημαντικά από τα αντίστοιχα ποσοστά κάποιου πίνακα θνησιμότητας, δηλαδή  $H_0 : q_x = q_{x0}$  έναντι  $H_1 : q_x \neq q_{x0}$ .

Όλα τα τεστ που χρησιμοποιούνται για τον έλεγχο της παραπάνω υπόθεσης είναι γνωστά από τη θεωρία ελέγχου υποθέσεων.

#### 4.4.1 $X^2$ τεστ καλής προσαρμογής

Για τον έλεγχο καλής προσαρμογής χρησιμοποιείται το γνωστό  $X^2$  τεστ. Ορίζουμε ως  $l_x$  τον αριθμό των ατόμων σε κίνδυνο ηλικίας  $x$  και ως  $d_x$  τον αριθμό των θανάτων στην ίδια ηλικία. Υπό τη μηδενική υπόθεση, θεωρούμε ότι το πλήθος των θανάτων  $d_x$  ακολουθεί τη διωνυμική κατανομή με παραμέτρους  $l_x$  και  $q_{x0}$ , δηλαδή

$$d_x \sim B_i(l_x, q_{x0}), \quad x = 1, 2, \dots, n.$$

Σε περίπτωση που ο αναμενόμενος αριθμός θανάτων  $l_x q_{x0}$  είναι μεγαλύτερος του 5, τότε ασυμπτωτικά το  $d_x$  θα ακολουθεί την κανονική κατανομή με μέση τιμή  $l_x q_{x0}$  και διακύμανση  $l_x q_{x0}(1 - q_{x0})$ . Τότε υπό τη μηδενική υπόθεση, η ποσότητα

$$X^2 = \sum_{x=1}^n \frac{(d_x - l_x q_{x0})^2}{l_x q_{x0}(1 - q_{x0})}$$

ακολουθεί τη  $\chi^2$  κατανομή με  $n$  βαθμούς ελευθερίας - υπό την προϋπόθεση ότι τα  $d_x$  είναι ανεξάρτητα μεταξύ τους - όπου  $n$  είναι ο αριθμός των ηλικιών ή των διαστημάτων ηλικιών. Η μηδενική υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας  $\alpha$  αν  $X^2 > \chi_{n,1-\alpha}^2$  όπου το  $\chi_{n,1-\alpha}^2$  βρίσκεται από τους πίνακες της  $\chi^2$  κατανομής και ορίζεται από τη σχέση  $P(X^2 \geq$

$$\chi_{n,1-a}^2 = 1 - a.$$

Σημειώνουμε επίσης ότι το  $X^2$  τεστ αδυνατεί να ελέγξει τα προβλήματα (α) - (γ), για τον έλεγχο των οποίων χρησιμοποιούνται τα παρακάτω τεστ.

#### 4.4.2 Έλεγχος μεμονωμένων τυποποιημένων αποκλίσεων

Για τον έλεγχο του προβλήματος (α) εφαρμόζεται το τεστ των μεμονωμένων τυποποιημένων αποκλίσεων. Υπό τη μηδενική υπόθεση, ισχύει ότι ο αριθμός των θανάτων στην ηλικία  $x$ ,  $x = 1, 2, \dots, n$ ,  $d_x$ , ακολουθεί τη διωνυμική κατανομή η οποία μπορεί να προσεγγιστεί από τη  $N(l_x q_{x0}, l_x q_{x0}(1 - q_{x0}))$ . Θεωρώντας ότι οι θάνατοι στις διάφορες ηλικίες είναι ανεξάρτητοι μεταξύ τους και ότι δεν υπάρχει ετερογένεια σε κάθε ηλικία, οι τυποποιημένες αποκλίσεις

$$\frac{d_x - l_x q_{x0}}{\sqrt{l_x q_{x0}(1 - q_{x0})}}$$

θα πρέπει να συμπεριφέρονται ως ανεξάρτητες παρατηρήσεις που προέρχονται από την τυπική κανονική κατανομή. Σε επίπεδο σημαντικότητας, έστω  $a = 0.05$ , απορρίπτεται η μηδενική υπόθεση όταν περισσότερο από το 5% των τυποποιημένων αποκλίσεων υπερβαίνει κατά απόλυτο τιμή την τιμή 1.96.

#### 4.4.3 Έλεγχος απόλυτων τυποποιημένων αποκλίσεων

Για το ίδιο πρόβλημα μπορούν επίσης να χρησιμοποιηθούν οι απόλυτες τυποποιημένες αποκλίσεις

$$Z = \left| \frac{d_x - l_x q_{x0}}{\sqrt{l_x q_{x0}(1 - q_{x0})}} \right|, \quad x = 1, 2, \dots, n$$

Επειδή  $P(|Z| \leq \frac{2}{3}) \approx 0$ ,  $Z \sim N(0, 1)$ , τότε υπό τη μηδενική υπόθεση, ο αριθμός, έστω  $R$ , των απόλυτων τυποποιημένων αποκλίσεων που υπερβαίνουν την τιμή  $\frac{2}{3}$  θα κατανέμεται ως διωνυμική κατανομή με παραμέτρους  $n$  και  $\frac{1}{2}$ . Η μηδενική υπόθεση θα απορρίπτεται αν ο αριθμός  $R$  ανήκει στην άνω  $a$  περιοχή της διωνυμικής κατανομής. Αν ο αριθμός των ηλικιών  $n$  είναι μεγαλύτερος του 20, τότε προσεγγιστικά το  $R$  θα ακολουθεί την κανονική κατανομή με μέσο  $\frac{n}{2}$  και διακύμανση  $\frac{n}{4}$  και η μηδενική υπόθεση θα απορρίπτεται αν η ποσότητα

$$T = \frac{2R - n}{\sqrt{n}}$$

ανήκει στην άνω  $a$  περιοχή της κατανομής  $N(0, 1)$ .

#### 4.4.4 Έλεγχος αθροιστικών (συσσωρευμένων) αποκλίσεων

Για τον έλεγχο του προβλήματος ( $\beta$ ) μπορούν να χρησιμοποιηθούν οι αθροιστικές αποκλίσεις. Υποθέτοντας ότι οι θάνατοι στις γειτονικές ηλικίες είναι ανεξάρτητοι μεταξύ τους, τότε υπό τη μηδενική υπόθεση, η αθροιστική απόκλιση ενός τυχαία επιλεγμένου διαστήματος ηλικιών, έστω  $[x_1, x_2]$ , η οποία ισούται με

$$\sum_{x=x_1}^{x_2} (d_x - l_x q_{x0}),$$

θα είναι μια κανονική τυχαία μεταβλητή με μέση τιμή 0 και διακύμανση  $\sum_{x=x_1}^{x_2} l_x q_{x0} (1 - q_{x0})$ .

Η μηδενική υπόθεση απορρίπτεται αν η απόλυτη τιμή της αθροιστικής απόκλισης σε όλο το εύρος των ηλικιών είναι μεγαλύτερη από το διπλάσιο της τετραγωνικής ρίζας της αντίστοιχης διακύμανσης ή αν οι αθροιστικές αποκλίσεις σε τμήματα του εύρους των ηλικιών είναι πολύ μεγάλες κατά απόλυτη τιμή. Σημειώνεται, ότι τα διαστήματα ηλικιών δεν πρέπει να αλληλοκαλύπτονται, διαφορετικά οι αποκλίσεις θα είναι θετικά συσχετισμένες και το αναμενόμενο ποσοστό των τεστ που δίνουν σημαντικό αποτέλεσμα δεν θα ξεπερνά το 5%.

#### 4.4.5 Προσημικός έλεγχος

Το πρόβλημα ( $\gamma$ ) μπορεί να ελεγχθεί μέσω του προσημικού ελέγχου. Υπό τη μηδενική υπόθεση, τα πρόσημα των μεμονωμένων αποκλίσεων είναι ανεξάρτητα και περίπου ισοκατανομημένα σε θετικά και αρνητικά. Συνεπώς ο αριθμός  $K$  των θετικών (ή αρνητικών) προσήμων σε μια σειρά  $n + 1$  αποκλίσεων θα ακολουθεί την κατανομή  $Bi(n, \frac{1}{2})$  και η μηδενική υπόθεση απορρίπτεται αν ο αριθμός  $K$  βρίσκεται στην κάτω ή άνω  $a$  περιοχή της  $Bi(n, \frac{1}{2})$ . Αν  $n \geq 20$ , το  $K$  προσεγγίζεται από την κατανομή  $N(\frac{n}{2}, \frac{n}{4})$  και η μηδενική υπόθεση απορρίπτεται αν η ποσότητα

$$L = \frac{2K - n}{\sqrt{n}},$$

ανήκει στην κάτω ή άνω  $a$  περιοχή της τυπικής κανονικής κατανομής.

Εναλλακτικό τεστ για το ίδιο πρόβλημα είναι το τεστ του Steven για τις ομαδοποιήσεις των προσήμων, που βασίζεται στην υπεργεωμετρική κατανομή και το διωνυμικό τεστ για την αλλαγή των προσήμων. Για περισσότερες πληροφορίες για τα παραπάνω τεστ παραπέμπουμε στους Benjamin and Pollard (1980). Οι ίδιοι συγγραφείς αναφέρουν, ότι καθώς όλα τα παραπάνω τεστ είναι ευκολότερα στην εφαρμογή τους από το  $X^2$  τεστ, είναι προτιμότερο να γίνονται πριν από το  $X^2$ , ο υπολογισμός του οποίου θα πρέπει να αποφεύγεται αν τα προηγούμενα τεστ δείχνουν ανεπαρκή ή ισχυρή προσαρμογή.

Οι παραπάνω έλεγχοι μπορούν να χρησιμοποιηθούν και για τον έλεγχο της εξομάλυνσης. Στην περίπτωση αυτή, ως μηδενική υπόθεση θεωρούμε την υπόθεση ότι η μέθοδος

εξομάλυνσης που χρησιμοποιείται δίνει ικανοποιητικά αποτελέσματα. Δηλαδή ότι τα αρχικά δεδομένα δεν διαφέρουν από τα εξομαλυμένα. Έτσι στους τύπους των ελέγχων αντικαθιστούμε το  $q_{x0}$  με  $\hat{q}_x$ , όπου  $\hat{q}_x$  είναι τα εξομαλυμένα ποσοστά. Όμως επειδή στις μεθόδους αυτές εκτιμούμε κάποιες παραμέτρους, οι βαθμοί ελευθερίας του  $X^2$  τεστ δεν είναι πλέον  $n$  αλλά αφαιρούμε από αυτούς τόσους όσες είναι οι παράμετροι που εκτιμούμε. Επίσης να σημειώσουμε ότι η στατιστική συνάρτηση  $X^2$  είναι ίδια με το μέτρο  $F_2$  της Υποενότητας 4.3.2 με βάρη  $w_x = \frac{l_x}{v_x(1-v_x)}$ , όπου  $v_x$  είναι οι εξομαλυμένες τιμές.

Οι έλεγχοι που παρουσιάσαμε, μπορούν να χρησιμοποιηθούν και για την εξομάλυνση που αναφέρεται στην ένταση θνησιμότητας  $\mu_x$ . Υποθέτοντας ότι  $d_x \sim P(r_x^c \mu_x)$ , μπορούμε να θεωρήσουμε ότι ασυμπτωτικά το  $d_x$  θα ακολουθεί την κανονική κατανομή με μέση τιμή και διακύμανση ίσα με  $r_x^c \mu_x$ , όπου  $r_x^c$  είναι ο κεντρικός χρόνος έκθεσης στον κίνδυνο.

Αν ο έλεγχος  $X^2$  υποδηλώνει ισχυρή προσαρμογή, πρέπει να απορρίψουμε την εξομάλυνση καθώς τότε η ομαλότητα θα είναι πάρα πολύ μικρή (υποεξομάλυνση), αν φυσικά η εξομάλυνση έχει γίνει με μια από τις μη παραμετρικές μεθόδους που θα δούμε παρακάτω. Η εξομάλυνση επίσης θα πρέπει να απορρίπτεται αν ο παραπάνω έλεγχος δείχνει κακή προσαρμογή γιατί στην περίπτωση αυτή η ομαλότητα θα είναι πάρα πολύ μεγάλη (υπερεξομάλυνση). Όπως έχουμε ήδη αναφέρει, η ομαλότητα και η καλή προσαρμογή είναι δυο αντικρουόμενα χαρακτηριστικά της εξομάλυνσης και για να είναι αυτή ικανοποιητική, θα πρέπει και τα δυο αυτά χαρακτηριστικά να υπάρχουν σε κάποιο «συγκεκριμένο» βαθμό.

Οι υπόλοιποι έλεγχοι χρησιμοποιούνται ως έχουν. Βέβαια οι Benjamin and Pollard (1980) αναφέρουν τροποποιήσεις αυτών των ελέγχων, αντικαθιστώντας τη διακύμανση  $l_x \hat{q}_x (1 - \hat{q}_x)$  με την  $\frac{n-k}{n} l_x \hat{q}_x (1 - \hat{q}_x)$ , όπου  $\hat{q}_x$  είναι τα εξομαλυμένα ποσοστά θνησιμότητας και  $k$  ο αριθμός των παραμέτρων που εκτιμώνται, χωρίς η τροποποίηση αυτή να επιφέρει σημαντικές αλλαγές στο αποτέλεσμα των ελέγχων.

## 4.5 Κατηγορίες μεθόδων εξομάλυνσης

Για την κατασκευή ενός μοντέλου θνησιμότητας το οποίο θα αναπαριστά, δηλαδή θα εκτιμά, το πραγματικό πρότυπο θνησιμότητας, υπάρχουν δυο τρόποι. Ο πρώτος τρόπος είναι να κατασκευάσουμε έναν πίνακα θνησιμότητας και ο δεύτερος είναι να υιοθετήσουμε την άποψη ότι το πρότυπο θνησιμότητας περιγράφεται από μια συνάρτηση. Οι δυο αυτοί τρόποι μας καθοδηγούν στο να θεωρήσουμε δυο κατηγορίες για τις μεθόδους εξομάλυνσης: τις μη παραμετρικές και τις παραμετρικές (London, 1985 και Haberman, 1998).

Στις παραμετρικές μεθόδους εξομάλυνσης, σκοπός μας είναι να προσαρμόσουμε μια μαθηματική συνάρτηση στις αρχικές εκτιμήσεις των ποσοστών θνησιμότητας  $\hat{q}_x$  ή της έντασης θνησιμότητας  $\hat{\mu}_x$ , έτσι ώστε να εκφράσουμε τη σχέση που υπάρχει μεταξύ των γειτονικών τιμών. Εξομάλυνση μπορεί να γίνει είτε προσαρμόζοντας μια συνάρτηση σε όλο το εύρος των

ηλικιών είτε περισσότερες συναρτήσεις σε υποδιαστήματα των δεδομένων. Οι παράμετροι των συναρτήσεων εκτιμώνται μέσω γνωστών μεθόδων όπως της μέγιστης πιθανοφάνειας, των ελαχίστων τετραγώνων κ.α. Παρόλο που οι παραμετρικές μέθοδοι δίνουν επαρκείς εκτιμητές, περιέχουν κάποια μεροληψία, καθώς δεν υπάρχει καμία συνάρτηση που να αναπαριστά επακριβώς τις πραγματικές τιμές των πιθανοτήτων θνησιμότητας  $q_x$  ή της έντασης θνησιμότητας  $\mu_x$ .

Οι μη παραμετρικές μέθοδοι εξομάλυνσης, εφαρμόζονται κυρίως σε πινακοποιημένα δεδομένα συνδυάζοντας δεδομένα σε διάφορες τιμές της ηλικίας  $x$ ,  $x = 1, 2, \dots, n$  χωρίς να προϋποθέτουν καμία συναρτησιακή μορφή για τα  $q_x$  ή  $\mu_x$ . Και αυτές οι μέθοδοι περιέχουν μεροληψία αλλά αυτή μπορεί να ελεγχθεί μειώνοντας τη δειγματική διακύμανση. Όσον αφορά το βαθμό ομαλότητας, αυτός μπορεί να ελεγχθεί με την επιλογή της κατάλληλης τιμής κάποιων παραμέτρων ενώ στην παραμετρική εξομάλυνση, η ομαλότητα θεωρείται δεδομένη και δεν μπορεί να αλλάξει εξαιτίας των συναρτησιακών μορφών που χρησιμοποιούνται.

Σύμφωνα με τον Greville (1983), η κατηγοριοποίηση των μεθόδων εξομάλυνσης μπορεί να γίνει και με άλλους τρόπους. Έτσι υπάρχουν οι διακριτές μέθοδοι, στις οποίες κάθε παρατηρούμενη τιμή ή αρχική εκτίμηση αντικαθίσταται με μια εξομαλυμένη τιμή και οι συνεχείς μέθοδοι, όπου στα παρατηρούμενα δεδομένα προσαρμόζεται μια ομαλή καμπύλη. Ένας άλλος διαχωρισμός είναι σε τοπικές (local) και ολικές (global) μεθόδους. Στις τοπικές μεθόδους, η εξομαλυμένη τιμή εξαρτάται από ορισμένες αρχικές εκτιμήσεις ενώ στις ολικές για κάθε εξομαλυμένη τιμή χρησιμοποιούνται όλες οι παρατηρούμενες. Επιπλέον υπάρχει και ο συνδυασμός των τεσσάρων παραπάνω κατηγοριών.

Στις παραμετρικές μεθόδους ανήκουν μέθοδοι που βασίζονται σε μοντέλα θνησιμότητας (Gompertz, Makeham, Oppermann, Theile and Steffenson, Beard, Barnett, Heligman and Pollard κ.α.), σε γενικευμένα γραμμικά μοντέλα (GLM), σε συναρτήσεις splines και η smooth - junction interpolation. Στις μη παραμετρικές μεθόδους εξομάλυνσης ανήκουν η γραφική μέθοδος, οι σταθμισμένοι κινούμενοι μέσοι όροι, η μέθοδος των Whittaker and Henderson, όπου τόσο η προσαρμογή όσο και η ομαλότητα λαμβάνονται υπόψιν, η μέθοδος πυρήνα και η εξομάλυνση με αναφορά σε έναν τυπικό πίνακα θνησιμότητας (κατανομή).

Στην επόμενη παράγραφο θα ασχοληθούμε με την εξομάλυνση με χρήση της Στατιστικής Θεωρίας Πληροφοριών και πιο συγκεκριμένα μέσω των μέτρων απόκλισης. Συγκεκριμένα, θα αναφερθούμε στον τρόπο που γίνεται η εξομάλυνση μέσω του μέτρου απόκλισης των Kullback - Leibler, που πρότειναν οι Zhang and Brockett (1987) ενώ στη συνέχεια θα παρουσιάσουμε το γενικό θεωρητικό πλαίσιο για τη χρήση των μέτρων απόκλισης, το οποίο περιλαμβάνει τα μέτρα των Kullback - Leibler, των Cressie - Read και τη διαφορά του Jensen.

## 4.6 Εξομάλυνση μέσω μέτρων απόκλισης

Οι πρώτοι που χρησιμοποίησαν ένα μέτρο απόκλισης, και συγκεκριμένα το μέτρο των Kullback - Leibler, για να κατασκευάσουν μια σειρά ομαλών πιθανοτήτων θνησιμότητας  $\{v_x\}$ , η οποία να είναι όσο το δυνατόν πιο κοντά στη σειρά των παρατηρούμενων τιμών  $\{u_x\}$ , ήταν οι Zhang and Brockett (1987). Επιπλέον υπέθεσαν ότι το πραγματικό αλλά άγνωστο πρότυπο πιθανότητας είναι (i) ομαλό, (ii) αυξανόμενο με την ηλικία  $x$ , δηλαδή μονότονο, (iii) απότομα αυξανόμενο στις μεγάλες ηλικίες, δηλαδή κυρτό. Επίσης υπέθεσαν ότι (iv) ο αριθμός των θανάτων στα εξομαλυμένα δεδομένα είναι ίσος με τον αριθμό των θανάτων στα παρατηρούμενα δεδομένα και (v) το σύνολο των εξομαλυμένων ηλικιών θανάτου ισούται με τις παρατηρούμενες συνολικές ηλικίες θανάτου. Ως συνολική ηλικία θανάτου εννοούμε το άθροισμα του γινομένου του αριθμού θανάτων σε κάθε ηλικία επί την αντίστοιχη ηλικία.

Μαθηματικά οι πέντε περιορισμοί γράφονται ως εξής:

$$(i) \sum_x (\Delta^3 v_x)^2 \leq M, \text{ όπου } M \text{ είναι μια προκαθορισμένη θετική σταθερά,}$$

$$(ii) \Delta v_x \geq 0, \text{ όπου } \Delta v_x = v_{x+1} - v_x,$$

$$(iii) \Delta^2 v_x \geq 0, \text{ όπου } \Delta^2 v_x = v_x - 2v_{x+1} + v_{x+2},$$

$$(iv) \sum_x l_x v_x = \sum_x l_x u_x, \text{ όπου } l_x \text{ είναι ο αριθμός των ατόμων σε κίνδυνο στην ηλικία } x,$$

και

$$(v) \sum_x x l_x v_x = \sum_x x l_x u_x.$$

Το θεωρητικό πλαίσιο για τη χρήση των μέτρων απόκλισης στο πρόβλημα της εξομάλυνσης έχει ως εξής:

Εύρεση του διανύσματος  $\mathbf{v}$ , το οποίο ελαχιστοποιεί το μέτρο απόκλισης

$$D(\mathbf{v}, \mathbf{u})$$

υπό ορισμένους ή όλους τους ακόλουθους περιορισμούς:

$$1. \sum_{x=1}^{n-3} (\Delta^3 v_x)^2 \leq M,$$

(ομαλότητα)

$$2. \Delta v_x \geq 0,$$

(μονοτονία)

$$3. \Delta^2 v_x \geq 0,$$

(κυρτότητα)

$$4. \sum_{x=1}^n l_x v_x = \sum_{x=1}^n l_x u_x, \quad 5. \sum_{x=1}^n x l_x v_x = \sum_{x=1}^n x l_x u_x,$$

(αναλογιστικοί περιορισμοί)

όπου  $\mathbf{v}$  είναι το διάνυσμα των εξομαλυμένων τιμών,  $\mathbf{u}$  είναι το διάνυσμα των υπό εξομάλυνση (αδρών) τιμών και  $l_x$  ο αριθμός των ατόμων σε κίνδυνο στην ηλικία  $x$ .

Είναι εύκολο να δούμε ότι οι παραπάνω πέντε περιορισμοί μπορούν να γραφούν στη μορφή  $g_i(\mathbf{v}) = \frac{1}{2} \mathbf{v}^T \mathbf{D}_i \mathbf{v} + \mathbf{b}_i^T \mathbf{v} + c_i \leq 0$ ,  $i = 1, 2, \dots, r$ , όπου  $\mathbf{D}_i$  είναι ένας θετικά ημιορισμένος πίνακας για κάθε  $i$  και  $\mathbf{b}_i$ ,  $c_i$  είναι σταθερές. Είναι εύκολο να καταλάβει κάποιος ότι σε αυτή την περίπτωση έχουμε  $r = 2(n + 1)$  περιορισμούς, όπου  $n$  είναι ο αριθμός των υπό εξομάλυνση ποσοστών θνησιμότητας (Brockett, 1991).

Συνεπώς, το πρόβλημα εφαρμογής των μέτρων απόκλισης στο πρόβλημα της εξομάλυνσης μπορεί να γραφεί ως:

Εύρεση του διανύσματος  $\mathbf{v}$ , το οποίο ελαχιστοποιεί το μέτρο απόκλισης

$$D(\mathbf{v}, \mathbf{u})$$

υπό περιορισμούς της μορφής  $\mathbf{v} \geq \mathbf{0}$  και  $g_i(\mathbf{v}) = \frac{1}{2} \mathbf{v}^T \mathbf{D}_i \mathbf{v} + \mathbf{b}_i^T \mathbf{v} + c_i \leq 0$ ,  $i = 1, 2, \dots, r$ , όπου  $\mathbf{v}$  είναι το διάνυσμα των εξομαλυμένων τιμών,  $\mathbf{u}$  είναι το διάνυσμα των υπό εξομάλυνση (αδρών) τιμών και  $l_x$  ο αριθμός των ατόμων σε κίνδυνο στην ηλικία  $x$ .

Πιο συγκεκριμένα, σε μορφή πινάκων οι περιορισμοί μπορούν να γραφούν ως:

(i)  $(\mathbf{A}\mathbf{v})^T (\mathbf{A}\mathbf{v}) = \mathbf{v}^T \mathbf{A}^T \mathbf{A} \mathbf{v}$ , όπου  $\mathbf{A}$  είναι ένας  $n \times (n - 3)$  πίνακας με γραμμές της μορφής  $(0, -1, 3, -3, 1, 0, \dots, 0)$ ,

(ii)  $\mathbf{B}\mathbf{v} = \mathbf{0}$ , όπου  $\mathbf{B}$  είναι ένας  $n \times (n - 1)$  πίνακας με γραμμές της μορφής  $(0, -1, 1, 0, 0, \dots, 0)$ ,

(iii)  $\mathbf{C}\mathbf{v} = \mathbf{0}$ , όπου  $\mathbf{C}$  είναι ένας  $n \times (n - 2)$  πίνακας με γραμμές της μορφής  $(0, 1, -2, 1, 0, \dots, 0)$ ,

(iv)  $\mathbf{d}^T \mathbf{v} = \mathbf{d}^T \mathbf{u}$ , όπου  $\mathbf{d} = (l_x, l_{x+1}, \dots, l_{x+n-1})^T$ , και

(v)  $\mathbf{e}^T \mathbf{v} = \mathbf{e}^T \mathbf{u}$ , όπου  $\mathbf{e} = (x l_x, (x + 1) l_{x+1}, \dots, (x + n - 1) l_{x+n-1})^T$ , αντίστοιχα.

Για περισσότερες πληροφορίες ο ενδιαφερόμενος αναγνώστης παραπέμπεται στους Zhang and Brockett (1987), τον Sachlas (2004) καθώς επίσης και σε επόμενα κεφάλαια της παρούσας διατριβής.

Με βάση τα αποτελέσματα της θεωρητικής μελέτης των μέτρων απόκλισης, τα οποία παρουσιάστηκαν στο προηγούμενο κεφάλαιο, προτείνουμε κατά την εξομάλυνση να χρησιμοποιείται ένας επιπλέον περιορισμός (Sachlas and Papaioannou, 2008, 2009a). Συγκεκριμένα προτείνουμε τη χρήση του περιορισμού

$$(vi) \sum_{x=1}^n v_x = \sum_{x=1}^n u_x,$$

για να ισχύουν οι καλές ιδιότητες των μέτρων πληροφορίας και ιδιαίτερα η (πλήρης) μη αρνητική ιδιότητα  $D(\mathbf{v}, \mathbf{u}) \geq 0$  με ισότητα εάν και μόνο εάν  $\mathbf{v} = \mathbf{u}$  (βλ. Πρόταση 3.1 και Basu et al. (1998)). Είναι εύκολο να δούμε μέσω αντιπαραδειγμάτων ότι οι περιορισμοί (iv) και (v) δεν συνεπάγονται τον περιορισμό (vi).

Στη συνέχεια θα παρουσιάσουμε ειδικά το πλαίσιο εφαρμογής της εξομάλυνσης μέσω των μέτρων απόκλισης των Kullback - Leibler, των Cressie - Read και του Jensen.

#### 4.6.1 Εξομάλυνση μέσω του μέτρου απόκλισης των Kullback - Leibler

Για να πάρουν τις εξομαλυμένες τιμές, οι Zhang and Brockett (1987) ελαχιστοποιούν το μέτρο απόκλισης των Kullback - Leibler μεταξύ των αδρών πιθανοτήτων θανάτου  $\{u_x\}$  και των νέων πιθανοτήτων θανάτου  $\{v_x\}$ ,

$$D^{KL}(\mathbf{v}, \mathbf{u}) = \sum_{x=1}^n v_x \ln \frac{v_x}{u_x},$$

υπό τους περιορισμούς (i) - (v) θεωρώντας όμως ένα δυϊκό πρόβλημα ελαχιστοποίησης. Έτσι αντί να ελαχιστοποιήσουν το  $D^{KL}(\mathbf{v}, \mathbf{u})$  υπό τους περιορισμούς  $\mathbf{v} \geq \mathbf{0}$  και  $g_i(\mathbf{v}) = \frac{1}{2} \mathbf{v}^T \mathbf{D}_i \mathbf{v} + \mathbf{b}_i^T \mathbf{v} + c_i \leq 0$ ,  $i = 1, 2, \dots, r$ , μεγιστοποιούν το δυϊκό πρόβλημα, που είναι:

$$\text{maximize } -\mathbf{v}^T \exp \left\{ \left[ - \sum_{i=1}^r y_i (\mathbf{A}_i^T \mathbf{w}_i + b_i) \right] \right\} + \mathbf{c}^T \mathbf{y} - \frac{1}{2} \sum_{i=1}^r y_i \|\mathbf{w}_i\|^2$$

υπό τους περιορισμούς  $\mathbf{v} \geq \mathbf{0}$  και  $\mathbf{w}_i \in \mathbf{R}^{m_i}$ .

Στο παραπάνω δυϊκό πρόβλημα,  $\mathbf{A}_i$  είναι ένας  $m_i \times n$  πίνακας, ο οποίος προκύπτει από τον πίνακα  $\mathbf{D}_i$  του πρωτεύοντος προβλήματος ως  $\mathbf{D}_i = \mathbf{A}_i^T \mathbf{A}_i$ , όπου  $m_i$  είναι η τάξη του  $\mathbf{D}_i$ ,  $\mathbf{w}_i$  είναι ένα  $m_i \times 1$  διάνυσμα πραγματικών αριθμών και  $\mathbf{y}$  είναι ένα  $r \times 1$  διάνυσμα διανυσματικών πολλαπλασιαστών Lagrange.

Επιλύοντας το παραπάνω δυϊκό πρόβλημα, μπορούμε εύκολα να υπολογίσουμε τις εξομαλυμένες τιμές  $v_x^*$  χρησιμοποιώντας την ισότητα  $\ln(v_x^*/u_x^*) = \zeta_x^*$ ,  $x = 1, 2, \dots, n$  δεδομένου ότι  $\zeta^* = - \sum_{i=1}^r y_i^* (\mathbf{A}_i^T \mathbf{w}_i^* + b_i)$ . Τα  $y_i^*$  και  $\mathbf{w}_i^*$  αποτελούν τη λύση του δυϊκού προβλήματος. Με τον τρόπο αυτό μπορούμε να θεωρήσουμε την προκύπτουσα εκτίμηση ελάχιστης διαχωριστικής πληροφορίας (MDI), ως τη σειρά των εξομαλυμένων τιμών που ικανοποιούν τους περιορισμούς και είναι ελάχιστα διαχωρίσιμη από τις αδρές τιμές.

Αν και τέτοιου είδους περιορισμοί, μαθηματικοί και αναλογιστικοί, βρίσκονται πίσω από την έννοια της εξομάλυνσης, ήταν η πρώτη φορά που περιορισμοί χρησιμοποιήθηκαν σε μέθοδο εξομάλυνσης. Ήταν η πρώτη επίσης φορά που ένα μέτρο απόκλισης διαφορετικό από το γνωστό  $X^2$ , αυτό των Kullback - Leibler, χρησιμοποιήθηκε ως μέτρο καλής προσαρμογής.



## 4.6.2 Εξομάλυνση μέσω των αποκλίσεων δύναμης των Cressie - Read

Ξεκινώντας από την ιδέα του Brockett να ελαχιστοποιήσει την απόκλιση των Kullback - Leibler ώστε να βρει την καλύτερα προσαρμοσμένη σειρά εξομαλυμένων τιμών  $\{u_x\}$  υπό τους περιορισμούς (i) έως (v), στην ενότητα αυτή εξετάζουμε τη χρήση του δείκτη απόκλισης δύναμης που όρισαν οι Cressie and Read (1984).

Οι αποκλίσεις δύναμης μπορούν να χρησιμοποιηθούν για εξομάλυνση ως εξής:

Ελαχιστοποιούμε ως προς το διάνυσμα των εξομαλυμένων τιμών  $\mathbf{v}$  την απόκλιση δύναμης τάξης  $\lambda$

$$D^{CR}(\mathbf{v}, \mathbf{u}) = \frac{1}{\lambda(\lambda + 1)} \sum_{i=1}^n v_i \left[ \left( \frac{v_i}{u_i} \right)^\lambda - 1 \right],$$

υπό όλους ή μερικούς από τους περιορισμούς (i) έως (v).

Το  $\lambda$  είναι μια παράμετρος πραγματικών αριθμών. Η ελαχιστοποίηση γίνεται για διάφορες τιμές της παραμέτρου  $\lambda$  και επιλέγουμε ως καλύτερη εξομάλυνση αυτή που δίνει ικανοποιητικά αποτελέσματα για την ομαλότητα και την προσαρμογή. Με αυτόν τον τρόπο ερμηνεύουμε την προκύπτουσα σειρά των εξομαλυμένων τιμών, ως τη σειρά που ικανοποιεί τους περιορισμούς και είναι ελάχιστα διαχωρίσιμη, υπό την άποψη της απόκλισης των Cressie and Read, από τη σειρά των αδρών δεδομένων  $\{u_x\}$

Στην περίπτωση που το  $\lambda \rightarrow 0$ , έχουμε ότι

$$D_{\lambda \rightarrow 0}^{CR}(\mathbf{v}, \mathbf{u}) \equiv D^{KL}(\mathbf{v}, \mathbf{u}) = \sum_{i=1}^n v_i \ln \frac{v_i}{u_i},$$

το οποίο είναι το μέτρο απόκλισης των Kullback - Leibler. Συνεπώς, στην περίπτωση αυτή, η χρήση του μέτρου απόκλισης δύναμης με  $\lambda = 0$  συμπίπτει με τη χρήση του μέτρου των Kullback - Leibler. Με άλλα λόγια, η μέθοδος εξομάλυνσης που πρότειναν οι Zhang and Brockett (1987) αποτελεί μέλος της οικογένειας των μεθόδων εξομάλυνσης μέσω των αποκλίσεων δύναμης.

## 4.6.3 Εξομάλυνση μέσω του μέτρου απόκλισης του Jensen

Μια τρίτη επιλογή είναι η χρήση ως μέτρου απόκλισης της διαφοράς του Jensen που εισήγαγαν οι Burbea and Rao (1982). Συγκεκριμένα, έχουμε τα εξής:

Ελαχιστοποιούμε ως προς το διάνυσμα των εξομαλυμένων τιμών  $\mathbf{v}$  τη διαφορά του Jensen

$$J(\mathbf{v}, \mathbf{u}) \equiv H\left(\frac{1}{2}(\mathbf{v} + \mathbf{u})\right) - \frac{1}{2}[H(\mathbf{v}) + H(\mathbf{u})],$$

υπό όλους ή μερικούς από τους περιορισμούς (i) έως (v). Υπενθυμίζουμε ότι  $H(\cdot)$  είναι η εντροπία του Shannon.

Στην επόμενη παράγραφο, θα αναλύσουμε για ευκολότερη κατανόηση τους περιορισμούς που χρησιμοποιούνται στην εξομάλυνση.

#### 4.7 Ανάλυση των περιορισμών

Οι πέντε περιορισμοί που πρότειναν οι Brockett and Zhang, γράφονται σε μορφή πινάκων ως εξής:

(i)  $(\mathbf{A}\mathbf{v})^T(\mathbf{A}\mathbf{v}) \leq M$ , όπου

$$\mathbf{A} = \begin{bmatrix} -1 & 3 & -3 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 3 & -3 & 1 & \cdots & 0 \\ 0 & 0 & -1 & 3 & -3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

αν  $z = 3$

(ii)  $\mathbf{B}\mathbf{v} \geq 0$ , όπου

$$\mathbf{B} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

(iii)  $\mathbf{C}\mathbf{v} \geq 0$ , όπου

$$\mathbf{C} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

(iv)  $\mathbf{d}^T\mathbf{v} = \mathbf{d}^T\mathbf{u}$ , όπου

$$\mathbf{d}^T = (l_x, l_{x+1}, \dots, l_{x+n})$$

και

(v)  $\mathbf{e}^T\mathbf{v} = \mathbf{e}^T\mathbf{u}$ , όπου

$$\mathbf{e}^T = (xl_x, (x+1)l_{x+1}, \dots, (x+1)l_{x+n}).$$

Σημειώνουμε ότι οι περιορισμοί της μορφής  $\mathbf{Z} = \mathbf{W}$  γράφονται ισοδύναμα ως  $\mathbf{Z} - \mathbf{W} \geq 0$  και  $-\mathbf{Z} + \mathbf{W} \geq 0$ .

Είναι εύκολο να δούμε ότι οι παραπάνω περιορισμοί γράφονται στη μορφή  $g_i(v) = \frac{1}{2}\mathbf{v}^T \mathbf{D}_i^T \mathbf{v} + \mathbf{b}_i^T \mathbf{v} + c_i \leq 0$ .

Πιο συγκεκριμένα, έχουμε

(i)  $(\mathbf{A}\mathbf{v})^T(\mathbf{A}\mathbf{v}) \leq M \Leftrightarrow \frac{1}{2}(\mathbf{A}\mathbf{v})^T(\mathbf{A}\mathbf{v}) \leq \frac{1}{2}M \Leftrightarrow \frac{1}{2}(\mathbf{A}\mathbf{v})^T(\mathbf{A}\mathbf{v}) - \frac{1}{2}M \leq 0$ , οπότε  $\mathbf{A}_1 = \mathbf{A}$ ,  $\mathbf{b}_1 = \mathbf{0}$  και  $c_1 = -\frac{M}{2}$ .

(ii)  $\mathbf{B}\mathbf{v} \geq 0 \Leftrightarrow -\mathbf{B}\mathbf{v} \leq 0$ , οπότε  $\mathbf{A}_{2j} = \mathbf{0}$ ,  $\mathbf{b}_{2j} = \eta$   $j$ -οστή γραμμή του πίνακα  $-\mathbf{B}$  και  $c_{2j} = 0$ ,  $j = 1, 2, \dots, n-1$ .

(iii)  $\mathbf{C}\mathbf{v} \geq 0 \Leftrightarrow -\mathbf{C}\mathbf{v} \leq 0$ , οπότε  $\mathbf{A}_{3j} = \mathbf{0}$ ,  $\mathbf{b}_{3j} = \eta$   $j$ -οστή γραμμή του πίνακα  $-\mathbf{C}$  και  $c_{3j} = 0$ ,  $j = 1, 2, \dots, n-2$ .

(iv) Ο περιορισμός  $\mathbf{d}^T \mathbf{v} = \mathbf{d}^T \mathbf{u}$  γράφεται ως  $\mathbf{d}^T \mathbf{v} - \mathbf{d}^T \mathbf{u} \geq 0$  και  $-\mathbf{d}^T \mathbf{v} + \mathbf{d}^T \mathbf{u} \geq 0$ . Με άλλα λόγια ο τέταρτος περιορισμός σπάει σε δυο περιορισμούς.

Για τον περιορισμό  $\mathbf{d}^T \mathbf{v} - \mathbf{d}^T \mathbf{u} \geq 0$ , έχουμε  $-\mathbf{d}^T \mathbf{v} + \mathbf{d}^T \mathbf{u} \leq 0$ , οπότε  $\mathbf{A}_{4.1} = \mathbf{0}$ ,  $\mathbf{b}_{4.1} = -\mathbf{d}^T$  και  $c_{4.1} = \mathbf{d}^T \mathbf{u}$ .

Για τον περιορισμό  $-\mathbf{d}^T \mathbf{v} + \mathbf{d}^T \mathbf{u} \geq 0$ , έχουμε  $\mathbf{d}^T \mathbf{v} - \mathbf{d}^T \mathbf{u} \leq 0$ , οπότε  $\mathbf{A}_{4.2} = \mathbf{0}$ ,  $\mathbf{b}_{4.2} = \mathbf{d}^T$  και  $c_{4.2} = -\mathbf{d}^T \mathbf{u}$ .

(v) Ο περιορισμός  $\mathbf{e}^T \mathbf{v} = \mathbf{e}^T \mathbf{u}$  επίσης γράφεται ως  $\mathbf{e}^T \mathbf{v} - \mathbf{e}^T \mathbf{u} \geq 0$  και  $-\mathbf{e}^T \mathbf{v} + \mathbf{e}^T \mathbf{u} \geq 0$ . Με άλλα λόγια και ο πέμπτος περιορισμός σπάει σε δυο περιορισμούς.

Για τον περιορισμό  $\mathbf{e}^T \mathbf{v} - \mathbf{e}^T \mathbf{u} \geq 0$ , έχουμε  $-\mathbf{e}^T \mathbf{v} + \mathbf{e}^T \mathbf{u} \leq 0$ , οπότε  $\mathbf{A}_{5.1} = \mathbf{0}$ ,  $\mathbf{b}_{5.1} = -\mathbf{e}^T$  και  $c_{5.1} = \mathbf{e}^T \mathbf{u}$ .

Για τον περιορισμό  $-\mathbf{e}^T \mathbf{v} + \mathbf{e}^T \mathbf{u} \geq 0$ , έχουμε  $\mathbf{e}^T \mathbf{v} - \mathbf{e}^T \mathbf{u} \leq 0$ , οπότε  $\mathbf{A}_{5.2} = \mathbf{0}$ ,  $\mathbf{b}_{5.2} = \mathbf{e}^T$  και  $c_{5.2} = -\mathbf{e}^T \mathbf{u}$ .

Σύμφωνα με τα παραπάνω, παρατηρούμε ότι εάν έχουμε  $n$  παρατηρήσεις, οι περιορισμοί που εμπλέκονται στο πρόβλημα της ελαχιστοποίησης του μέτρου απόκλισης, είτε αυτό είναι το μέτρο των Kullback - Leibler είτε το μέτρο των Cressie and Read είτε η διαφορά του Jensen, είναι  $2(n+1)$ , αριθμός μεγάλος ακόμη και για μικρό  $n$ . Για παράδειγμα, εάν  $n = 15$  οι περιορισμοί που προκύπτουν είναι 32. Το γεγονός αυτό, δυσκολεύει το πρόβλημα της ελαχιστοποίησης του μέτρου απόκλισης και για το λόγο αυτό είναι προτιμότερο να εργαστούμε με το δυϊκό πρόβλημα, όπως θα δούμε στο επόμενο κεφάλαιο.

Με βάση τα παραπάνω, το λαγκρανζιανό δυϊκό πρόβλημα γίνεται

$$(D) \quad \sup_{\lambda \in \mathbb{R}_+^m, \mathbf{y}_i \in \mathbb{R}^{n_i}} \left\{ \frac{1}{(r+1)^2} \sum_{j=1}^n d_j \left( \frac{1}{r+1} - r\mathbf{z}_j \right)^{\frac{1}{r}} [r(r+1)\mathbf{z}_j - 1] - \frac{1}{2} \frac{\|\mathbf{y}_1\|^2}{\lambda_1} + \lambda^T \mathbf{c} \right\},$$

όπου  $\mathbf{z}_j = \left( \mathbf{y}_1^T \mathbf{A}_1 + \sum_{i=2}^m \lambda_i \mathbf{b}_i^T \right)_j$  και ο δείκτης  $j$  δηλώνει το  $j$ -οστό στοιχείο του διανύσματος  $\mathbf{z}$ .  $\mathbf{y}_1$  είναι ένα  $n_1 \times 1$  διάνυσμα,  $\lambda$  είναι ένα  $2(n+1) \times 1$  διάνυσμα και  $\mathbf{c}^T =$

$(-\frac{M}{2}, \underbrace{0, \dots, 0}_{n-1}, \underbrace{0, \dots, 0}_{n-2}, \mathbf{d}^T \mathbf{u}, -\mathbf{d}^T \mathbf{u}, \mathbf{e}^T \mathbf{u}, -\mathbf{e}^T \mathbf{u})$ . Η διάσταση  $n_i$  του διανύσματος  $\mathbf{y}_1$  ισούται με  $n - z$ , όπου  $z = 2, 3$  ή  $4$  ανάλογα με την τάξη των διαφορών που θα χρησιμοποιήσουμε στον πρώτο περιορισμό.

Ο περιορισμός που εμείς προτείνουμε ως απαραίτητο κατά την ελαχιστοποίηση των αποκλίσεων, είτε αυτή είναι η απόκλιση των Kullback - Leibler είτε η απόκλιση δύναμης των Cressie and Read είτε η διαφορά του Jensen,

$$(vi) \sum_{x=1}^n v_x = \sum_{x=1}^n u_x$$

μπορεί επίσης να γραφεί υπό μορφή πινάκων και συνεπώς στη μορφή  $g_i(\mathbf{v})$  ως  $\mathbf{1}^T \mathbf{v} = \mathbf{1}^T \mathbf{u}$ , όπου  $\mathbf{1} = (1, 1, \dots, 1)^T$ .

Σε αντιστοιχία με τους πέντε περιορισμούς του Brockett, έχουμε τα παρακάτω:

(vi) Ο περιορισμός  $\mathbf{1}^T \mathbf{v} = \mathbf{1}^T \mathbf{u}$  επίσης γράφεται ως  $\mathbf{1}^T \mathbf{v} - \mathbf{1}^T \mathbf{u} \geq 0$  και  $-\mathbf{1}^T \mathbf{v} + \mathbf{1}^T \mathbf{u} \geq 0$ . Με άλλα λόγια και ο έκτος περιορισμός σπάει σε δυο περιορισμούς.

Για τον περιορισμό  $\mathbf{1}^T \mathbf{v} - \mathbf{1}^T \mathbf{u} \geq 0$ , έχουμε  $-\mathbf{1}^T \mathbf{v} + \mathbf{1}^T \mathbf{u} \leq 0$ , οπότε  $\mathbf{A}_{6.1} = \mathbf{0}$ ,  $\mathbf{b}_{6.1} = -\mathbf{1}^T$  και  $c_{6.1} = \mathbf{1}^T \mathbf{u} = \sum_x u_x$ .

Για τον περιορισμό  $-\mathbf{1}^T \mathbf{v} + \mathbf{1}^T \mathbf{u} \geq 0$ , έχουμε  $\mathbf{1}^T \mathbf{v} - \mathbf{1}^T \mathbf{u} \leq 0$ , οπότε  $\mathbf{A}_{6.2} = \mathbf{0}$ ,  $\mathbf{b}_{6.2} = \mathbf{1}^T$  και  $c_{6.2} = -\mathbf{1}^T \mathbf{u} = -\sum_x u_x$ .

Συνεπώς, η ενσωμάτωση του περιορισμού (vi) στο πρόβλημα της ελαχιστοποίησης του μέτρου απόκλισης, ισοδυναμεί με την αύξηση των περιορισμών του πρωτεύοντος προβλήματος (P) κατά δυο.

## 4.8 Συμπεράσματα Κεφαλαίου

Στο κεφάλαιο αυτό έγινε μια σύντομη περιγραφή της εξομάλυνσης ποσοστών θνησιμότητας. Συγκεκριμένα περιγράψαμε τα βασικά χαρακτηριστικά της (την ομαλότητα και την καλή προσαρμογή), αναφέραμε τους στατιστικούς ελέγχους που χρησιμοποιούνται για τον έλεγχο της εξομάλυνσης ενώ επίσης αναφέραμε εν συντομία τις μεθόδους εξομάλυνσης που έχουν παρουσιαστεί στη βιβλιογραφία. Στο τέλος του κεφαλαίου παρουσιάσαμε το γενικό πλαίσιο χρήσης των μέτρων απόκλισης καθώς επίσης και το ειδικό πλαίσιο χρήσης καθενός από τα τρία μέτρα απόκλισης με τα οποία ασχολούμαστε στη διατριβή αυτή. Επιπλέον αναλύσαμε τους περιορισμούς που χρησιμοποιούνται κατά την ελαχιστοποίηση των μέτρων απόκλισης παρουσιάζοντας τους πίνακες και τα διανύσματα που εμπλέκονται σε αυτούς.

## ΚΕΦΑΛΑΙΟ 5

# ΕΛΑΧΙΣΤΟΠΟΙΗΣΗ ΜΕΤΡΩΝ ΑΠΟΚΛΙΣΗΣ ΚΑΙ ΔΥΙΚΟΤΗΤΑ

### 5.1 Εισαγωγή

Τα μέτρα αποκλίσεων παρουσιάζονται σχεδόν παντού στα μαθηματικά, τις πιθανότητες και τη στατιστική. Εκφράζουν την «απόσταση» μεταξύ δυο συναρτήσεων ή διανυσμάτων. Υπάρχουν πολλά μέτρα απόκλισης ή diversity στη βιβλιογραφία. Τα πιο γνωστά είναι αυτά των Kullback-Leibler, Csiszar ή αλλιώς  $\phi$ -απόκλιση και Cressie and Read διαφορετικά γνωστό ως απόκλιση δύναμης. Ένα λιγότερο γνωστό μέτρο απόκλισης είναι η διαφορά του Jensen.

Τα μέτρα απόκλισης δεν είναι μετρικές (metrics) καθώς δεν ικανοποιούν την συμμετρική (symmetric) ιδιότητα και για το λόγο αυτό μερικές φορές ονομάζονται μέτρα κατευθυνόμενης απόκλισης.

Υπάρχουν περιπτώσεις όπου μέτρα απόκλισης εμπλέκουν μη πιθανοτικές κατανομές, είτε συναρτήσεις πυκνότητας πιθανότητας είτε διανύσματα πιθανότητας. Μια τέτοια περίπτωση είναι η εξομάλυνση ποσοστών θνησιμότητας, η οποία είναι ένα πρόβλημα ελαχιστοποίησης που εμπλέκει αποκλίσεις.

Η ελαχιστοποίηση μιας μετρικής, μιας απόστασης ή μιας απόκλισης έχει μια κυρίαρχη θέση στη Στατιστική Επιστήμη. Εμφανίζεται σχεδόν παντού και αυτό επιβεβαιώνει την ισχυρή σχέση μεταξύ της Στατιστικής, του Μαθηματικού Προγραμματισμού και της Επιχειρησιακής Έρευνας. Μερικά παραδείγματα είναι τα ελάχιστα τετράγωνα, οι βελτιστοποιήσεις  $L^1$ ,  $L^2$ ,  $L^\infty$ , η ελάχιστη διαχωριστική πληροφορία (minimum discrimination information), το κριτήριο AIC του Akaike, η εκτίμηση ελάχιστης απόκλισης (minimum divergence estimation), η εκτίμηση απόστασης Hellinger (Hellinger distance estimation), η εκτίμηση της εκ των προτέρων συνάρτησης του Bayes (Bayes prior estimation).

Όλα αυτά τα προβλήματα εμπλέκουν αποκλίσεις με κατανομές πιθανότητας, είτε συναρτήσεις πυκνότητας πιθανότητας είτε πεπερασμένα/άπειρα διανύσματα πιθανότητας. Το

σύνηθες πλαίσιο (set up) είναι το ακόλουθο: Η  $g_0$  είναι γνωστή ή εμπειρική (δηλαδή βασισμένη σε δεδομένα) κατανομή και ψάχνουμε την  $f$  που ελαχιστοποιεί την κατευθυνόμενη απόκλιση  $D(f, g_0)$  της  $f$  από την  $g_0$  υπό κάποιους απαραίτητους περιορισμούς για την  $f$  γιατί διαφορετικά η λύση είναι η  $f = g_0$ . Υπάρχουν περιπτώσεις όπου το πλαίσιο αντιστρέφεται: η  $f$  είναι γνωστή, βασισμένη σε δεδομένα ή εκτιμημένη, δηλαδή  $f = f_0$ , η  $g$  είναι άγνωστη και το αντικείμενο είναι η ελαχιστοποίηση της  $D(f_0, g)$  με τη  $g$  να ικανοποιεί κάποιους περιορισμούς.

Στο κεφάλαιο αυτό θα περιγράψουμε τρόπους ελαχιστοποίησης των μέτρων απόκλισης υπό διάφορους περιορισμούς. Ειδική βαρύτητα δίνουμε στην περίπτωση της ελαχιστοποίησης των μέτρων απόκλισης υπό τετραγωνικούς περιορισμούς, καθώς τέτοιοι περιορισμοί παρουσιάζονται στο πρόβλημα της εξομάλυνσης. Η ελαχιστοποίηση των μέτρων γίνεται μέσω της Θεωρίας της Δυϊκότητας (Duality Theory). Στο τέλος του κεφαλαίου παρουσιάζουμε τις ακριβείς λύσεις των προβλημάτων ελαχιστοποίησης των μέτρων των Kullback - Leibler, Cressie and Read και Jensen υπό τετραγωνικούς περιορισμούς.

## 5.2 Ελαχιστοποίηση αποκλίσεων

Η ελαχιστοποίηση των αποκλίσεων ακολουθεί δυο πρότυπα. Το ένα μπορεί να ονομαστεί *παραμετρικό* και το άλλο *μη παραμετρικό*.

Για το πρώτο θεωρούμε την  $f(x)$  γνωστή και την  $g(x, \theta) \equiv g_\theta$  γνωστή αλλά εξαρτώμενη από μια άγνωστη παράμετρο  $\theta$ , όπου το  $\theta$  μπορεί να είναι μονοδιάστατη παράμετρος ή διάνυσμα. Τότε προσπαθούμε να εκτιμήσουμε την παράμετρο  $\theta$  ελαχιστοποιώντας, με ή χωρίς περιορισμούς, την

$$D(f, g_\theta).$$

Αυτό οδηγεί σε εκτίμηση ελάχιστης απόκλισης ή απόστασης. Για παράδειγμα, εάν  $D(f, g_\theta)$  είναι η απόκλιση των Kullback-Leibler και πάρουμε την  $f$  να είναι η  $\hat{f}$ , δηλαδή εμπειρική ή βασισμένη σε δεδομένα, τότε έχουμε να ελαχιστοποιήσουμε ως προς την παράμετρο  $\theta$ , το

$$\int \hat{f}(x) \ln \left[ \hat{f}(x)/g(x, \theta) \right] d\mu,$$

όπου  $\mu$  είναι ένα κατάλληλο μέτρο. Με αυτό τον τρόπο παίρνουμε εκτιμητές ελάχιστης απόκλισης Kullback-Leibler, οι οποίοι συμπίπτουν (a.s.) με τους εκτιμητές μέγιστης πιθανοφάνειας για διακριτά (π.χ. πολυωνυμικά) μοντέλα. Εάν η απόκλιση είναι η απόσταση του Kolmogorov ή η μετρική  $L^\infty$ , τότε πάλι προσπαθούμε να ελαχιστοποιήσουμε ως προς το  $\theta$ , την

$$\sup |F_n(x) - F(x, \theta)|,$$

όπου  $F$  είναι η αθροιστική συνάρτηση κατανομής (cdf) της  $X$  και  $F_n$  η αντίστοιχη εμπει-

ρική αθροιστική συνάρτηση κατανομής. Ένα κλασικό παράδειγμα εκτίμησης ελάχιστης απόκλισης είναι η εκτίμηση ελαχίστων τετραγώνων στα γραμμικά μοντέλα, η οποία έχει αντιμετωπιστεί στη βιβλιογραφία τόσο αλγεβρικά όσο και γεωμετρικά. Το  $\chi^2$ , η απόσταση Hellinger, η απόκλιση  $\phi$  του Csiszar και οι αποκλίσεις δύναμης των Cressie and Read έχουν χρησιμοποιηθεί για εκτίμηση ελάχιστης απόστασης ή απόκλισης της παραμέτρου  $\theta$ .

Εδώ μπορεί κανείς να παρατηρήσει τη χρήση της «απόκλισης των δεδομένων από το μοντέλο». Δεν υπάρχει συγκεκριμένος λόγος να προσπαθήσουμε να ελαχιστοποιήσουμε την  $D(\hat{f}, g_\theta)$  ως προς το  $\theta$ . Μπορούμε να ελαχιστοποιήσουμε την  $D(g_\theta, \hat{f})$  και αυτό οδηγεί σε διαφορετικούς εκτιμητές για το  $\theta$ . Εκτιμητές τέτοιου τύπου δεν έχουν ερευνηθεί στη βιβλιογραφία. Για αναφορές στο θέμα αυτό, ο ενδιαφερόμενος αναγνώστης παραπέμπεται στους Read and Cressie (1988), Vos (1992), Cutler and Cordero-Brana (1996) και Pardo (2006).

Για το μη παραμετρικό πρόβλημα έχουμε την ελαχιστοποίηση της  $D(f, g)$  με περιορισμούς για την  $f$  με σκοπό τον προσδιορισμό της  $f$ , η οποία είναι όσο το δυνατόν πιο κοντά στη γνωστή  $g$ . Εάν  $g = \hat{g}$ , δηλαδή η εκτίμηση βασίζεται σε δεδομένα, τότε προσπαθούμε να εκτιμήσουμε το στοχαστικό μοντέλο που είναι πιο κοντά στα δεδομένα. Κλασικό παράδειγμα είναι η ελάχιστη διαχωριστική κατανομή, η οποία παράγεται από την ελαχιστοποίηση της απόκλισης των Kullback-Leibler

$$\int f \ln(f/g) dx,$$

με περιορισμούς στην  $f$  και οδηγεί στην ελάχιστη διαχωριστική πληροφορία (minimum discrimination information).

Η λύση του Kullback παράγεται μέσω Λαγκρανζιανής Θεωρίας και λογισμού μεταβολών (calculus of variation) ενώ στη διακριτή περίπτωση, μπορούμε να χρησιμοποιήσουμε αποτελέσματα μαθηματικού προγραμματισμού. Η βέλτιστη λύση  $f^*$  είναι εκθετική.

Στην επιλογή μοντέλων, μέσω προσαρμογής διάφορων μοντέλων, επιδιώκουμε να εκτιμήσουμε το μοντέλο  $\hat{g}$ , το οποίο είναι όσο το δυνατόν πιο κοντά στην αληθινή υποτιθέμενη κατανομή  $f$  των δεδομένων. Για το λόγο αυτό ελαχιστοποιούμε την

$$D(f, \hat{g}) \text{ ως προς } \hat{g}.$$

Ένα κλασικό παράδειγμα είναι το κριτήριο πληροφορίας του Akaike AIC (Akaike, 1973).

Ένα άλλο σημαντικό παράδειγμα χρήσης των αποκλίσεων στη Στατιστική είναι μέσω της έννοιας των *στατιστικών συναρτήσεων πληροφορίας ή απόκλισης*: Έχουμε τις συνηθισμένες στατιστικές συναρτήσεις αποκλίσεων  $D(\hat{f}, g)$  με την  $g$  γνωστή ή την  $D(f, \hat{g})$  με την  $f$  γνωστή ή την  $D(\hat{f}, \hat{g})$ . Επιπλέον έχουμε στατιστικές συναρτήσεις ελάχιστης διαχωριστικής πληροφορίας [Kullback, (1959, σελ. 82, 85)], οι οποίες προκύπτουν από διάφορες κλάσεις μοντέλων πιθανότητας. Για πολλά προβλήματα οι δυο προσεγγίσεις συμπίπτουν. Υπάρχει

μεγάλη βιβλιογραφία για το θέμα και οι αποκλίσεις παρέχουν ένα εργαλείο για τη διαμόρφωση στατιστικών συναρτήσεων για ελέγχους υποθέσεων, διαστήματα εμπιστοσύνης, για τη σύγκριση και επιλογή μοντέλων και γενικότερα τη διαμόρφωση μιας εφαρμοσμένης στατιστικής συμπερασματολογίας. Για μια σημαντική πρόσφατη αναφορά στο θέμα, ο ενδιαφερόμενος παραπέμπεται στον Pardo (2006).

### 5.3 Ελαχιστοποίηση αποκλίσεων στην Αναλογιστική Επιστήμη

Όπως είδαμε στο προηγούμενο κεφάλαιο, τα μέτρα πληροφορίας παίζουν σημαντικό ρόλο στην Αναλογιστική Επιστήμη. Για την κανονική κατανομή, το μέτρο του Fisher, που είναι το αντίστροφο της διακύμανσης, είναι ένα λογικό μέτρο της αβεβαιότητας. Σύμφωνα με τον Brockett (1991) τα αναλογιστικά μοντέλα έχουν γίνει περίπλοκα και διάφοροι περιορισμοί σχετικά με την κανονικότητα έχουν εισέλθει σε αυτά. Για το λόγο αυτό έχει παρατηρηθεί μια απομάκρυνση από τα βασισμένα στη διακύμανση και την κανονική κατανομή μέτρα, όπως είναι αυτό του Fisher. Από το 1948 μάλιστα ο Wiener παρατήρησε ότι η εντροπία του Shannon θα αντικαθιστούσε το μέτρο του Fisher.

Στην Αναλογιστική Επιστήμη, βασικός στόχος μας είναι να επιλέξουμε μεταξύ δυο υποψήφιων κατανομών πιθανότητας  $p$  και  $q$  για ένα στοχαστικό μοντέλο, με βάση την πληροφορία που συλλέγουμε από τις παρατηρήσεις μιας τυχαίας μεταβλητής  $X$  που περιγράφει το μοντέλο. Το βασικότερο λοιπόν μέτρο απόστασης μεταξύ δυο κατανομών είναι το μέτρο των Kullback - Leibler, το οποίο δίνεται από τον τύπο

$$I^{KL}(\mathbf{p}^*, \mathbf{q}^*) = \int_{-\infty}^{+\infty} p^*(x) \ln \frac{p^*(x)}{q^*(x)} dx$$

για τη συνεχή περίπτωση, και

$$I^{KL}(\mathbf{p}^*, \mathbf{q}^*) = \sum_i p_i \ln \frac{p_i}{q_i}$$

για τη διακριτή περίπτωση. Το μέτρο αυτό είναι η μέση τιμή του λογαρίθμου του λόγου πιθανοφανειών (odds - ratio), ο οποίος είναι επαρκής στατιστική συνάρτηση, και δίνει, κατά μια ερμηνεία, την αναμενόμενη πληροφορία που υπάρχει σε μια παρατήρηση μιας τυχαίας μεταβλητής  $X$  για το διαχωρισμό μεταξύ των δυο υποψήφιων κατανομών πιθανότητας για το στοχαστικό μοντέλο.

Εάν κάποια χαρακτηριστικά των κατανομών, όπως ροπές, που συμβολίζουν καθαρά ασφάλιστρα, μέσο χρόνο ζωής κλπ, ποσοστημόρια, για παράδειγμα αναμενόμενα ποσοστά θνησιμότητας ή πιθανότητες επιβίωσης, ή άλλα που μπορούν να εκφραστούν ως αναμενόμενες τιμές, είναι ή υποτίθεται ότι είναι γνωστά, τότε αυτά μπορούν να χρησιμοποιηθούν ως



περιορισμοί στην ανάλυσή μας.

## 5.4 Το κανονικό MDI πρόβλημα

Επειδή βασικός μας στόχος - στην Αναλογιστική Επιστήμη - είναι να βρούμε την κατανομή  $\mathbf{p}^*$ , η οποία θα είναι όσο το δυνατόν πιο όμοια με την  $\mathbf{q}^*$ , θα πρέπει να ελαχιστοποιήσουμε ως προς  $\mathbf{p}^*$  το μέτρο  $I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$ . Η ελάχιστη τιμή ονομάζεται Ελάχιστη Διαχωριστική Πληροφορία (Minimum Discriminant Information, MDI), την οποία πρώτος εισήγαγε ο Kullback το 1954.

Επειδή, όπως είδαμε από το Θεώρημα 2.1, ισχύει  $I^{KL}(\mathbf{p}^*, \mathbf{q}^*) \geq 0$  με ισότητα αν και μόνο αν  $p^*(x) = q^*(x)$ , τότε η κατανομή που δίνει η, υπό ορισμένους γραμμικούς περιορισμούς, ελαχιστοποίηση του παραπάνω μέτρου, δηλαδή η MDI εκτίμηση  $\hat{\mathbf{p}}^*$  είναι όσο το δυνατόν πιο κοντά (ελάχιστα διαχωρίσιμη) στην  $\mathbf{q}^*$  και ταυτόχρονα ικανοποιεί τους περιορισμούς που ίσως η  $\mathbf{q}^*$  να μην ικανοποιεί (Brockett, 1991).

Στη συνεχή περίπτωση έχουμε την ελαχιστοποίηση του μέτρου

$$I^{KL}(p^*, q^*) = \int_{-\infty}^{\infty} p^*(x) \ln \left( \frac{p^*(x)}{q^*(x)} \right) dx,$$

υπό τους περιορισμούς

$$\int_{-\infty}^{\infty} p^*(x) dx = \theta_0 = 1$$

και

$$\int_{-\infty}^{\infty} a_j(x) p^*(x) dx = \theta_j, \quad j = 1, 2, \dots, m.$$

Με την επιλογή  $a_j(x) = x^j$ , εισάγουμε στο πρόβλημα της ελαχιστοποίησης περιορισμούς της μορφής ροπών, όπως για παράδειγμα μέσες τιμές και διακυμάνσεις. Επίσης μπορούμε να ορίσουμε την  $a_j(x)$  ως τη δείκτρια συνάρτηση ενός διαστήματος, οπότε εισάγουμε στο πρόβλημα περιορισμούς που εμπλέκουν πιθανότητες (για παράδειγμα πιθανότητες επιβίωσης ή stop loss values).

Στη διακριτή περίπτωση, ελαχιστοποιούμε το μέτρο

$$I^{KL}(\mathbf{p}^*, \mathbf{q}^*) = \sum_{i=1}^n p_i^* \ln \left( \frac{p_i^*}{q_i^*} \right),$$

όπου  $\mathbf{q}^* = (q_1^*, \dots, q_n^*)^T$  είναι η τυπική διακριτή κατανομή πιθανότητας και  $\mathbf{p}^* = (p_1^*, \dots, p_n^*)^T$  η διακριτή κατανομή με την οποία θέλουμε να αντικαταστήσουμε την  $\mathbf{q}^*$ . Η ελαχιστοποίηση

του μέτρου των Kullback - Leibler γίνεται υπό τους περιορισμούς

$$\sum_{i=1}^n p_i^* = 1,$$

$$\sum_j a_{ij} p_i^* = j_j, \quad j = 1, 2, \dots, m,$$

και

$$p_i^* \geq 0, \quad i = 1, 2, \dots, n.$$

Υπό μορφή πινάκων, οι περιορισμοί γράφονται ως  $\mathbf{A}\mathbf{p}^* = \boldsymbol{\theta}$  και  $\mathbf{p}^* \geq \mathbf{0}$ , όπου  $\mathbf{A}$  είναι ένας  $(m+1) \times n$  πίνακας όπου στην πρώτη γραμμή έχει όλο μονάδες και  $\theta_0 = 1$  για να ικανοποιείται ο περιορισμός  $\sum_{i=1}^n p_i^* = 1$ , το οποίο πιστοποιεί ότι το  $\mathbf{p}^*$  είναι κατανομή πιθανότητας. Βέβαια όπως αναφέρει ο Brockett (1991), όταν έχουμε πιθανότητες θνησιμότητας (mortality rates), οι οποίες δεν αθροίζουν στη μονάδα, για παράδειγμα όταν δεν χρησιμοποιούμε ολόκληρο τον πίνακα αλλά ένα κομμάτι του, τότε μπορούμε να διαγράψουμε τον πρώτο περιορισμό χωρίς να δημιουργηθεί κανένα πρόβλημα στην ανάλυσή μας.

Σημειώνουμε ότι στο παραπάνω πρόβλημα ελαχιστοποίησης, οι πιθανότητες  $q_i^*$  και οι τιμές των παραμέτρων  $\theta_j$  είναι σταθερές. Οι τιμές τόσο των  $q_i^*$  όσο και των  $\theta_j$  μπορεί να είναι εξωγενώς προσδιορισμένες ή προσδιορισμένες μέσω του δείγματος. Η ελαχιστοποίηση γίνεται για τον έλεγχο της υπόθεσης  $H_0 : \mathbf{p}^* = \mathbf{q}^*$ , δηλαδή ότι οι παρατηρούμενες και αναμενόμενες κατανομές πιθανότητας δεν είναι στατιστικά διαχωρίσιμες.

Έστω ότι  $\tilde{\mathbf{p}}^*$  είναι το βέλτιστο διάνυσμα οπότε  $I^{KL}(\tilde{\mathbf{p}}^*, \mathbf{q}^*)$  είναι η ελάχιστη υπό τους περιορισμούς τιμή του προβλήματος. Εάν θεωρήσουμε ότι η παρατηρούμενη κατανομή πιθανότητας προέρχεται από ένα τυχαίο δείγμα μεγέθους  $N$  από την κατανομή πιθανότητας  $\mathbf{q}^*$ , τότε αποδεικνύεται ότι η ποσότητα  $2NI^{KL}(\tilde{\mathbf{p}}^*, \mathbf{q}^*)$  ασυμπτωτικά ακολουθεί  $\chi^2$  κατανομή με βαθμούς ελευθερίας που εξαρτώνται από το μέγεθος των διανυσμάτων  $n$  και τον αριθμό των γραμμικώς ανεξάρτητων περιορισμών (Kullback, 1959).

## 5.5 Αρχή της Μέγιστης Εντροπίας

Σε πολλές περιπτώσεις δεν υπάρχει γνωστή κατανομή πιθανότητας  $\mathbf{q}^*$  από την οποία να υπολογίσουμε την  $\mathbf{p}^*$ . Η πιο λογική πρακτική που ακολουθούμε είναι να θεωρήσουμε ότι όλες οι τιμές της  $\mathbf{q}^*$  είναι ισοπίθανες. Δηλαδή  $q_i^* = \frac{1}{n}$  για όλα τα  $i$  στη διακριτή περίπτωση. Με άλλα λόγια, θεωρούμε ως τυπική κατανομή την ομοιόμορφη. Συνεπώς είναι εύκολο να δούμε ότι στην περίπτωση αυτή ισχύει

$$I^{KL}(\mathbf{p}^*, \mathbf{q}^*) = -H(\mathbf{p}^*) + \ln n,$$

και συνεπώς η ελαχιστοποίηση του μέτρου  $I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$  είναι ισοδύναμη με τη μεγιστοποίηση της εντροπίας, δηλαδή ισχύει ότι

$$\min \sum_{i=1}^n p_i^* \ln \left( \frac{p_i^*}{q_i^*} \right) \Leftrightarrow \max \sum_i p_i^* \ln \frac{1}{p_i^*} \quad \text{για} \quad q_i = \frac{1}{n}.$$

Η ποσότητα  $\sum_i p_i^* \ln p_i^*$  είναι η αρνητική τιμή της εντροπίας της κατανομής  $\mathbf{p}^*$  και η κατανομή που λύνει την παραπάνω σχέση, ονομάζεται Κατανομή Μέγιστης Εντροπίας (Maximum Entropy, ME, Distribution). Αντιλαμβανόμαστε ότι, στην περίπτωση αυτή, η ελαχιστοποίηση του  $I^{KL}(\mathbf{p}^*|\mathbf{q}^*)$  είναι ισοδύναμη με τη μεγιστοποίηση της εντροπίας  $H(\mathbf{p}^*)$ . Παρατηρούμε, ότι η εντροπία είναι ειδική περίπτωση του μέτρου των Kullback - Leibler και μετρά την απόσταση μιας κατανομής  $\mathbf{p}^*$  από την ομοιόμορφη κατανομή. Το μοντέλο της μέγιστης εντροπίας, μπορεί να θεωρηθεί ότι επιλέγει την πιο «αβέβαια» κατανομή υπό τους περιορισμούς (Brockett, 1991).

Ο Brockett (1991) αναφέρει ότι το μοντέλο *MDI* είναι γενικότερη και περισσότερο εύκαμπτη έννοια, αφού η συνάρτηση  $\mathbf{q}^*$  μπορεί να παριστάνει οποιαδήποτε κατανομή και όχι μόνο την ομοιόμορφη όπως συμβαίνει στο *ME* μοντέλο. Επίσης το *MDI* μοντέλο έχει σοβαρό στατιστικό υπόβαθρο σε αντίθεση με την εντροπία που προέρχεται από την αντίστοιχη έννοια της θερμοδυναμικής.

Η Αρχή βέβαια της Μέγιστης Εντροπίας (Maximum Entropy Principle) είναι ευρέως χρησιμοποιούμενη στην πράξη και πολλές από τις γνωστές κατανομές, τόσο συνεχείς όσο και διακριτές, έχουν χαρακτηριστεί ως κατανομές Μέγιστης Εντροπίας (βλέπε Karur, 1989; Zografos, 1999). Εφαρμογές της Αρχής της Μέγιστης Εντροπίας έχουμε ήδη αναφέρει στην Ενότητα 2.4.

Το *ME* μοντέλο χρησιμοποιείται στην Αναλογιστική Επιστήμη όταν υποθέτουμε σταθερή ένταση θνησιμότητας και ομοιόμορφη κατανομή θανάτων. Πιο συγκεκριμένα έστω ότι η μόνη πληροφορία που διαθέτουμε είναι η πιθανότητα θανάτου σε ένα διάστημα ηλικιών της μορφής  $[a, b]$ . Η πληροφορία αυτή μετατρέπεται σε έναν περιορισμό της μορφής  $\int_{-\infty}^{+\infty} a_0(x)p^*(x)dx = \theta_0$ , όπου

$$a_0 = \begin{cases} 0, & x < a \\ 1, & a \leq x < b \\ 0, & b \geq x, \end{cases}$$

όπου  $x$  είναι η ηλικία και η  $\theta_0$  δεδομένη πιθανότητα θανάτου. Στην περίπτωση αυτή η *ME* κατανομή που ικανοποιεί τον συγκεκριμένο περιορισμό είναι η ομοιόμορφη. Σε περίπτωση που χρησιμοποιήσουμε και άλλους περιορισμούς, δηλαδή περισσότερη πληροφορία, οδηγούμαστε σε άλλα μοντέλα επιβίωσης.

Στη Βιοστατιστική το *ME* μοντέλο χρησιμοποιείται στη φαρμακοκινητική και σε επιδημικά μοντέλα. Για περισσότερες πληροφορίες ο ενδιαφερόμενος αναγνώστης παραπέμπεται

στον Karur (1989).

## 5.6 Στοιχεία Μαθηματικού Προγραμματισμού

Ένα πρόγραμμα μαθηματικού προγραμματισμού με περιορισμούς, είναι ένα πρόγραμμα της μορφής

$$(P) \quad \min \{f_0(\mathbf{x}) \mid \mathbf{x} \in X, f_i(\mathbf{x}) \leq b_i, \quad i = 1, \dots, m \\ h_i(\mathbf{x}) = c_i, \quad i = 1, \dots, k\}. \quad (5.1)$$

Το διάνυσμα  $\mathbf{x} = (x_1, \dots, x_n)$  είναι η μεταβλητή βελτιστοποίησης (optimization variable) του προβλήματος, η συνάρτηση  $f_0 : R^n \rightarrow R$  είναι η αντικειμενική συνάρτηση (objective function) ή συνάρτηση κόστους (cost function), οι συναρτήσεις  $f_i : R^n \rightarrow R, i = 1, \dots, m$ , είναι συναρτήσεις περιορισμών ανισότητας (inequality constraint functions), οι συναρτήσεις  $h_i : R^n \rightarrow R, i = 1, \dots, k$ , είναι συναρτήσεις περιορισμών ισότητας (equality constraint functions) και οι σταθερές  $b_1, \dots, b_m, c_1, \dots, c_k$  είναι τα όρια ή τα φράγματα για τους περιορισμούς. Ένα διάνυσμα  $\mathbf{x}^*$  ονομάζεται βέλτιστο (optimal), ή λύση του προβλήματος (P) που δίνεται στην Εξίσωση 5.1, εάν η αντικειμενική συνάρτηση έχει τη μικρότερη τιμή μεταξύ όλων των διανυσμάτων που ικανοποιούν τους περιορισμούς. Εάν δεν υπάρχουν περιορισμοί λέμε ότι το πρόβλημα (P) είναι χωρίς περιορισμούς (unconstrained).

Το (P) καλείται κυρτό ή πρόγραμμα κυρτού προγραμματισμού, εάν

- το  $X$  είναι κυρτό υποσύνολο του  $R^n$ ,
- οι  $f_0, \dots, f_m$  είναι κυρτές συναρτήσεις πραγματικών αριθμών στο  $X$ , και
- δεν υπάρχει κανένας περιορισμός ισότητας.

Εναλλακτικά αντί περιορισμών ισότητας μπορούμε να έχουμε γραμμικούς περιορισμούς ισότητας. Έτσι οι περιορισμοί αυτοί μπορούν να πέσουν εύκολα στην περίπτωση των περιορισμών ανισότητας αντικαθιστώντας το  $R^n$  με το αφινικό (affine) σύνολο που δίνεται από τους γραμμικούς περιορισμούς ισότητας. Σημειώνουμε ότι ένα σύνολο  $S$  ονομάζεται αφινικό εάν περιέχει τη γραμμή που περνάει από δυο οποιαδήποτε από τα σημεία του, δηλαδή εάν  $x, y \in S$  τότε το  $ax + (1 - a)y$  ανήκει στο  $S$  για όλες τις πραγματικές  $a$ . Με άλλα λόγια, ένα αφινικό σύνολο περιέχει το γραμμικό συνδυασμό δυο οποιονδήποτε σημείων του, με την προϋπόθεση ότι οι συντελεστές στο γραμμικό συνδυασμό αθροίζουν στη μονάδα (Boyd and Vandenberghe, 2006).

Η  $x \in R^n$  ονομάζεται μεταβλητή βελτιστοποίησης (optimization variable) ενώ η συνάρτηση  $f_0 : R^n \rightarrow R$  είναι η αντικειμενική συνάρτηση (objective function) ή συνάρτηση

κόστους (cost function). Οι ανισότητες  $f_i(x) \leq 0$  ονομάζονται περιορισμοί ανισότητας (inequality constraints) και οι αντίστοιχες συναρτήσεις  $f_i : R^n \rightarrow R$  ονομάζονται συναρτήσεις περιορισμού ανισότητας (inequality constraint functions). Επιπλέον, οι ισότητες  $h_i(x) = 0$  ονομάζονται περιορισμοί ισότητας (equality constraints) και οι συναρτήσεις  $h_i : R^n \rightarrow R$  καλούνται συναρτήσεις περιορισμού ισότητας (equality constraint functions). Εάν δεν υπάρχουν περιορισμοί, ισχύει δηλαδή  $m = p = 0$ , λέμε ότι το πρόβλημα  $(P)$  είναι χωρίς περιορισμούς (unconstrained).

Το σύνολο των σημείων για τα οποία ορίζεται η αντικειμενική και όλες οι συναρτήσεις περιορισμών, συμβολίζεται με

$$D = \bigcap_{i=0}^m \text{dom} f_i \cap \bigcap_{i=1}^k \text{dom} h_i$$

και καλείται domain του προβλήματος βελτιστοποίησης  $(P)$ . Ένα σημείο  $\mathbf{x} \in D$  είναι εφικτό (feasible) εάν ικανοποιεί τους περιορισμούς  $f_i(\mathbf{x}) \leq b_i, i = 1, \dots, m$ , και  $h_i(\mathbf{x}) = c_i, i = 1, \dots, k$ .

Το πρόβλημα  $(P)$  λέγεται ότι είναι εφικτό εάν υπάρχει τουλάχιστον ένα εφικτό σημείο, διαφορετικά ονομάζεται μη εφικτό. Το σύνολο όλων των εφικτών σημείων καλείται εφικτό σύνολο (feasible set) ή σύνολο περιορισμών (constraint set).

Η βέλτιστη τιμή (optimal value)  $p^*$  του προβλήματος  $(P)$  ορίζεται ως

$$p^* = \inf \{f_0(\mathbf{x}) | f_i(\mathbf{x}) \leq b_i, i = 1, \dots, m, h_i(\mathbf{x}) = c_i, i = 1, \dots, k\}.$$

Το  $p^*$  επιτρέπεται να παίρνει τιμές  $\pm\infty$  ενώ εάν το πρόβλημα δεν είναι εφικτό, έχουμε  $p^* = \infty$ . Εάν υπάρχουν εφικτά σημεία  $x_k$  με  $f_0(x_k) \rightarrow -\infty$  όταν  $k \rightarrow \infty$ , τότε  $p^* = -\infty$ , και λέμε ότι το πρόβλημα  $(P)$  δεν είναι φραγμένο από κάτω.

### 5.6.1 Βέλτιστα και τοπικά βέλτιστα σημεία

Ένα σημείο  $\mathbf{x}^*$  ονομάζεται βέλτιστο σημείο ή λύνει το πρόβλημα  $(P)$  εάν το  $\mathbf{x}^*$  είναι εφικτό και  $f_0(\mathbf{x}^*) = p^*$ . Το σύνολο όλων των εφικτών σημείων είναι το βέλτιστο σύνολο (optimal set), το οποίο συμβολίζεται με

$$X_{\text{opt}} = \{\mathbf{x} | f_i(\mathbf{x}) \leq b_i, i = 1, \dots, m, h_i(\mathbf{x}) = c_i, i = 1, \dots, k, f_0(\mathbf{x}) = p^*\}.$$

Εάν υπάρχει ένα βέλτιστο σημείο για το πρόβλημα  $(P)$ , λέμε ότι η βέλτιστη τιμή επιτυγχάνεται και το πρόβλημα είναι επιλύσιμο. Εάν το  $X_{\text{opt}}$  είναι κενό, λέμε αντίστοιχα ότι η βέλτιστη τιμή δεν επιτυγχάνεται, πράγμα που συμβαίνει πάντα όταν το πρόβλημα δεν είναι φραγμένο από κάτω. Ένα εφικτό σημείο  $\mathbf{x}$  με  $f_0(\mathbf{x}) \leq p^* + \epsilon$  (όπου  $\epsilon > 0$ ) καλείται  $\epsilon$ -υποβέλτιστο ( $\epsilon$ -suboptimal) και το σύνολο όλων των  $\epsilon$ -υποβέλτιστων σημείων ονομάζεται το  $\epsilon$ -υποβέλτιστο

σύνολο για το πρόβλημα  $(P)$ .

Λέμε ότι ένα εφικτό σημείο  $\mathbf{x}$  είναι τοπικό εφικτό σημείο εάν υπάρχει ένα  $R > 0$ , τέτοιο ώστε

$$f_0(\mathbf{x}) = \inf \{f_0(\mathbf{z}) \mid f_i(\mathbf{z}) \leq b_i, i = 1, \dots, m, h_i(\mathbf{z}) = c_i, i = 1, \dots, k, \|\mathbf{z} - \mathbf{x}\|_2 \leq R\}$$

ή, με άλλα λόγια, το  $\mathbf{x}$  λύνει το πρόβλημα βελτιστοποίησης

$$\begin{aligned} &\text{ελαχιστοποίησε την} && f_0(\mathbf{z}) \\ &\text{υπό τους περιορισμούς} && f_i(\mathbf{z}) \leq b_i, i = 1, \dots, m \\ &&& h_i(\mathbf{z}) = c_i, i = 1, \dots, k \\ &&& \|\mathbf{z} - \mathbf{x}\|_2 \leq R \end{aligned}$$

με τη μεταβλητή  $\mathbf{z}$ . Αυτό σημαίνει ότι το  $\mathbf{x}$  ελαχιστοποιεί την  $f_0$  πάνω σε γειτονικά σημεία στο εφικτό σύνολο.

Εάν το  $\mathbf{x}$  είναι εφικτό σημείο και  $f_i(\mathbf{x}) = 0$ , τότε λέμε ότι ο  $i$ -στός περιορισμός ανισότητας  $f_i(\mathbf{x}) \leq 0$  είναι ενεργός στο  $\mathbf{x}$ . Εάν  $f_i(\mathbf{x}) < 0$ , τότε λέμε ότι ο περιορισμός  $f_i(\mathbf{x}) \leq 0$  είναι ανενεργός. Οι περιορισμοί ισότητας είναι ενεργοί σε όλα τα εφικτά σημεία. Ένας περιορισμός λέμε ότι είναι πλεονάζων (redundant) εάν όταν τον αφαιρέσουμε δεν αλλάζει το εφικτό σύνολο.

### 5.6.2 Πρόγραμμα Κυρτού Προγραμματισμού

Ένα πρόβλημα κυρτού προγραμματισμού έχει τη μορφή

$$\begin{aligned} (CP) \quad &\text{ελαχιστοποίησε την} && f_0(\mathbf{x}) \\ &\text{υπό τους περιορισμούς} && f_i(\mathbf{x}) \leq 0, i = 1, \dots, m \\ &&& \mathbf{a}_i^T(\mathbf{x}) = c_i, i = 1, \dots, k, \end{aligned}$$

όπου οι συναρτήσεις  $f_0, \dots, f_m : R^n \rightarrow R$  είναι κυρτές, ικανοποιούν δηλαδή τη σχέση

$$f_i(\alpha\mathbf{x} + \beta\mathbf{y}) \leq \alpha f_i(\mathbf{x}) + \beta f_i(\mathbf{y})$$

για όλα τα  $\mathbf{x}, \mathbf{y} \in R^n$  και όλα τα  $\alpha, \beta \in R$  με  $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$ . Συγκρίνοντας το πρόβλημα  $(CP)$  με το τυπικό γενικής μορφής πρόβλημα  $(P)$ , το κυρτό πρόβλημα έχει τρεις επιπλέον απαιτήσεις:

- η αντικειμενική συνάρτηση πρέπει να είναι κυρτή,
- οι συναρτήσεις περιορισμών ανισότητας πρέπει να είναι κυρτές, και

- οι συναρτήσεις περιορισμών ισότητας πρέπει να είναι αφινικές.

Σημειώνουμε ότι μια συνάρτηση  $f : R^n \rightarrow R^m$  είναι αφινική εάν είναι το άθροισμα μιας γραμμικής συνάρτησης και μιας σταθεράς, δηλαδή εάν είναι της μορφής  $f(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$ , όπου  $\mathbf{A} \in R^{m \times n}$  και  $\mathbf{b} \in R^m$  (Boyd and Vandenberghe, 2006). Εναλλακτικά, η  $f$  αφινική εάν και μόνο εάν είναι ταυτόχρονα κυρτή και κοίλη.

Το κυρτό πρόγραμμα έχει μια σημαντική ιδιότητα. Το εφικτό σύνολο ενός προβλήματος κυρτής βελτιστοποίησης είναι κυρτό καθώς είναι η τομή του domain του προβλήματος

$$D = \bigcap_{i=0}^m \text{dom } f_i,$$

το οποίο είναι ένα κυρτό σύνολο, με  $m$  (κυρτά) υποεπίπεδα σύνολα  $\{\mathbf{x} | f_i(\mathbf{x}) \leq 0\}$  και  $p$  υπερεπίπεδα  $\{\mathbf{x} | \mathbf{a}_i^T \mathbf{x} = b_i\}$ . Επιπλέον, μπορούμε να υποθέσουμε χωρίς απώλεια της γενικότητας, ότι  $\mathbf{a}_i \neq \mathbf{0}$ : εάν  $\mathbf{a}_i = \mathbf{0}$  και  $b_i = 0$  για ορισμένα  $i$ , τότε ο  $i$ -οστός περιορισμός ισότητας μπορεί να διαγραφεί, εάν πάλι  $\mathbf{a}_i = \mathbf{0}$  και  $b_i \neq 0$ , ο  $i$ -οστός περιορισμός ισότητας είναι ασταθής (inconsistent) και το πρόβλημα δεν είναι εφικτό. Συνεπώς, σε ένα πρόβλημα κυρτής βελτιστοποίησης, ελαχιστοποιούμε μια κυρτή αντικειμενική συνάρτηση πάνω σε ένα κυρτό σύνολο.

Εάν η  $f_0$  είναι ψευδοκυρτή (quasiconvex) αντί κυρτής, λέμε ότι το πρόβλημα (CP) είναι (τυπικής μορφής) πρόβλημα ψευδοκυρτής βελτιστοποίησης. Καθώς τα σύνολα υποεπιπέδων μιας κυρτής ή ψευδοκυρτής συνάρτησης είναι κυρτά, συμπεραίνουμε ότι για ένα πρόβλημα κυρτής ή ψευδοκυρτής βελτιστοποίησης τα  $\epsilon$ -υποβέλτιστα σύνολα είναι κυρτά. Πιο συγκεκριμένα, το βέλτιστο σύνολο είναι κυρτό. Εάν η αντικειμενική συνάρτηση είναι αυστηρώς κυρτή, τότε το βέλτιστο σύνολο περιέχει το πολύ ένα σημείο.

### 5.6.3 Προβλήματα τετραγωνικής βελτιστοποίησης

Το πρόβλημα κυρτής βελτιστοποίησης (CP) ονομάζεται τετραγωνικό πρόγραμμα (QP) εάν η αντικειμενική συνάρτηση είναι (κυρτή) τετραγωνική και οι συναρτήσεις περιορισμών είναι αφινικές. Ένα τετραγωνικό πρόγραμμα μπορεί να εκφραστεί στη μορφή

$$\begin{aligned} (QP) \quad & \text{ελαχιστοποίησε την} \quad \frac{1}{2} \mathbf{x}^T \mathbf{P} \mathbf{x} + \mathbf{q}^T \mathbf{x} + r \\ & \text{υπό τους περιορισμούς} \quad \mathbf{G} \mathbf{x} \leq \mathbf{h} \\ & \quad \quad \quad \mathbf{A} \mathbf{x} = \mathbf{b}, \end{aligned}$$

όπου  $\mathbf{P} \in S_+^n$ ,  $\mathbf{G} \in R^{m \times n}$  και  $\mathbf{A} \in R^{p \times n}$  ενώ ο συμβολισμός  $\leq$  σημαίνει την ανά συνιστώσα ανισότητα (componentwise inequality). Σε ένα τετραγωνικό πρόγραμμα ελαχιστοποιούμε μια κυρτή τετραγωνική συνάρτηση πάνω σε ένα πολύεδρο. Εάν η αντικειμενική συνάρτηση του προβλήματος (CP) καθώς επίσης και οι συναρτήσεις περιορισμών ανισότητας είναι

(κυρτές) τετραγωνικές, όπως στο

$$(QCQP) \quad \begin{aligned} & \text{ελαχιστοποίησε την} \quad \frac{1}{2} \mathbf{x}^T \mathbf{P}_0 \mathbf{x} + \mathbf{q}_0^T \mathbf{x} + r_0 \\ & \text{υπό τους περιορισμούς} \quad \frac{1}{2} \mathbf{x}^T \mathbf{P}_i \mathbf{x} + \mathbf{q}_i^T \mathbf{x} + r_i \leq 0, \quad i = 1, \dots, m \\ & \mathbf{A} \mathbf{x} = \mathbf{b}, \end{aligned}$$

όπου  $\mathbf{P}_i \in S_+^n$ ,  $i = 0, 1, \dots, m$ , το πρόβλημα καλείται τετραγωνικό πρόγραμμα με τετραγωνικούς περιορισμούς (QCQP). Σε ένα (QCQP) πρόγραμμα, ελαχιστοποιούμε μια κυρτή τετραγωνική συνάρτηση πάνω σε μια εφικτή περιοχή, η οποία είναι η τομή ελλειψοειδών (όταν  $\mathbf{P}_i > 0$ ).

Τα τετραγωνικά προγράμματα περιλαμβάνουν ως ειδική περίπτωση τα γραμμικά προγράμματα όταν ισχύει  $\mathbf{P} = \mathbf{0}$  στο (QP). Επίσης περιλαμβάνουν τα τετραγωνικά προγράμματα όταν  $\mathbf{P}_i = \mathbf{0}$  στο (QCQP), για  $i = 1, \dots, m$ .

#### 5.6.4 Λαγκρανζιανή συνάρτηση και λαγκρανζιανή δυϊκότητα

Το αποτέλεσμα του Κυρτού Θεωρήματος επί της Εναλλακτικής (Convex Theorem on Alternative) [Boyd and Vandenberghe (2006)] φέρνει στην προσοχή μας τη συνάρτηση

$$L(\boldsymbol{\lambda}) = \inf_{\mathbf{x} \in X} \left[ f_0(\mathbf{x}) + \sum_{j=1}^m \lambda_j g_j(\mathbf{x}) \right], \quad (5.2)$$

η οποία είναι ισοδύναμη και προέρχεται από το άθροισμα

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f_0(\mathbf{x}) + \sum_{j=1}^m \lambda_j g_j(\mathbf{x}). \quad (5.3)$$

Η Εξίσωση 5.3 έχει ιδιαίτερη σημασία στον μαθηματικό προγραμματισμό και ονομάζεται Λαγκρανζιανή συνάρτηση (Lagrange function) του προγράμματος ελαχιστοποίησης υπό περιορισμούς ανισότητας (inequality constrained optimization program)

$$(IC) \quad f(\mathbf{x}) \rightarrow \min g_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, m, \quad \mathbf{x} \in X.$$

Η Λαγκρανζιανή συνάρτηση ενός προγράμματος βελτιστοποίησης είναι μια πολύ σημαντική ποσότητα, καθώς οι περισσότερες συνθήκες βελτιστοποίησης εκφράζονται σε όρους αυτής της συνάρτησης. Στη συνέχεια παραθέτουμε το Θεώρημα της Δυϊκότητας του Κυρτού Προγραμματισμού.

**Θεώρημα 5.1.** Θεωρούμε ένα αυθαίρετο πρόγραμμα βελτιστοποίησης με περιορισμούς ανισότητας (IC). Τότε



(i) Το *infimum* της Λαγκρανζιανής συνάρτησης στο  $x \in X$  είναι για κάθε  $\lambda \geq 0$  ένα κάτω φράγμα για τη βέλτιστη τιμή στο (IC), έτσι ώστε η βέλτιστη τιμή στο πρόγραμμα βελτιστοποίησης

$$(IC^*) \quad \sup_{\lambda \geq 0} L(\lambda),$$

να είναι επίσης ένα κάτω φράγμα για τη βέλτιστη τιμή στο (IC).

(ii) (Θεώρημα Κυρτής Δυϊκότητας) Εάν το πρόγραμμα (IC) είναι

- κυρτό,
- κάτω φραγμένο, και
- ικανοποιεί τις συνθήκες του Slater,

τότε η βέλτιστη τιμή στο (IC\*) επιτυγχάνεται και είναι ίση με τη βέλτιστη τιμή στο (IC).

Το Θεώρημα 5.1 (Boyd and Vandenberghe, 2006) καθιερώνει συγκεκριμένη σχέση μεταξύ δυο προγραμμάτων βελτιστοποίησης, του «πρωτεύοντος» προγράμματος (primal problem)

$$(IC) \quad f(\mathbf{x}) \rightarrow \min g_j(\mathbf{x}) \leq 0, \quad j = 1, \dots, m, \quad \mathbf{x} \in X.$$

και του «Λαγκρανζιανού δυϊκού» προγράμματος (dual problem)

$$(IC^*) \quad \sup_{\lambda \geq 0} L(\lambda), \quad L(\lambda) = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda).$$

Οι μεταβλητές  $\lambda$  του δυϊκού προβλήματος ονομάζονται Λαγκρανζιανοί πολλαπλασιαστές του πρωτεύοντος προβλήματος. Το παραπάνω θεώρημα αναφέρει ότι η βέλτιστη τιμή στο δυϊκό πρόβλημα είναι μικρότερη ή ίση με την αντίστοιχη στο πρωτεύον πρόβλημα, και υπό ορισμένες περιστάσεις (το πρωτεύον πρόβλημα είναι κυρτό, φραγμένο από κάτω και ικανοποιεί τις συνθήκες του Slater), οι βέλτιστες τιμές στα προγράμματα είναι ίσες.

Έχουμε αναφέρει ότι οι βέλτιστες τιμές στα προγράμματα (IC) και (IC\*) είναι ίσες μεταξύ τους, υπό ορισμένες υποθέσεις κυρτότητας και ομαλότητας (regularity). Υπάρχει και ένας άλλος τρόπος για να πούμε εάν οι βέλτιστες τιμές είναι ίσες μεταξύ τους - αυτό ισχύει όταν η Λαγκρανζιανή συνάρτηση έχει ένα σαγματικό σημείο (saddle point), δηλαδή υπάρχει ένα ζεύγος  $\mathbf{x}^* \in X$ ,  $\lambda^* \geq 0$  τέτοιο ώστε η συνάρτηση  $L(\mathbf{x}, \lambda)$  επιτυγχάνει το ελάχιστό της στο  $\mathbf{x} \in X$  και το μέγιστό της στο  $\lambda^* \geq 0$  και ισχύει η σχέση

$$L(\mathbf{x}, \lambda^*) \geq L(\mathbf{x}^*, \lambda^*) \geq L(\mathbf{x}^*, \lambda) \quad \forall \mathbf{x} \in X, \lambda \geq 0.$$

Είναι εύκολο να δειχθεί η παρακάτω πρόταση.

**Πρόταση 5.1.** Το ζεύγος  $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$  είναι σαγματικό σημείο της Λαγκρανζιανής συνάρτησης  $L$  του προγράμματος (IC) εάν και μόνο εάν το  $\mathbf{x}^*$  είναι βέλτιστη λύση του (IC), το  $\boldsymbol{\lambda}^*$  είναι βέλτιστη λύση του προγράμματος (IC\*) και οι βέλτιστες τιμές στα δυο προβλήματα είναι ίσες.

Μια ισοδύναμη διατύπωση των παραπάνω έχει ως εξής: Θεωρούμε το παρακάτω πρόβλημα βελτιστοποίησης:

$$\begin{aligned} & \text{ελαχιστοποίησε την } f_0(\mathbf{x}) \\ & \text{υπό τους περιορισμούς } f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m; \\ & \quad \quad \quad h_i(\mathbf{x}) = 0; \quad i = 1, \dots, k; \end{aligned} \quad (5.4)$$

όπου  $\mathbf{x} \in R^n$ . Θεωρούμε ότι το domain

$$D = \bigcap_{i=0}^m \text{dom} f_i \cap \bigcap_{i=1}^k \text{dom} h_i$$

δεν είναι κενό και συμβολίζουμε τη βέλτιστη λύση του παραπάνω προβλήματος με  $\mathbf{p}^*$ . Υποθέτουμε ότι το πρόβλημα 5.4 δεν είναι κυρτό.

Η βασική ιδέα στη Λαγκρανζιανή δυϊκότητα είναι να λάβουμε υπόψη μας τους περιορισμούς του προβλήματος 5.4 αυξάνοντας την αντικειμενική συνάρτηση με ένα σταθμισμένο άθροισμα των συναρτήσεων των περιορισμών. Η Λαγκρανζιανή συνάρτηση  $L : R^n \times R_m \times R_p \rightarrow R$  που σχετίζεται με το πρόβλημα 5.4 έχει ως εξής:

$$L(\mathbf{x}; \boldsymbol{\lambda}; \boldsymbol{\nu}) = f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^k \nu_i h_i(\mathbf{x});$$

με  $\text{dom} L = D \times R^m \times R^p$ . Αναφερόμαστε στο  $\lambda_i$  ως ο πολλαπλασιαστής Lagrange που σχετίζεται με τον  $i$ -οστό περιορισμό ανισότητας  $f_i(\mathbf{x}) \leq 0$ . Αντίστοιχα, αναφερόμαστε στο  $\nu_i$  ως ο πολλαπλασιαστής Lagrange που σχετίζεται με τον  $i$ -οστό περιορισμό ισότητας  $h_i(\mathbf{x}) = 0$ . Τα διανύσματα  $\boldsymbol{\lambda}$  και  $\boldsymbol{\nu}$  ονομάζονται δυϊκές μεταβλητές ή διανύσματα των πολλαπλασιαστών Lagrange που σχετίζονται με το πρόβλημα 5.4.

Η Λαγκρανζιανή δυϊκή συνάρτηση (ή απλά δυϊκή συνάρτηση)  $g : R^m \times R^p \rightarrow R$  ορίζεται ως την ελάχιστη τιμή της Λαγκρανζιανής συνάρτησης πάνω στα  $\mathbf{x}$ : για  $\boldsymbol{\lambda} \in R^m$ ,  $\boldsymbol{\nu} \in R^p$ ,

$$g(\boldsymbol{\lambda}; \boldsymbol{\nu}) = \inf_{\mathbf{x} \in D} L(\mathbf{x}; \boldsymbol{\lambda}; \boldsymbol{\nu}) = \inf_{\mathbf{x} \in D} \left( f_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^k \nu_i h_i(\mathbf{x}) \right).$$

Όταν η Λαγκρανζιανή συνάρτηση δεν είναι φραγμένη από κάτω ως προς  $\mathbf{x}$ , η δυϊκή συνάρτηση παίρνει την τιμή  $-\infty$ . Καθώς η δυϊκή συνάρτηση είναι το ανά σημείο ελάχιστο μιας

οικογένειας αφινικών συναρτήσεων του  $(\boldsymbol{\lambda}; \boldsymbol{\nu})$ , είναι κοίλη, ακόμα και όταν το πρόβλημα 5.4 δεν είναι κυρτό.

Τέλος, για το λαγκρανζιανό δυϊκό πρόβλημα ισχύουν τα εξής: Για κάθε ζεύγος  $(\boldsymbol{\lambda}; \boldsymbol{\nu})$  με  $\boldsymbol{\lambda} \geq \mathbf{0}$ , η Λαγκρανζιανή συνάρτηση δίνει ένα κάτω φράγμα για τη βέλτιστη τιμή  $p^*$  του προβλήματος βελτιστοποίησης 5.4. Έτσι έχουμε ένα κάτω όριο που εξαρτάται από κάποιες παραμέτρους  $\boldsymbol{\lambda}, \boldsymbol{\nu}$ . Μια φυσική ερώτηση είναι ποιο είναι το καλύτερο κάτω φράγμα που μπορούμε να πάρουμε από τη Λαγκρανζιανή δυϊκή συνάρτηση.

Το ερώτημα αυτό οδηγεί στο πρόβλημα βελτιστοποίησης

$$\begin{aligned} & \text{ελαχιστοποίησε την } g(\boldsymbol{\lambda}; \boldsymbol{\nu}) \\ & \text{υπό τους περιορισμούς } \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \tag{5.5}$$

Το πρόβλημα αυτό είναι δυϊκό πρόβλημα που σχετίζεται με το πρόβλημα 5.4. Υπό το πλαίσιο αυτό το αρχικό πρόβλημα 5.4 καλείται πρωτεύον πρόβλημα. Ο όρος δυϊκός εφικτός, για να περιγράψει ένα ζεύγος  $(\boldsymbol{\lambda}; \boldsymbol{\nu})$  με  $(\boldsymbol{\lambda} \geq \mathbf{0} \text{ και } g(\boldsymbol{\lambda}; \boldsymbol{\nu}) > -\infty)$ , σημαίνει ότι το  $(\boldsymbol{\lambda}; \boldsymbol{\nu})$  είναι εφικτό για το πρόβλημα 5.5. Αναφερόμαστε στο  $(\boldsymbol{\lambda}^*; \boldsymbol{\nu}^*)$  ως το δυϊκό βέλτιστο ή τους βέλτιστους δυϊκούς πολλαπλασιαστές εάν είναι βέλτιστοι για το πρόβλημα 5.5. Το Λαγκρανζιανό δυϊκό πρόβλημα 5.5 είναι ένα πρόβλημα κυρτής βελτιστοποίησης, καθώς η αντικειμενική συνάρτηση που θέλουμε να μεγιστοποιήσουμε είναι κοίλη και οι περιορισμοί κυρτοί. Αυτό ισχύει είτε το πρωτεύον πρόβλημα 5.4 είναι κυρτό είτε όχι.

Στις επόμενες παραγράφους θα εξάγουμε ορισμένα χρήσιμα αποτελέσματα που αφορούν την ελαχιστοποίηση των μέτρων απόκλισης με τα οποία ασχολούμαστε στη διατριβή αυτή. Για την εξαγωγή των αποτελεσμάτων θα στηριχθούμε σε έννοιες του κυρτού μαθηματικού προγραμματισμού και της λαγκρανζιανής δυϊκότητας που αναφέραμε παραπάνω.

## 5.7 Ελαχιστοποίηση του Διακριτού Μέτρου των Kullback - Leibler με γραμμικούς και τετραγωνικούς περιορισμούς

Στη διακριτή περίπτωση, η αντικειμενική συνάρτηση την οποία θέλουμε να ελαχιστοποιήσουμε είναι η

$$I(\mathbf{x}, \mathbf{d}) = \sum_{j=1}^n x_j \ln \frac{x_j}{d_j}$$

με γραμμικούς και τετραγωνικούς περιορισμούς, όπου  $\mathbf{x} = (x_1, \dots, x_n)^T$  και  $\mathbf{d} = (d_1, \dots, d_n)^T$  είναι δυο πεπερασμένα μέτρα από τα οποία τα  $d_i$  είναι γνωστά ενώ τα  $x_i$  είναι άγνωστα.

Ισοδύναμα μπορούμε να αναζητήσουμε την ελαχιστοποίηση της ακόλουθης έκφρασης του

μέτρου των Kullback - Leibler

$$I(\mathbf{x}, \mathbf{d}) = \sum_{j=1}^n x_j \ln \frac{x_j}{ed_j}$$

ακολουθώντας το σκεπτικό των Zhang and Brockett (1987), οι οποίοι έχουν αναπτύξει σχετική θεωρία. Ο παράγοντας  $e$  στον παρονομαστή του νεπέρειου λογαρίθμου της αντικειμενικής συνάρτησης δεν παίζει κανένα ρόλο όταν το  $\sum_{i=1}^n x_i$  είναι σταθερό, παρόλα αυτά διευκολύνει στη δυϊκότητα (duality).

Οι Zhang and Brockett (1987) δίνουν το πρωτεύον (primal) πρόβλημα στην περίπτωση των γραμμικών περιορισμών, το οποίο είναι το

$$(P_L) \quad \min I(\mathbf{x}, \mathbf{d})$$

υπό τους περιορισμούς  $\mathbf{A}^T \mathbf{x} \leq \mathbf{v}, \mathbf{x} \geq \mathbf{0}$

όπου  $\mathbf{A}$  είναι ένας  $n \times k$  πίνακας,  $\mathbf{v}$  είναι ένα  $n \times 1$  διάνυσμα σταθερών ποσοτήτων, το  $T$  δηλώνει την αντιστροφή του πίνακα ή του διανύσματος και η ανισότητα νοείται ως ανισότητα ανά συνιστώσα του διανύσματος.

Οι Brockett, Charnes and Cooper (1980) αποδεικνύουν ότι το δυϊκό (dual) πρόβλημα του πρωτεύοντος προβλήματος είναι το

$$(D_L) \quad \max -\mathbf{x}^T \mathbf{e}^{i\mathbf{A}\mathbf{z}} + \mathbf{v}^T \mathbf{z}$$

υπό τον περιορισμό  $\mathbf{z} \leq \mathbf{0}$ .

Το  $i\mathbf{A}$  συμβολίζει την  $i$ -οστή γραμμή του πίνακα  $\mathbf{A}$ , το  $\mathbf{z}$  είναι ένα  $k \times 1$  διάνυσμα άγνωστων σταθερών ποσοτήτων και το  $\mathbf{e}^{\mathbf{x}}$ , με  $\mathbf{x}$  ένα  $n \times 1$  διάνυσμα, συμβολίζει το διάνυσμα  $(e^{x_1}, \dots, e^{x_n})^T$ .

Στην περίπτωση αυτή ισχύει η δυϊκή ανισότητα

$$\sum_{i=1}^n x_i \ln \frac{x_i}{ed_i} \geq -\mathbf{x}^T \mathbf{e}^{i\mathbf{A}\mathbf{z}} + \mathbf{v}^T \mathbf{z}$$

όταν ισχύουν ταυτόχρονα οι περιορισμοί  $\mathbf{A}^T \mathbf{x} \leq \mathbf{v}, \mathbf{x} \geq \mathbf{0}$  και  $\mathbf{z} \leq \mathbf{0}$ .

Οι Zhang and Brockett (1987) βασιζόμενοι στα προηγούμενα, διατύπωσαν το αντίστοιχο πρόβλημα στην περίπτωση που οι περιορισμοί είναι της μορφής τετραγωνικών ανισοτήτων, όπως συμβαίνει στην εξομάλυνση των πινάκων θνησιμότητας. Συγκεκριμένα το πρωτεύον πρόβλημα στην περίπτωση αυτή είναι το

$$(P) \quad \min I(\mathbf{x}, \mathbf{d})$$

υπό τους περιορισμούς  $\mathbf{x} \geq \mathbf{0}$  και  $g_i(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{D}_i \mathbf{x} + \mathbf{b}_i^T \mathbf{x} + c_i \leq 0, i = 1, 2, \dots, r$

όπου ο  $\mathbf{D}_i$  είναι ένας θετικά ημιορισμένος πίνακας για κάθε  $i$  και τα  $\mathbf{b}_i, c_i$  είναι σταθερές.

Παρατηρούμε ότι το πρόβλημα (P) περιέχει σαν ειδικές περιπτώσεις τα προβλήματα (P<sub>L</sub>) και (D<sub>L</sub>) με περιορισμούς της μορφής γραμμικών ισοτήτων και γραμμικών ανισοτήτων.

Τελικά κατέληξαν ότι το δυϊκό του προβλήματος (P) είναι το

$$(D) \quad \max -\mathbf{x}^T \exp \left\{ -\sum_{i=1}^r y_i (\mathbf{A}_i^T \mathbf{w}_i + \mathbf{b}_i) \right\} + \mathbf{c}^T \mathbf{y} - \frac{1}{2} \sum_{i=1}^r \rho_i$$

υπό τους περιορισμούς  $\mathbf{y} \geq \mathbf{0}$  και  $\mathbf{w}_i \in R^{m_i}$ .

Το  $\mathbf{A}_i$  είναι ένας  $m_i \times n$  πίνακας για τον οποίο ισχύει  $\mathbf{D}_i = \mathbf{A}_i^T \mathbf{A}_i$ , όπου  $m_i$  είναι η τάξη του πίνακα  $\mathbf{D}_i$ ,  $\mathbf{w}_i$  είναι ένα  $m_i \times 1$  διάνυσμα πραγματικών αριθμών,  $\mathbf{y}$  είναι ένα  $r \times 1$  διάνυσμα άγνωστων μεταβλητών και  $\rho_i = y_i \|\mathbf{w}_i\|^2$ .

Παρατηρούμε ότι το δυϊκό πρόβλημα (D) δεν έχει κανέναν περιορισμό πέραν της μη αρνητικότητας ενός μικρού αριθμού μεταβλητών και συνεπώς επιλύεται ευκολότερα από το πρωτεύον πρόβλημα (P). Επίσης παρατηρούμε ότι το πρόβλημα (D) είναι μια επέκταση του προβλήματος ( $D_L$ ) με την έννοια ότι όλοι οι περιορισμοί του ( $D_L$ ) είναι γραμμικοί, δηλαδή αν  $\mathbf{D}_i = \mathbf{0}$  τότε το (D) είναι όμοιο με το ( $D_L$ ).

Συνοψίζοντας την παραπάνω μέθοδο, μπορούμε να πούμε ότι για να βρούμε το δυϊκό πρόβλημα ενός προβλήματος με κυρτής μορφής περιορισμούς, αρκεί να γνωρίζουμε τη δυϊκή μορφή ενός κυρτού προβλήματος με γραμμικούς περιορισμούς. Συγκεκριμένα, με ένα σύνολο βοηθητικών υπερεπιπέδων προσεγγίζουμε την εφικτή περιοχή, βρίσκουμε τη δυϊκή έκφραση για το προκύπτον πρόβλημα που έχει γραμμικούς περιορισμούς και στη συνέχεια παίρνουμε όρια και απλοποιούμε την προκύπτουσα έκφραση. Για περισσότερες πληροφορίες για την εξαγωγή του προβλήματος (D), παραπέμπουμε στους Zhang and Brockett (1987).

Ο Teboulle (1989) παρήγαγε το ίδιο δυϊκό πρόβλημα για το πρόβλημα που εισήγαξαν οι Zhang and Brockett (1987), μέσω απλής εφαρμογής της Λαγκρανζιανής δυϊκότητας (Boyd and Vandenberghe, 2006). Πιο συγκεκριμένα, μέσω μιας απλής διάσπασης της έκφρασης μετέτρεψε το πρόβλημα (P) σε ένα αντίστοιχο κυρτό πρόγραμμα με γραμμικούς και τετραγωνικούς περιορισμούς.

Κάθε πίνακας  $\mathbf{D}_i$  είναι ημιθετικά ορισμένος, συνεπώς μπορεί να εκφρασθεί ως  $\mathbf{D}_i = \mathbf{A}_i^T \mathbf{A}_i$ , όπου  $\mathbf{A}_i$  είναι ένας  $n_i \times n$  πίνακας με  $n_i$  τον βαθμό (rank) του  $\mathbf{D}_i$ ,  $i = 1, 2, \dots, m$ . Έτσι έχουμε  $g_i(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A}_i^T \mathbf{A}_i \mathbf{x} + \mathbf{b}_i^T \mathbf{x} + c_i$ . Ορίζοντας τις νέες μεταβλητές  $\mathbf{u}_i = \mathbf{A}_i \mathbf{x}$ ,  $\mathbf{u}_i \in R^{n_i}$ ,  $i = 1, \dots, m$ , το πρόβλημα (P) είναι αντίστοιχο με το παρακάτω κυρτό πρόβλημα με γραμμικής ισότητας και τετραγωνικής ανισότητας περιορισμούς.

$$(P_{KL}) \quad \min_{\mathbf{x}, \mathbf{u}_i} \sum_{j=1}^n x_j \ln \frac{x_j}{ed_j}$$

υπό τους περιορισμούς  $\frac{1}{2} \mathbf{u}^T \mathbf{u}_i + \mathbf{b}_i^T \mathbf{x} + c_i \leq 0$ ,  $\mathbf{A}_i \mathbf{x} = \mathbf{u}_i$ ,  $\mathbf{u}_i \in R^{n_i}$ ,  $i = 1, \dots, m$ ,  $\mathbf{x} \geq \mathbf{0}$ .

Ορίζοντας τα  $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_m) \in R^{n_1} \times \dots \times R^{n_m}$ ,  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_m) \in R^{n_1} \times \dots \times R^{n_m}$  και  $N = (n_1 + \dots + n_m)$  το Λαγκρανζιανό δυϊκό πρόβλημα του προβλήματος ( $P_{KL}$ ) δίνεται στο παρακάτω θεώρημα.

**Θεώρημα 5.2.** Το Λαγκρανζιανό δυϊκό πρόβλημα του προβλήματος ( $P_{KL}$ ) είναι το

$$(D_{KL}) \quad \sup_{\lambda \in R_+^m, \mathbf{y}_i \in R^{n_i}} \left\{ \lambda^T \mathbf{c} - \sum_{j=1}^n d_j \exp \left\{ -\sum_{i=1}^m (\lambda_i \mathbf{b}_i + \mathbf{A}_i^T \mathbf{y}_i) \right\} - \frac{1}{2} \sum_{i=1}^m \frac{\|\mathbf{y}_i\|^2}{\lambda_i} \right\}.$$

Απόδειξη. (Σκιαγράφηση) Η λαγκρανζιανή συνάρτηση για το πρόβλημα ( $P_{KL}$ ) είναι η  $L : R_+^n \times R_N \times R_+^m \times R_N \rightarrow R$ ,

$$L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) = \boldsymbol{\lambda}^T \mathbf{c} + \sum_{j=1}^n x_j \ln \frac{x_j}{ed_j} + \sum_{i=1}^m (\lambda_i \mathbf{b}_i + \mathbf{A}_i^T \mathbf{y}_i)^T \mathbf{x} + \sum_{i=1}^m \left( \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right). \quad (5.6)$$

Η λαγκρανζιανή δυϊκή αντικειμενική συνάρτηση (lagrangian dual objective function) είναι η

$$h(\boldsymbol{\lambda}, \mathbf{y}) = \inf L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) : \mathbf{x} \geq \mathbf{0}, \mathbf{u}_i \in R^{n_i}, \quad (5.7)$$

και το δυϊκό πρόβλημα ( $D_{KL}$ ) που σχετίζεται με το πρόβλημα ( $P_{KL}$ ) ορίζεται ως

$$(D) \quad \sup \{ h(\boldsymbol{\lambda}, \mathbf{y}) : \boldsymbol{\lambda} \in R_+^m, \mathbf{y}_i \in R^{n_i} \}.$$

Η λαγκρανζιανή συνάρτηση της Σχέσης 5.6 είναι διαχωρίσιμη (seperable) στις δυο μεταβλητές απόφασης  $\mathbf{x}$  και  $\mathbf{u}$ . Αυτό επιτρέπει την εξαγωγή μιας σαφούς μορφής για τη δυϊκή αντικειμενική συνάρτηση  $h(\boldsymbol{\lambda}, \mathbf{y})$ . Έτσι από τις Σχέσεις 5.6 και 5.7, έχουμε

$$h(\boldsymbol{\lambda}, \mathbf{y}) = \mathbf{c}^T \boldsymbol{\lambda} + \sum_{j=1}^n \inf_{x_j \geq 0} \left\{ x_j \ln \left( \frac{x_j}{ed_j} + \sum_{i=1}^m (\lambda_i \mathbf{b}_i + \mathbf{A}_i^T \mathbf{y}_i)_j x_j \right) \right\} + \sum_{i=1}^m \inf_{\mathbf{u}_i \in R^{n_i}} \left\{ \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right\}. \quad (5.8)$$

Το πρώτο infimum στην παραπάνω σχέση υπολογίζεται εύκολα και φτάνει την ελάχιστη τιμή του,

$$- \sum_{j=1}^n d_j \exp \left\{ - \sum_{i=1}^m (\lambda_i \mathbf{b}_i + \mathbf{A}_i^T \mathbf{y}_i)_j \right\}, \quad (5.9)$$

στο βέλτιστο σημείο

$$x_j^* = d_j \exp \left\{ - \sum_{i=1}^m (\lambda_i \mathbf{b}_i + \mathbf{A}_i^T \mathbf{y}_i)_j \right\}. \quad (5.10)$$

Η συνθήκη βελτιστοποίησης για το δεύτερο infimum της Σχέσης 5.8 είναι η  $\lambda \mathbf{u}_i - \mathbf{y}_i = 0$ ,  $i = 1, \dots, m$ . Εάν  $\lambda_i > 0$  τότε  $\mathbf{u}_i^* = \mathbf{y}_i / \lambda_i$  και το infimum φτάνει την ελάχιστη τιμή του

$$- \frac{1}{2} \sum_{i=1}^m \frac{\|\mathbf{y}_i\|^2}{\lambda_i}. \quad (5.11)$$

Εάν  $\lambda_i = 0$  και  $\mathbf{y}_i \neq \mathbf{0}$  τότε  $h(\boldsymbol{\lambda}, \mathbf{y}) = -\infty$  και συνεπώς ένα τέτοιο  $\mathbf{y}_i$  δεν μπορεί να θεωρηθεί υποψήφιο για το supremum στη δυϊκή αντικειμενική συνάρτηση, οπότε στη συγκεκριμένη

περίπτωση θέτουμε  $\mathbf{y}_i = \mathbf{0}$ . Αντικαθιστώντας τις Σχέσεις 5.9 και 5.11 στη Σχέση 5.8 παίρνουμε το επιθυμητό αποτέλεσμα. ■

Παρατηρούμε ότι ο Teboulle χρησιμοποιώντας μια απλή διαμέριση του προβλήματος  $(P_{KL})$  παρήγαγε ένα σαφές δυϊκό πρόβλημα, το οποίο συμπίπτει με το δυϊκό πρόβλημα που παρουσίασαν οι Zhang and Brockett. Οι καταστάσεις δυϊκότητας μεταξύ των προβλημάτων  $(P_{KL})$  και  $(D_{KL})$  μπορούν να εξαχθούν μέσω της εφαρμογής αποτελεσμάτων κυρτής δυϊκότητας.

**Θεώρημα 5.3.** (α) Εάν το πρόβλημα  $(P_{KL})$  είναι εφικτό, τότε το  $\inf(P_{KL})$  επιτυγχάνεται και  $\min(P_{KL}) = \sup(D_{KL})$ . Επιπλέον, εάν υπάρχει ένα  $\mathbf{x} \in R^n$  τέτοιο ώστε  $\mathbf{x} > \mathbf{0}$ ,  $g(\mathbf{x}_i) < 0$ ,  $i = 1, \dots, m$  τότε το  $\sup(D_{KL})$  επιτυγχάνεται και  $\min(P_{KL}) = \max(D_{KL})$ .

(β) Εάν το  $\mathbf{x}^*$  είναι η λύση του προβλήματος  $(P_{KL})$  και τα  $\mathbf{y}_i^* \in R^{n_i}$ ,  $\lambda^* \in R_+^{n_i}$  λύνουν το πρόβλημα  $(D_{KL})$ , τότε

$$x_j^* = d_j \exp \left\{ - \sum_{i=1}^m (\lambda_i^* \mathbf{b}_i + \mathbf{A}_i^T \mathbf{y}_i^*)_j \right\}, \quad j = 1, \dots, n.$$

Στην επόμενη παράγραφο, θα παρουσιάσουμε παρόμοια αποτελέσματα δυϊκότητας για την απόκλιση δύναμης των Cressie - Read. Σημειώνουμε ότι για λόγους συμβολισμών, και επειδή οι πολλαπλασιαστές Lagrange συμβολίζονται στη βιβλιογραφία με το ελληνικό γράμμα  $\lambda$ , στην Ενότητα 5.8 έχουμε, σε αντίθεση με τα προηγούμενα κεφάλαια, συμβολίσει την τάξη των αποκλίσεων δύναμης των Cressie - Read με  $r$ .

## 5.8 Δυϊκή θεωρία για το πρόβλημα ελαχιστοποίησης της απόκλισης των Cressie - Read υπό τετραγωνικούς περιορισμούς

Το πρόβλημα ελαχιστοποίησης της απόκλισης των Cressie - Read υπό τετραγωνικούς περιορισμούς ορίζεται ως:

Εύρεση του  $\mathbf{x} \in R^n$ , το οποίο επιλύει το πρωτεύον πρόβλημα

$$(P_{CR}) \quad \min \frac{1}{r(r+1)} \sum_{j=1}^n x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right]$$

υπό τους περιορισμούς

$$g_i(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{D}_i \mathbf{x} + \mathbf{b}_i^T \mathbf{x} + c_i \leq 0, \quad i = 1, 2, \dots, m, \quad \mathbf{x} \geq \mathbf{0},$$

όπου  $\mathbf{d} = (d_1, d_2, \dots, d_n)^T$  είναι ένα γνωστό διάνυσμα με αυστηρώς θετικά στοιχεία,  $\mathbf{D}_i$  είναι ένας θετικά ημιορισμένος πίνακας για κάθε  $i$ ,  $\mathbf{b}_i \in R^n$  και  $c_i$  είναι γνωστές σταθερές όχι ταυτόχρονα ίσες με μηδέν. Για λόγους συμβολισμού, και επειδή με  $\lambda$  συνήθως στη βιβλιογραφία συμβολίζονται οι πολλαπλασιαστές Lagrange, στην ενότητα αυτή η παράμετρος δύναμης του μέτρου απόκλισης των Cressie - Read θα συμβολίζεται με  $r$ .

Στη συνέχεια, θα προσπαθήσουμε να εξάγουμε μια δυϊκή αναπαράσταση του πρωτεύοντος προβλήματος ( $P_{CR}$ ) μέσω Λαγκρανζιανής δυϊκότητας. Αυτό θα γίνει χρησιμοποιώντας μια απλή διαμέριση για να μετατρέψουμε το ( $P_{CR}$ ) σε ένα ισοδύναμο κυρτό πρόγραμμα με γραμμικούς και τετραγωνικούς περιορισμούς.

Επειδή ο  $\mathbf{D}_i$  είναι ένας ημιθετικά ορισμένος  $n \times n$  πίνακας, μπορούμε να τον εκφράσουμε ως  $\mathbf{D}_i = \mathbf{A}_i^T \mathbf{A}_i$ , όπου  $\mathbf{A}_i$  είναι ένας  $n_i \times n$  πίνακας και  $n_i$  είναι η τάξη (rank) του  $\mathbf{D}_i$ ,  $i = 1, 2, \dots, m$ . Στην περίπτωση αυτή, οι περιορισμοί μπορούν να γραφούν ως  $g_i(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A}_i^T \mathbf{A}_i \mathbf{x} + \mathbf{b}_i^T \mathbf{x} + c_i$ . Ορίζοντας τις νέες μεταβλητές  $\mathbf{u}_i = \mathbf{A}_i \mathbf{x}$ ,  $\mathbf{u}_i \in R^{n_i}$ ,  $i = 1, 2, \dots, m$ , το πρόβλημα ( $P_{CR}$ ) είναι ισοδύναμο με το ακόλουθο κυρτό πρόγραμμα με περιορισμούς γραμμικής ισότητας και τετραγωνικής ανισότητας:

$$(P_{CR}^*) \quad \min_{\mathbf{x}, \mathbf{u}_i} \frac{1}{r(r+1)} \sum_{j=1}^n x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right]$$

υπό τους περιορισμούς

$$\frac{1}{2} \mathbf{u}_i^T \mathbf{u}_i + \mathbf{b}_i^T \mathbf{x} + c_i \leq 0, \quad \mathbf{A}_i \mathbf{x} = \mathbf{u}_i, \quad \mathbf{u}_i \in R^{n_i}, \quad i = 1, 2, \dots, m, \quad \mathbf{x} \geq \mathbf{0}.$$

Έστω  $\mathbf{u} = (u_1, u_2, \dots, u_m)^T \in R^{n_1} \times \dots \times R^{n_m}$ ,  $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)^T \in R^{n_1} \times \dots \times R^{n_m}$  και  $N = n_1 + \dots + n_m$ .

**Θεώρημα 5.4.** Το Λαγκρανζιανό δυϊκό πρόβλημα του ( $P_{CR}$ ) δίνεται από το

$$(D) \quad \sup_{\boldsymbol{\lambda} \in R_+^m, \mathbf{y}_i \in R^{n_i}} \left\{ \sum_{j=1}^n d_j \left( \frac{1}{r+1} - r \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right)^{\frac{1}{r}} \cdot \left( r(r+1) \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j - 1 \right) - \frac{1}{2} \sum_{i=1}^m \frac{\|\mathbf{y}_i\|^2}{\lambda_i} + \boldsymbol{\lambda}^T \mathbf{c} \right\}.$$

*Απόδειξη.* Η Λαγκρανζιανή συνάρτηση του πρωτεύοντος προβλήματος ( $P_{CR}$ ) όπως δείχνεται στο Παράρτημα Α, είναι η  $L : R_+^m \times R^N \times R_+^m \times R^N \rightarrow R$  με

$$L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) = \frac{1}{r(r+1)} \sum_{j=1}^n x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right] + \boldsymbol{\lambda}^T \mathbf{c} + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i) \mathbf{x} + \sum_{i=1}^m \left( \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right). \quad (5.12)$$

Η Λαγκρανζιανή δυϊκή αντικειμενική συνάρτηση του προβλήματος ( $P_{CR}$ ) δίνεται από την

$$h(\boldsymbol{\lambda}, \mathbf{y}) = \inf_{\mathbf{x} \geq \mathbf{0}, \mathbf{u}_i \in R^{n_i}} L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}),$$

ενώ το δυϊκό πρόβλημα που σχετίζεται με το ( $P_{CR}$ ) ορίζεται ως

$$(D_{CR}) \quad \sup_{\boldsymbol{\lambda} \in R_+^m, \mathbf{y}_i \in R^{n_i}} h(\boldsymbol{\lambda}, \mathbf{y}).$$

Με  $\boldsymbol{\lambda}$  συμβολίζουμε το διάνυσμα των Λαγκρανζιανών πολλαπλασιαστών, του πρωτεύοντος προβλήματος, που σχετίζεται με τους περιορισμούς  $\frac{1}{2} \mathbf{u}_i^T \mathbf{u}_i + \mathbf{b}_i^T \mathbf{x} + c_i \leq 0$  ενώ με  $\mathbf{y}$  συμβο-



λίζουμε το διάνυσμα των διανυσματικών Λαγκρανζιανών πολλαπλασιαστών που σχετίζονται με τους περιορισμούς  $\mathbf{A}_i \mathbf{x} = \mathbf{u}_i$ ,  $\mathbf{u}_i \in R^{n_i}$ ,  $i = 1, 2, \dots, m$ . Στη συνέχεια χρησιμοποιούμε το γεγονός ότι η Λαγκρανζιανή συνάρτηση που ορίζεται στην Εξίσωση 5.12 είναι διαχωρίσιμη (separable) σε δυο μεταβλητές απόφασης,  $\mathbf{x}$  και  $\mathbf{u}$  (Boyd and Vandenberghe, 2006), και βασιζόμενοι σε αυτό, εξάγουμε μια αναλυτική μορφή για τη δυϊκή αντικειμενική συνάρτηση  $h(\boldsymbol{\lambda}, \mathbf{y})$  (Teboulle, 1989) ως εξής:

$$h(\boldsymbol{\lambda}, \mathbf{y}) = \frac{1}{r(r+1)} \sum_{j=1}^n \inf_{x_j \geq 0} \left\{ x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right] + r(r+1) \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j x_j \right\} + \mathbf{c}^T \boldsymbol{\lambda} + \sum_{i=1}^m \inf_{\mathbf{u}_i \in R^{n_i}} \left\{ \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right\}. \quad (5.13)$$

Για το πρώτο infimum στην Εξίσωση 5.13, είναι εύκολο να δούμε (βλέπε Παράρτημα Α) ότι φτάνει τη μικρότερη τιμή του

$$\frac{1}{(r+1)^2} \sum_{j=1}^n d_j \left( \frac{1}{r+1} - r \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right)^{\frac{1}{r}} \left( r(r+1) \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j - 1 \right) \quad (5.14)$$

στο βέλτιστο σημείο (βλέπε Εξίσωση Α.2 του Παραρτήματος Α)

$$x_j^* = d_j \left[ \frac{1}{r+1} - r \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right]^{\frac{1}{r}}. \quad (5.15)$$

Για το δεύτερο infimum στην Εξίσωση 5.13, έχουμε ότι η συνθήκη βελτιστοποίησής της είναι η:  $\lambda_i \mathbf{u}_i - \mathbf{y}_i = \mathbf{0}$ ,  $i = 1, 2, \dots, m$ . Όποτε ισχύει  $\lambda_i > 0$ , τότε ισχύει ότι  $\mathbf{u}_i^* = \mathbf{y}_i / \lambda_i$  και το infimum στην Εξίσωση 5.13 φτάνει την ελάχιστη τιμή του

$$-\frac{1}{2} \sum_{i=1}^m \frac{\|\mathbf{y}_i\|^2}{\lambda_i}. \quad (5.16)$$

Όταν  $\lambda_i = 0$  και  $\mathbf{y}_i \neq \mathbf{0}$ , ισχύει ότι  $h(\boldsymbol{\lambda}, \mathbf{y}) = -\infty$ , και συνεπώς ένα τέτοιο  $\mathbf{y}_i$  δεν μπορεί να θεωρηθεί ως το supremum στη δυϊκή αντικειμενική συνάρτηση  $h(\boldsymbol{\lambda}, \mathbf{y})$ , οπότε στην περίπτωση αυτή θέτουμε  $\mathbf{y}_i = \mathbf{0}$ . Αντικαθιστώντας τις Εξισώσεις 5.14 και 5.16 στην Εξίσωση 5.13 παίρνουμε το επιθυμητό αποτέλεσμα. ■

**Θεώρημα 5.5.** (α) Εάν το πρόβλημα  $(P_{CR})$  είναι εφικτό (feasible) τότε το  $\inf(P_{CR})$  επιτυγχάνεται και ισχύει  $\min(P_{CR}) = \sup(D_{CR})$ . Επιπλέον, εάν υπάρχει ένα  $\mathbf{x} \in R^n$  που ικανοποιεί τις συνθήκες  $\mathbf{x} > \mathbf{0}$ ,  $g_i(\mathbf{x}) < 0$ ,  $i = 1, \dots, m$ , τότε το  $\sup(D_{CR})$  επιτυγχάνεται και ισχύει  $\min(P_{CR}) = \max(D_{CR})$ .

(β) Εάν το  $\mathbf{x}^*$  αποτελεί λύση του πρωτεύοντος προβλήματος ( $P_{CR}$ ) και το  $\mathbf{y}_i^* \in R^{n_i}$ ,  $\boldsymbol{\lambda}^* \in R_+^m$  είναι λύση του δυϊκού προβλήματος ( $D_{CR}$ ), τότε

$$x_j^* = d_j \left[ \frac{1}{r+1} - r \sum_{i=1}^m (\lambda_i^* \mathbf{b}_i^T + \mathbf{y}_i^{*T} \mathbf{A}_i)_j \right]^{\frac{1}{r}}, j = 1, 2, \dots, n.$$

Απόδειξη. Η απόδειξη του θεωρήματος μπορεί να γίνει μέσω τυπικών αποτελεσμάτων δυϊκότητας (βλέπε για παράδειγμα τους Laurent (1972), Rockafellar (1970) ή Boyd and Vandenberghe (2006)).

(α) Εξαιτίας της μη αρνητικότητας των περιορισμών, δηλαδή  $\boldsymbol{\lambda} \in R_+^m$ , του δυϊκού προβλήματος ( $D_{CR}$ ), αυτό ικανοποιεί την strongest constraint qualification, η οποία υπαινίσσεται την έλλειψη του δυϊκού χάσματος (lack of duality gap) και την πραγματοποίηση του πρωτεύοντος infimum. Συνεπώς, το πρώτο μέρος ακολουθεί άμεσα. Το δεύτερο μέρος ακολουθεί από τον ορισμό της δυϊκότητας.

(β) Η συνθήκη βελτιστοποίησης για το  $\mathbf{x} = \mathbf{x}^*$  να αποτελεί τη λύση της ελαχιστοποίησης της  $h(\boldsymbol{\lambda}, \mathbf{y})$  που δίνεται στην Εξίσωση 5.13, είναι η βέλτιστη λύση που δίνεται στην Εξίσωση 5.15 και συνεπώς το επιθυμητό αποτέλεσμα ακολουθεί.

Από το μέρος (α) έχουμε ότι υπάρχει ένα σαγματικό σημείο  $(\mathbf{x}^*, \mathbf{y}_i^*, \boldsymbol{\lambda}^*)$  και συνεπώς ισχύει  $\min_{\mathbf{x} \geq \mathbf{0}} L(\mathbf{x}, \mathbf{y}_i^*, \boldsymbol{\lambda}^*) = L(\mathbf{x}^*, \mathbf{y}_i^*, \boldsymbol{\lambda}^*)$  (Ben-Tal, 1979). ■

## 5.9 Αποτελέσματα Λαγκρανζιανής δυϊκότητας για τη διαφορά του Jensen

Στην ενότητα αυτή θα παρουσιάσουμε αποτελέσματα λαγκρανζιανής δυϊκότητας για τη διαφορά του Jensen.

Το πρόβλημα ελαχιστοποίησης της διαφοράς του Jensen υπό τετραγωνικούς περιορισμούς ορίζεται ως:

Εύρεση του  $\mathbf{x} \in R^n$ , το οποίο επιλύει το πρωτεύον πρόβλημα

$$(P_J) \quad \min - \sum_{j=1}^n \frac{1}{2} (x_j + d_j) \ln \left( \frac{1}{2} (x_j + d_j) \right) + \frac{1}{2} \left[ \sum_{j=1}^n x_j \ln x_j + \sum_{j=1}^n d_j \ln d_j \right]$$

υπό τους περιορισμούς

$$g_i(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{D}_i \mathbf{x} + \mathbf{b}_i^T \mathbf{x} + c_i \leq 0, i = 1, 2, \dots, m, \mathbf{x} \geq \mathbf{0},$$

όπου  $\mathbf{d} = (d_1, d_2, \dots, d_n)^T$  είναι ένα δοθέν διάνυσμα με αυστηρώς θετικά στοιχεία,  $\mathbf{D}_i$  είναι ένας δεδομένος θετικά ημιορισμένος πίνακας για κάθε  $i$ ,  $\mathbf{b}_i \in R^n$  και  $c_i$  είναι δεδομένες σταθερές όχι ταυτόχρονα ίσες με μηδέν.

Στη συνέχεια, θα προσπαθήσουμε να εξάγουμε τη δυϊκή αναπαράσταση του πρωτεύοντος προβλήματος ( $P_J$ ) με τη βοήθεια της λαγκρανζιανής δυϊκότητας. Και σε αυτή την περίπτωση,

αυτό θα επιτευχθεί μέσω μιας απλής διαμέρισης για τη μετατροπή του προβλήματος ( $P_J$ ) σε ένα ισοδύναμο κυρτό πρόγραμμα με γραμμικούς και τετραγωνικούς περιορισμούς.

Επειδή ο  $\mathbf{D}_i$  είναι ένας ημιθετικά ορισμένος  $n \times n$  πίνακας, μπορούμε να τον εκφράσουμε ως  $\mathbf{D}_i = \mathbf{A}_i^T \mathbf{A}_i$ , όπου  $\mathbf{A}_i$  είναι ένας  $n_i \times n$  πίνακας και  $n_i$  είναι η τάξη του  $\mathbf{D}_i$ ,  $i = 1, 2, \dots, m$ . Στην περίπτωση αυτή οι περιορισμοί μπορούν να γραφούν ως  $g_i(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A}_i^T \mathbf{A}_i \mathbf{x} + \mathbf{b}_i^T \mathbf{x} + c_i$ . Ορίζοντας τις νέες μεταβλητές  $\mathbf{u}_i = \mathbf{A}_i \mathbf{x}$ ,  $\mathbf{u}_i \in R^{n_i}$ ,  $i = 1, 2, \dots, m$ , το πρόβλημα ( $P_{CR}$ ) είναι αντίστοιχο με το ακόλουθο κυρτό πρόγραμμα με περιορισμούς γραμμικής ισότητας και τετραγωνικής ανισότητας:

$$(P_J^*) \quad \min_{\mathbf{x}, \mathbf{u}_i} - \sum_{j=1}^n \frac{1}{2} (x_j + d_j) \ln \left( \frac{1}{2} (x_j + d_j) \right) + \frac{1}{2} \left[ \sum_{j=1}^n x_j \ln x_j + \sum_{j=1}^n d_j \ln d_j \right]$$

υπό τους περιορισμούς

$$\frac{1}{2} \mathbf{u}_i^T \mathbf{u}_i + \mathbf{b}_i^T \mathbf{x} + c_i \leq 0, \quad \mathbf{A}_i \mathbf{x} = \mathbf{u}_i, \quad \mathbf{u}_i \in R^{n_i}, \quad i = 1, 2, \dots, m, \quad \mathbf{x} \geq \mathbf{0}.$$

Έστω  $\mathbf{u} = (u_1, u_2, \dots, u_m)^T \in R^{n_1} \times \dots \times R^{n_m}$ ,  $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)^T \in R^{n_1} \times \dots \times R^{n_m}$  και  $N = n_1 + \dots + n_m$ .

**Θεώρημα 5.6.** Το λαγκρανζιανό δυϊκό πρόβλημα του ( $P_J$ ) δίνεται από το

$$(D_J) \quad \sup_{\lambda \in R_+^m, \mathbf{y}_i \in R^{n_i}} \left\{ - \sum_{j=1}^n \frac{d_j}{2e^{2s_j} - 1} \left[ e^{2s_j} \ln \left( \frac{d_j e^{2s_j}}{2e^{2s_j} - 1} \right) - \frac{1}{2} \ln \left( \frac{d_j}{2e^{2s_j} - 1} \right) - s_j \right] - \frac{1}{2} \sum_{i=1}^m \frac{\|\mathbf{y}_i\|^2}{\lambda_i} + \boldsymbol{\lambda}^T \mathbf{c} + \mathbf{d}^T \mathbf{z} \right\},$$

όπου  $\mathbf{z}^T = (\ln d_1, \ln d_2, \dots, \ln d_n)$

Απόδειξη. Η λαγκρανζιανή συνάρτηση για το πρόβλημα ( $P_J$ ) είναι η

$$\begin{aligned} L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) &= - \sum_{j=1}^n \frac{1}{2} (x_j + d_j) \ln \left( \frac{1}{2} (x_j + d_j) \right) + \frac{1}{2} \left[ \sum_{j=1}^n x_j \ln x_j + \sum_{j=1}^n d_j \ln d_j \right] \\ &\quad + \sum_{i=1}^m \lambda_i \left( \frac{1}{2} \mathbf{u}_i^T \mathbf{u}_i + \mathbf{b}_i^T \mathbf{x} + c_i \right) + \sum_{i=1}^m \mathbf{y}_i^T (\mathbf{A}_i \mathbf{x} - \mathbf{u}_i) \\ &= - \sum_{j=1}^n \frac{1}{2} (x_j + d_j) \ln \left( \frac{1}{2} (x_j + d_j) \right) + \frac{1}{2} \left[ \sum_{j=1}^n x_j \ln x_j + \sum_{j=1}^n d_j \ln d_j \right] \\ &\quad + \sum_{i=1}^m \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i + \sum_{i=1}^m \lambda_i \mathbf{b}_i^T \mathbf{x} + \sum_{i=1}^m \lambda_i c_i + \sum_{i=1}^m \mathbf{y}_i^T \mathbf{A}_i \mathbf{x} - \sum_{i=1}^m \mathbf{y}_i^T \mathbf{u}_i \\ &= - \sum_{j=1}^n \frac{1}{2} (x_j + d_j) \ln \left( \frac{1}{2} (x_j + d_j) \right) + \frac{1}{2} \left[ \sum_{j=1}^n x_j \ln x_j + \sum_{j=1}^n d_j \ln d_j \right] \\ &\quad + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i) \mathbf{x} + \boldsymbol{\lambda}^T \mathbf{c} + \sum_{i=1}^m \left( \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right), \end{aligned}$$

όπου  $\mathbf{y}_i \in R^{n_i}$ ,  $i = 1, 2, \dots, m$  είναι διανυσματικοί λαγκρανζιανοί πολλαπλασιαστές ενώ η λαγκρανζιανή δυϊκή αντικειμενική συνάρτηση του προβλήματος  $(P_J)$  δίνεται από την

$$h(\boldsymbol{\lambda}, \mathbf{y}) = \inf_{\mathbf{x} \geq \mathbf{0}, \mathbf{u}_i \in R^{n_i}} L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}).$$

Το δυϊκό πρόβλημα που σχετίζεται με το  $(P_J)$  ορίζεται ως

$$(D_J) \quad \sup_{\boldsymbol{\lambda} \in R_+^m, \mathbf{y}_i \in R^{n_i}} h(\boldsymbol{\lambda}, \mathbf{y}).$$

Με  $\boldsymbol{\lambda}$  ορίζουμε το διάνυσμα των πολλαπλασιαστών Lagrange του πρωτεύοντος προβλήματος που σχετίζεται με τους περιορισμούς  $\frac{1}{2}\mathbf{u}_i^T \mathbf{u}_i + \mathbf{b}_i^T \mathbf{x} + c_i \leq 0$  ενώ με  $\mathbf{y}$  συμβολίζουμε το διάνυσμα των διανυσματικών πολλαπλασιαστών Lagrange που σχετίζονται με τους περιορισμούς  $\mathbf{A}_i \mathbf{x} = \mathbf{u}_i$ ,  $\mathbf{u}_i \in R^{n_i}$ ,  $i = 1, 2, \dots, m$ . Στη συνέχεια χρησιμοποιούμε το γεγονός ότι η λαγκρανζιανή συνάρτηση είναι διαχωρίσιμη στις δυο μεταβλητές απόφασης,  $\mathbf{x}$  και  $\mathbf{u}$  (Boyd and Vandenberghe, 2006), και βασιζόμενοι σε αυτό εξάγουμε την ακριβή μορφή της δυϊκής αντικειμενικής συνάρτησης  $h(\boldsymbol{\lambda}, \mathbf{y})$  (Teboulle, 1989) ως εξής:

$$\begin{aligned} h(\boldsymbol{\lambda}, \mathbf{y}) &= - \sum_{j=1}^n \inf_{x_j \geq 0} \left\{ \frac{1}{2}(x_j + d_j) \ln \left( \frac{1}{2}(x_j + d_j) \right) + \frac{1}{2}x_j \ln x_j + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j x_j \right\} \\ &\quad + \sum_{i=1}^m \inf_{\mathbf{u}_i \in R^{n_i}} \left\{ \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right\} + \mathbf{c}^T \boldsymbol{\lambda} + \mathbf{d}^T \mathbf{z}. \end{aligned} \quad (5.17)$$

Ας συμβολίσουμε το πρώτο μέρος της δεξιάς πλευράς της  $L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y})$  με

$$f(\mathbf{x}; \boldsymbol{\lambda}, \mathbf{y}) = - \sum_{j=1}^n \frac{1}{2}(x_j + d_j) \ln \left( \frac{1}{2}(x_j + d_j) \right) + \frac{1}{2} \sum_{j=1}^n x_j \ln x_j + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i) \mathbf{x}. \quad (5.18)$$

Έχουμε ότι

$$\begin{aligned} \frac{\partial}{\partial x_j} f(\mathbf{x}; \boldsymbol{\lambda}, \mathbf{y}) &= -\frac{1}{2} \left\{ (x_j + d_j)' \ln \left( \frac{1}{2}(x_j + d_j) \right) \right. \\ &\quad \left. + (x_j + d_j) \left( \ln \left( \frac{1}{2}(x_j + d_j) \right) \right)' \right\} \\ &\quad + \frac{1}{2} \left[ (x_j)' \ln x_j + x_j (\ln x_j)' \right] + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \\ &= -\frac{1}{2} \left\{ \ln \left( \frac{1}{2}(x_j + d_j) \right) + (x_j + d_j) \frac{2}{x_j + d_j} \frac{1}{2} \right\} \\ &\quad + \frac{1}{2} (\ln x_j + 1) + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \\ &= -\frac{1}{2} \ln \left( \frac{1}{2}(x_j + d_j) \right) + \frac{1}{2} \ln x_j + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j, \end{aligned}$$

όπου ο τόνος ( $\dot{\phantom{x}}$ ) συμβολίζει την παράγωγο. Για να βρούμε το βέλτιστο σημείο, θέτουμε την παραπάνω ισότητα ίση με μηδέν, οπότε έχουμε

$$\begin{aligned}
\frac{\partial}{\partial x_j} f(\mathbf{x}; \boldsymbol{\lambda}, \mathbf{y}) &= 0 \Leftrightarrow \\
-\frac{1}{2} \ln \left( \frac{1}{2}(x_j + d_j) \right) + \frac{1}{2} \ln x_j &= - \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \Leftrightarrow \\
\frac{x_j + d_j}{2x_j} &= \exp \left\{ 2 \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right\} \Leftrightarrow \\
d_j &= 2x_j \exp \left\{ 2 \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right\} - x_j \Leftrightarrow \\
d_j &= x_j \left( 2 \exp \left\{ 2 \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right\} - 1 \right) \Leftrightarrow \\
x_j &= \frac{d_j}{2 \exp \left\{ 2 \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right\} - 1}. \tag{5.19}
\end{aligned}$$

Αντικαθιστώντας τη Σχέση 5.19 στην  $L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y})$ , και θέτοντας  $s_j = \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j$ ,  $j = 1, 2, \dots, n$  έχουμε ότι

$$\begin{aligned}
& - \sum_{j=1}^n \frac{1}{2} \left( \frac{d_j}{2e^{2s_j} - 1} + d_j \right) \ln \left( \frac{1}{2} \left( \frac{d_j}{2e^{2s_j} - 1} + d_j \right) \right) \\
& + \frac{1}{2} \sum_{j=1}^n \frac{d_j}{2e^{2s_j} - 1} \ln \left( \frac{d_j}{2e^{2s_j} - 1} \right) + \sum_{j=1}^n s_j \frac{d_j}{2e^{2s_j} - 1} \\
& = - \sum_{j=1}^n \frac{d_j e^{2s_j}}{2e^{2s_j} - 1} \ln \left( \frac{d_j e^{2s_j}}{2e^{2s_j} - 1} \right) + \frac{1}{2} \sum_{j=1}^n \frac{d_j}{2e^{2s_j} - 1} \ln \left( \frac{d_j}{2e^{2s_j} - 1} \right) + \sum_{j=1}^n \frac{d_j s_j}{2e^{2s_j} - 1} \\
& = - \sum_{j=1}^n \left\{ \frac{d_j e^{2s_j}}{2e^{2s_j} - 1} \ln \left( \frac{d_j e^{2s_j}}{2e^{2s_j} - 1} \right) - \frac{1}{2} \frac{d_j}{2e^{2s_j} - 1} \ln \left( \frac{d_j}{2e^{2s_j} - 1} \right) - \frac{d_j s_j}{2e^{2s_j} - 1} \right\} \\
& = - \sum_{j=1}^n \frac{d_j}{2e^{2s_j} - 1} \left\{ e^{2s_j} \ln \left( \frac{d_j e^{2s_j}}{2e^{2s_j} - 1} \right) - \frac{1}{2} \ln \left( \frac{d_j}{2e^{2s_j} - 1} \right) - s_j \right\},
\end{aligned}$$

το οποίο είναι η ελάχιστη τιμή του πρώτου infimum στη Σχέση 5.17.

Θέτοντας  $g(\mathbf{u}; \boldsymbol{\lambda}, \mathbf{y})$  για τον τελευταίο όρο της λαγκρανζιανής συνάρτησης  $L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y})$ , δηλαδή

$$g(\mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) = \sum_{i=1}^m \left( \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right) \quad (5.20)$$

έχουμε ότι

$$\frac{\partial}{\partial \mathbf{u}_i} g(\mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) = \lambda_i \mathbf{u}_i - \mathbf{y}_i.$$

Για να βρούμε το βέλτιστο σημείο, θέτουμε την παραπάνω εξίσωση ίση με μηδέν, οπότε έχουμε ότι

$$\begin{aligned} \frac{\partial}{\partial \mathbf{u}_i} g(\mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) &= \mathbf{0} \Leftrightarrow \\ \lambda_i \mathbf{u}_i - \mathbf{y}_i &= \mathbf{0} \Leftrightarrow \\ \lambda_i \mathbf{u}_i &= \mathbf{y}_i, \end{aligned}$$

το οποίο σημαίνει

$$\mathbf{u}_i = \frac{1}{\lambda_i} \mathbf{y}_i. \quad (5.21)$$

Αντικαθιστώντας τη Σχέση 5.21 στη Σχέση 5.20 έχουμε ότι

$$\begin{aligned} \sum_{i=1}^m \left( \frac{1}{2} \lambda_i \left( \frac{1}{\lambda_i} \mathbf{y}_i \right)^T \left( \frac{1}{\lambda_i} \mathbf{y}_i \right) - \mathbf{y}_i^T \left( \frac{1}{\lambda_i} \mathbf{y}_i \right) \right) &= \sum_{i=1}^m \left( \frac{1}{2} \frac{1}{\lambda_i} \mathbf{y}_i^T \mathbf{y}_i - \frac{1}{\lambda_i} \mathbf{y}_i^T \mathbf{y}_i \right) \\ &= \sum_{i=1}^m \frac{1}{\lambda_i} \mathbf{y}_i^T \mathbf{y}_i \left( \frac{1}{2} - 1 \right) \\ &= -\frac{1}{2} \sum_{i=1}^m \frac{1}{\lambda_i} \mathbf{y}_i^T \mathbf{y}_i \\ &= -\frac{1}{2} \sum_{i=1}^m \frac{\|\mathbf{y}_i\|^2}{\lambda_i}, \end{aligned}$$

το οποίο είναι η ελάχιστη τιμή του δεύτερου infimum στη Σχέση 5.17. ■

**Θεώρημα 5.7.** (a) Εάν το πρόβλημα  $(P_J)$  είναι εφικτό τότε το  $\inf(P_J)$  επιτυγχάνεται και ισχύει η σχέση  $\min(P_J) = \sup(D_J)$ . Επιπλέον, εάν υπάρχει ένα  $\mathbf{x} \in R^n$  που ικανοποιεί τις συνθήκες  $\mathbf{x} > \mathbf{0}$ ,  $g_i(\mathbf{x}) < 0$ ,  $i = 1, \dots, m$ , τότε το  $\sup(D_J)$  επιτυγχάνεται και  $\min(P_J) = \max(D_J)$ .

(b) Εάν το  $\mathbf{x}^*$  αποτελεί λύση του πρωτεύοντος προβλήματος  $(P_J)$  και το  $\mathbf{y}_i^* \in R^{n_i}$ ,  $\boldsymbol{\lambda}^* \in R_+^m$  είναι λύση του δυϊκού προβλήματος  $(D_J)$ , τότε

$$x_j^* = \frac{d_j}{2 \exp \left\{ 2 \sum_{i=1}^m (\lambda_i^* \mathbf{b}_i^T + \mathbf{y}_i^{*T} \mathbf{A}_i)_j \right\} - 1}.$$

Απόδειξη. Η απόδειξη του θεωρήματος γίνεται μέσω τυπικών αποτελεσμάτων δυϊκότητας (βλέπε για παράδειγμα Laurent (1972), Rockafellar (1970) ή Boyd and Vandenberghe (2006)).

(a) Εξαιτίας της μη αρνητικότητας των περιορισμών, δηλαδή  $\boldsymbol{\lambda} \in R_+^m$ , του δυϊκού προβλήματος ( $D_J$ ), αυτό ικανοποιεί την strongest constraint qualification, η οποία υπαινίσσεται την έλλειψη του δυϊκού χάσματος και την εξασφάλιση του πρωτεύοντος infimum. Συνεπώς το πρώτο μέλος ακολουθεί άμεσα. Το δεύτερο μέλος ακολουθεί από τον ορισμό της δυϊκότητας.

(b) Η συνθήκη βελτιστοποίησης για το  $\mathbf{x} = \mathbf{x}^*$  να αποτελεί τη λύση της ελαχιστοποίησης της  $h(\boldsymbol{\lambda}, \mathbf{y})$  που δίνεται από τη Σχέση 5.17 είναι η βέλτιστη λύση

$$x_j^* = \frac{d_j}{2 \exp \left\{ 2 \sum_{i=1}^m (\lambda_i^* \mathbf{b}_i^T + \mathbf{y}_i^{*T} \mathbf{A}_i)_j \right\} - 1}.$$

και συνεπώς το επιθυμητό αποτέλεσμα ακολουθεί.

Από το μέρος (a) έχουμε ότι υπάρχει ένα σαγματικό σημείο  $(\mathbf{x}^*, \mathbf{y}_i^*, \boldsymbol{\lambda}^*)$  και συνεπώς ισχύει  $\min_{\mathbf{x} \geq \mathbf{0}} L(\mathbf{x}, \mathbf{y}_i^*, \boldsymbol{\lambda}^*) = L(\mathbf{x}^*, \mathbf{y}_i^*, \boldsymbol{\lambda}^*)$  (Ben-Tal and Charnes, 1979). ■

## 5.10 Συμπεράσματα Κεφαλαίου

Αντικείμενο του παρόντος κεφαλαίου, ήταν η παρουσίαση τρόπων ελαχιστοποίησης των μέτρων απόκλισης υπό διάφορους περιορισμούς και η εφαρμογή τους στην ελαχιστοποίηση των μέτρων των Kullback - Leibler, των Cressie and Read και της διαφοράς του Jensen υπό τετραγωνικούς περιορισμούς.

Στην περίπτωση που έχουμε γραμμικούς και τετραγωνικούς περιορισμούς ανισότητας, η επίλυση του προβλήματος ελαχιστοποίησης των μέτρων απόκλισης είναι δύσκολη, λόγω του μεγάλου όγκου των περιορισμών. Βέβαια στις μέρες μας όπου η τεχνολογία έχει εξελιχθεί σε σημαντικό βαθμό, είναι εύκολο μέσω του ηλεκτρονικού υπολογιστή να επιλύσουμε σε σύντομο χρονικό διάστημα προβλήματα τέτοιου είδους. Παρόλα αυτά, από μαθηματικής άποψης κρίθηκε σκόπιμο να μελετήσουμε τα μέτρα απόκλισης από την πλευρά της Λαγκρανζιανής δυϊκότητας. Στο τέλος του κεφαλαίου δώσαμε ιδιαίτερη βάση στην εξαγωγή ορισμένων δυϊκών αποτελεσμάτων για το πρόβλημα της ελαχιστοποίησης της απόκλισης των Cressie - Read τάξης  $r$  και της διαφοράς του Jensen υπό τετραγωνικούς και γραμμικούς περιορισμούς ανισότητας. Το σημαντικότερο εύρημα της μελέτης μας είναι, πέραν της λύσης του δυϊκού προβλήματος, ότι το δυϊκό πρόβλημα και στις δυο περιπτώσεις αποδείχθηκε ότι δεν έχει περιορισμούς. Συνεπώς είναι πολύ πιο εύκολο να εφαρμοσθεί. Τα αποτελέσματα αυτά είναι σημαντικά για την αναλογιστική επιστήμη και ειδικότερα για το πρόβλημα της εξομάλυνσης

των ποσοστών θνησιμότητας.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΠΑ



## ΚΕΦΑΛΑΙΟ 6

# ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΕΞΟΜΑΛΥΝΣΗ ΠΟΣΟΣΤΩΝ ΘΝΗΣΙΜΟΤΗΤΑΣ

### 6.1 Εισαγωγή

Το πρόβλημα της εξομάλυνσης, όπως έχουμε ήδη αναφέρει, ως σκοπό έχει να βρει τις τιμές  $u_x$ , οι οποίες ικανοποιούν τους μαθηματικούς και αναλογιστικούς περιορισμούς (i) έως (v) που παρουσιάσαμε στην Ενότητα 4.6, και είναι οι ελάχιστα διαχωρίσιμες από τις αρχικές εκτιμήσεις  $u_x$ , οι οποίες αποτελούν κατά κανόνα μια ανώμαλη σειρά με πολλές διακυμάνσεις. Στο κεφάλαιο αυτό θα εφαρμόσουμε την υπό περιορισμούς ελαχιστοποίηση της απόκλισης των Cressie - Read και της διαφοράς του Jensen και θα συγκρίνουμε τα αποτελέσματά της με αντίστοιχα που προκύπτουν από την ελαχιστοποίηση της απόκλισης των Kullback - Leibler και της γνωστής μεθόδου εξομάλυνσης των Whittaker and Henderson. Σημειώνουμε ότι η ελαχιστοποίηση των μέτρων απόκλισης μπορεί εύκολα να γίνει με οποιονδήποτε έτοιμο κώδικα μη γραμμικού προγραμματισμού. Στην παρούσα διατριβή, για την εξαγωγή των αποτελεσμάτων, χρησιμοποιήσαμε το πρόγραμμα LINGO ver. 10.

Η εξομάλυνση των ποσοστών θνησιμότητας εξακολουθεί να τυγχάνει την προσοχή των ερευνητών. Πέρα από τα κλασσικά βιβλία των Benjamin and Pollard (1980) και London (1985), εξαιρετικά άρθρα ανασκόπησης είναι αυτά των Haberman (1998) και Wang (1998). Στις εφαρμογές οι πλέον δημοφιλείς μέθοδοι είναι αυτές των Helligman - Pollard, Whittaker και Henderson, οι συναρτήσεις splines και οι εκτιμητές πυρήνα. Ορισμένα πρόσφατα άρθρα είναι αυτά του Broffitt (1996), ο οποίος εξέτασε την πολυδιάστατη Whittaker εξομάλυνση, των Haberman and Renshaw (1999), οι οποίοι περιέγραψαν μια μέθοδο προσαρμογής καμπυλών στις διαφορές των λογαριθμημένων αδρών ποσοστών θνησιμότητας (log crude mortality) παρακινήμενοι από τη λεγόμενη προσέγγιση της διϊκής μοντελοποίησης (dual modelling approach) στην εξομάλυνση που περιέγραψαν οι Renshaw et al. (1996). Ο Nielsen (2003) ανασκόπησε ορισμένα άρθρα σχετικά με την εξομάλυνση και πρόβλεψη

μέσω εκτιμητών πυρήνα με εφαρμογές στην Αναλογιστική Επιστήμη, τη Βιοστατιστική και τα Χρηματοοικονομικά. Οι Debon et al. (2006) συνέκριναν μη παραμετρικές μεθόδους εξομάλυνσης και έβγαλαν το συμπέρασμα ότι η καλύτερη μέθοδος είναι η εξομάλυνση μέσω Γενικευμένων Προσθετικών Μοντέλων (GAM) με συναρτήσεις splines. Οι Neves and Migon (2007) παρουσίασαν μοντέλα μπεϋζιανής εξομάλυνσης, χρησιμοποιώντας τεχνικές Markov Chain Monte Carlo (MCMC).

Για τη διερεύνησή μας, θα χρησιμοποιήσουμε τρία διαφορετικά σύνολα δεδομένων με πιθανότητες θανάτου. Το πρώτο σύνολο προέρχεται από τον London (1985), αρχικά από τον Miller (1949), και θα το συμβολίζουμε με L85. Το δεύτερο προέρχεται από την Ένωση Αναλογιστών του Χόνγκ Κόνγκ (Actuarial Society of Hong Kong) αναφέρεται σε άντρες ασφαλισμένους για περίοδο μεγαλύτερη των 2 ετών και θα συμβολίζεται με HK01M. Το τρίτο σύνολο προέρχεται από την ίδια Κοινότητα, αναφέρεται σε γυναίκες ασφαλισμένες για περίοδο μεγαλύτερη των 2 ετών και θα συμβολίζεται με HK01F. Τα παραπάνω σύνολα δεδομένων είναι διαφορετικού μεγέθους. Ειδικότερα, το σύνολο L85 αποτελείται από 20 πιθανότητες θανάτου (υπολογισμένες από συνολικά 79880 παρατηρήσεις) που ανήκουν στις ηλικίες 75 με 94 ετών. Από το HK01M έχουμε χρησιμοποιήσει 16 πιθανότητες θανάτου (υπολογισμένες από συνολικά 13678 παρατηρήσεις) για τις ηλικίες 70 με 85 ετών ενώ από το σύνολο HK01F έχουμε χρησιμοποιήσει 20 πιθανότητες θανάτου (υπολογισμένες από συνολικά 18341 παρατηρήσεις) για τις ηλικίες 70 με 89 ετών (Sachlas and Papaioannou, 2008, 2009a).

Θα ξεκινήσουμε την ανάλυσή μας, με την ελαχιστοποίηση της απόκλισης δύναμης των Cressie and Read υπό τους πέντε περιορισμούς που πρότειναν οι Brockett and Zhang. Στη συνέχεια, θα εφαρμόσουμε την ελαχιστοποίηση του ίδιου μέτρου προσθέτοντας τον περιορισμό που προτείναμε στην Ενότητα 4.6 ενώ στη συνέχεια, από τους έξι περιορισμούς θα αφαιρέσουμε τους δυο αναλογιστικούς περιορισμούς. Στη συνέχεια θα ελαχιστοποιήσουμε τη διαφορά του Jensen πάλι υπό τις τρεις περιπτώσεις των περιορισμών. Επιπλέον θα διερευνήσουμε τη συμπεριφορά της εκτεταμένης απόκλισης των Cressie and Read, την οποία αναφέραμε στην Ενότητα 3.3.

## 6.2 Ελαχιστοποίηση απόκλισης C - R υπό 5 περιορισμούς

Έχουμε διεξάγει αρκετές εξομαλύνσεις για κάθε σύνολο δεδομένων, χρησιμοποιώντας διαφορετική τιμή για την παράμετρο  $\lambda$  και τους περιορισμούς της ομαλότητας, μονοτονίας, κυρτότητας και τους δυο αναλογιστικούς περιορισμούς. Μεταξύ αυτών είναι οι τιμές 1, 0,  $-1$ ,  $-(1/2)$  και  $-2$ , οι οποίες δίνουν τη στατιστική συνάρτηση  $\chi^2$  του Pearson, την απόκλιση των Kullback - Leibler, τη τροποποιημένη στατιστική συνάρτηση του λόγου πιθανοφάνειας, τη στατιστική συνάρτηση των Freeman - Tukey  $F^2$  και την τροποποιημένη στατιστική συ-

Πίνακας 6.1: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων L85 (5 περιορισμοί)

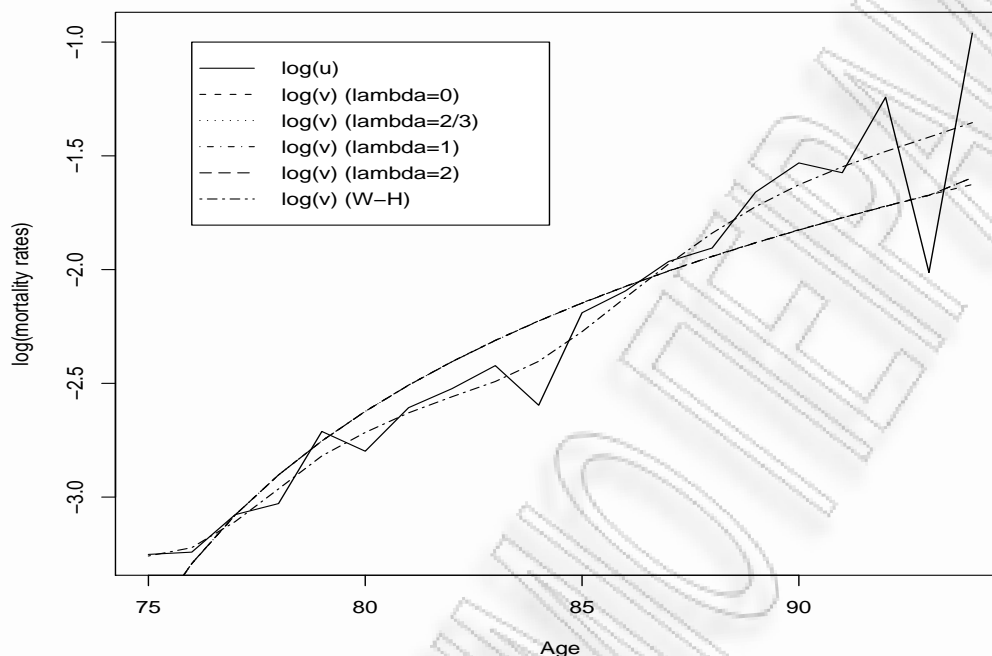
$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
75	0.03870	0.02828	0.02833	0.02833	0.02833	0.03844
76	0.03909	0.03715	0.03718	0.03718	0.03718	0.03987
77	0.04609	0.04601	0.04603	0.04603	0.04603	0.04470
78	0.04839	0.05488	0.05488	0.05488	0.05488	0.05161
79	0.06644	0.06374	0.06372	0.06372	0.06372	0.05967
80	0.06091	0.07261	0.07257	0.07257	0.07257	0.06615
81	0.07376	0.08147	0.08142	0.08142	0.08142	0.07210
82	0.08019	0.09034	0.09027	0.09027	0.09027	0.07739
83	0.08874	0.09920	0.09911	0.09911	0.09911	0.08276
84	0.07455	0.10807	0.10796	0.10796	0.10796	0.09044
85	0.11195	0.11693	0.11681	0.11681	0.11681	0.10302
86	0.12320	0.12580	0.12566	0.12566	0.12566	0.11965
87	0.14031	0.13466	0.13450	0.13450	0.13450	0.13879
88	0.14879	0.14353	0.14335	0.14335	0.14335	0.15886
89	0.19020	0.15239	0.15220	0.15220	0.15220	0.17847
90	0.21628	0.16126	0.16104	0.16104	0.16105	0.19650
91	0.20718	0.17012	0.16989	0.16989	0.16989	0.21261
92	0.28866	0.17899	0.17874	0.17874	0.17874	0.22747
93	0.13362	0.18785	0.18759	0.18759	0.18759	0.24224
94	0.38265	0.19672	0.20276	0.20276	0.20274	0.25817

νάρτηση του Neyman  $\chi^2$ , αντίστοιχα. Έχουμε επιπλέον χρησιμοποιήσει την τιμή 2/3 που πρότειναν οι Cressie and Read (1984). Σημειώνουμε ότι η τιμή της παραμέτρου  $M$  στον πρώτο περιορισμό, είναι διαφορετική σε κάθε σύνολο δεδομένων, και έχει υπολογιστεί μέσω εξομάλυνσης με τη μέθοδο των Whittaker - Henderson, με  $h$  τη μέση τιμή των βαρών  $w_x$ , οι οποίες είναι  $h = 80786,8$  για το σύνολο L85,  $h = 29354,6$  για το σύνολο HK01M και  $h = 54547$  για το σύνολο HK01F. Οι κώδικες που γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, για την εύρεση των εξομαλυμένων τιμών, παρατίθενται στο Παράρτημα Γ.1.

Τα αποτελέσματα για  $\lambda = 0, 2/3, 1, 2$  και τη μέθοδο των Whittaker - Henderson δίνονται στους Πίνακες 6.1, 6.2 και 6.3 για τα σύνολο L85, HK01M και HK01F, αντίστοιχα. Στα Σχήματα 6.1, 6.2 και 6.3 δίνουμε τα αποτελέσματα κάθε εξομάλυνσης σε λογαριθμική κλίμακα. Οι εξομαλύνσεις μέσω της απόκλισης των Cressie - Read δίνουν ισοδύναμα αποτελέσματα με αυτά της Whittaker - Henderson μεθόδου, η οποία είναι μια ευρέως αποδεκτή μέθοδος εξομάλυνσης (London, 1985).

Είναι αναμενόμενο και λογικό ότι διαφορετικές επιλογές για την τιμή της παραμέτρου  $\lambda$  οδηγούν σε διαφορετικές εξομαλυμένες τιμές. Όπως έχουμε ήδη αναφέρει, τα δυο βασικά στοιχεία της εξομάλυνσης είναι η ομαλότητα και η καλή προσαρμογή. Συνεπώς, για να συγκρίνουμε τα αποτελέσματα των διάφορων εξομαλύνσεων που έχουμε διεξάγει, υπολογίσαμε,

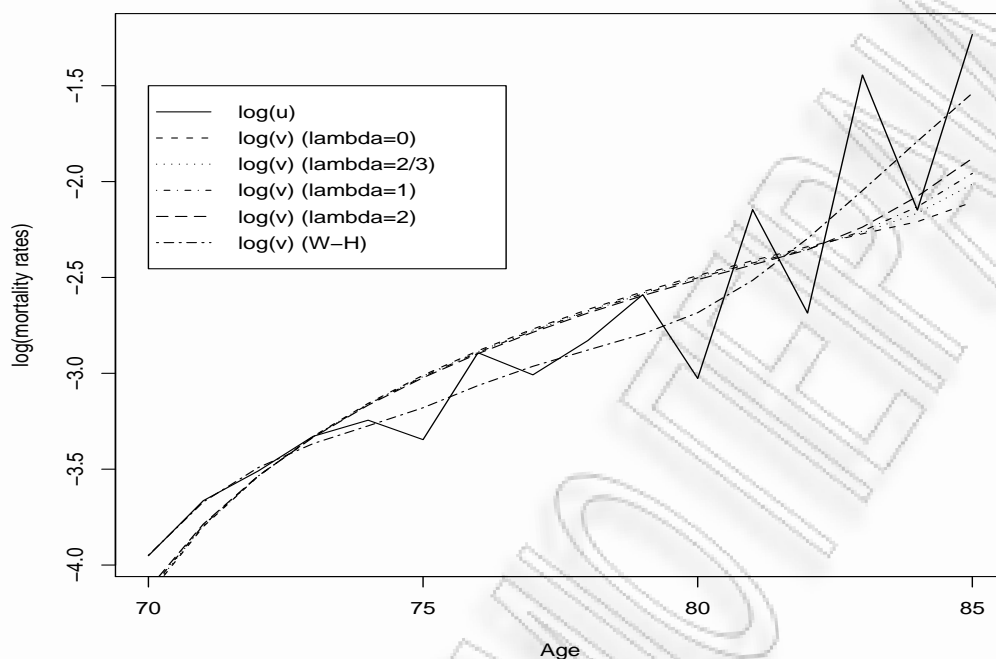
Σχήμα 6.1: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (L85 - 5 περιορισμοί)



Πίνακας 6.2: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων ΗΚ01Μ (5 περιορισμοί)

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
70	0.01923	0.01567	0.01583	0.01596	0.01618	0.01928
71	0.02563	0.02239	0.02248	0.02255	0.02268	0.02546
72	0.02992	0.02911	0.02913	0.02915	0.02917	0.03045
73	0.03585	0.03584	0.03578	0.03574	0.03567	0.03451
74	0.03899	0.04256	0.04243	0.04234	0.04217	0.03785
75	0.03523	0.04928	0.04908	0.04894	0.04867	0.04160
76	0.05543	0.05601	0.05573	0.05553	0.05516	0.04671
77	0.04939	0.06273	0.06239	0.06213	0.06166	0.05164
78	0.05906	0.06945	0.06904	0.06872	0.06816	0.05621
79	0.07503	0.07617	0.07569	0.07532	0.07466	0.06110
80	0.04848	0.08290	0.08234	0.08192	0.08115	0.06830
81	0.11692	0.08962	0.08899	0.08851	0.08765	0.08085
82	0.06816	0.09634	0.09564	0.09511	0.09483	0.10057
83	0.23598	0.10307	0.10365	0.10454	0.10674	0.12914
84	0.11659	0.10979	0.11477	0.11869	0.12536	0.16700
85	0.29152	0.12154	0.13346	0.14123	0.15274	0.21458

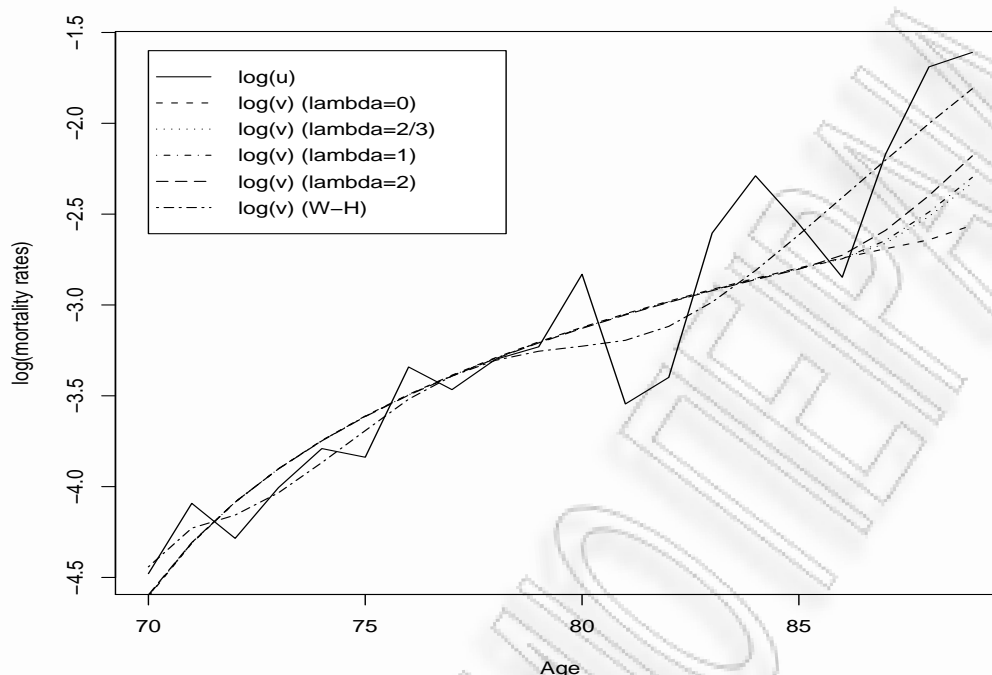
Σχήμα 6.2: Αδρές  $u_x$  και εξομαλυσμένες  $v_x$  τιμές (HK01M - 5 περιορισμοί)



Πίνακας 6.3: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01F (5 περιορισμοί)

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
70	0.01134	0.01000	0.01005	0.01005	0.01009	0.01178
71	0.01670	0.01341	0.01343	0.01344	0.01346	0.01457
72	0.01378	0.01681	0.01682	0.01682	0.01683	0.01566
73	0.01822	0.02021	0.02021	0.02020	0.02020	0.01771
74	0.02258	0.02362	0.02359	0.02359	0.02356	0.02091
75	0.02153	0.02702	0.02698	0.02697	0.02693	0.02492
76	0.03539	0.03042	0.03037	0.03035	0.03030	0.02964
77	0.03124	0.03383	0.03375	0.03374	0.03367	0.03366
78	0.03688	0.03723	0.03714	0.03712	0.03704	0.03669
79	0.03954	0.04063	0.04053	0.04050	0.04040	0.03859
80	0.05896	0.04404	0.04391	0.04389	0.04377	0.03968
81	0.02887	0.04744	0.04730	0.04727	0.04714	0.04098
82	0.03342	0.05084	0.05069	0.05065	0.05051	0.04426
83	0.07395	0.05425	0.05407	0.05404	0.05387	0.05057
84	0.10138	0.05765	0.05746	0.05742	0.05724	0.06018
85	0.07776	0.06105	0.06085	0.06080	0.06061	0.07324
86	0.05800	0.06446	0.06423	0.06419	0.06547	0.08995
87	0.11429	0.06786	0.06991	0.07100	0.07509	0.11056
88	0.18465	0.07126	0.08101	0.08302	0.09090	0.13513
89	0.20000	0.07723	0.09748	0.10053	0.11314	0.16371

Σχήμα 6.3: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (HK01F - 5 περιορισμοί)



μετά την εξομάλυνση, το μέτρο  $F = \sum_{x=1}^n w_x (u_x - v_x)^2$ , που χρησιμοποιούν οι Whittaker - Henderson. Σημειώνουμε ότι σαν βάρη χρησιμοποιήσαμε τα  $w_x = \frac{l_x}{v_x(1-v_x)}$ , όπου  $l_x$  είναι ο αριθμός των ατόμων σε κίνδυνο στην ηλικία  $x$ . Το μέτρο  $S = \sum_{x=1}^{n-3} (\Delta^3 v_x)^2$  χρησιμοποιήθηκε για να μετρήσει την ομαλότητα των εξομαλυμένων τιμών.

Επιπρόσθετα, η συμπεριφορά - απόδοση των μεθόδων που προτείνουμε, θα αξιολογηθεί μέσω των στατιστικών συναρτήσεων log - likelihood, deviance και του  $\chi^2$  καλής προσαρμογής. Υποθέτοντας ότι οι θάνατοι  $d_x$  ακολουθούν τη διωνυμική κατανομή,  $Bi(l_x, q_x)$ , όπου  $q_x$  είναι η πραγματική αλλά άγνωστη πιθανότητα θανάτου για την ηλικία  $x$ , έχουμε ότι η log - likelihood εξαιρώντας τις σταθερές ισούται με

$$\log L(q) = \sum_{i=1}^n [d_x \log q_x + (l_x - d_x) \log(1 - q_x)].$$

Γνωρίζοντας τη συνάρτηση log - likelihood μπορούμε να υπολογίσουμε τη deviance,

$$D(v) = 2 \log L(u) - 2 \log L(v).$$

Τέλος, μπορούμε να μετρήσουμε την απόσταση μεταξύ των παρατηρούμενων και των ανα-

Πίνακας 6.4: Τιμές των μέτρων ομαλότητας και προσαρμογής (5 περιορισμοί)

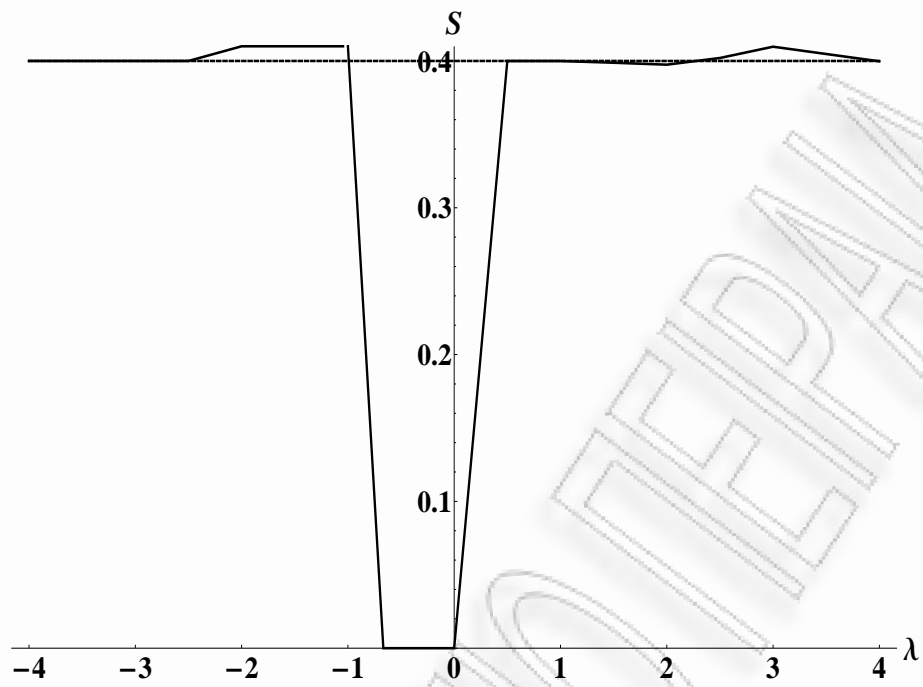
	L85				
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
$S$	$3.15 \times 10^{-13}$	0.000039	0.000039	0.000039	0.000039
$F$	210.084	205.632	205.632	205.646	61.122
Deviance	197.712	194.206	194.206	194.217	61.752
log-likelihood	-17938.14	-17936.38	-17936.38	-17936.39	-17870.16
$\chi^2$	210.11	205.66	205.66	205.67	61.13
	HK01M				
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
$S$	0.000025	0.000025	0.000025	0.000025	0.000025
$F$	41.403	38.690	36.919	34.213	19.964
Deviance	36.37	34.33	32.95	30.81	18.29
log-likelihood	-2090.91	-2089.89	-2089.20	-2088.13	-2081.87
$\chi^2$	41.29	38.54	36.75	34.01	19.73
	HK01F				
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
$S$	$6.55 \times 10^{-6}$	0.00015	0.00015	0.00015	0.00015
$F$	15.636	14.820	14.696	14.265	9.571
Deviance	15.39	14.81	14.72	14.36	9.24
log-likelihood	-1796.46	-1796.17	-1796.12	-1795.95	-1793.38
$\chi^2$	16.32	15.45	15.32	14.85	9.82

μενόμενων θανάτων με την αντίστοιχη  $\chi^2$  στατιστική συνάρτηση,

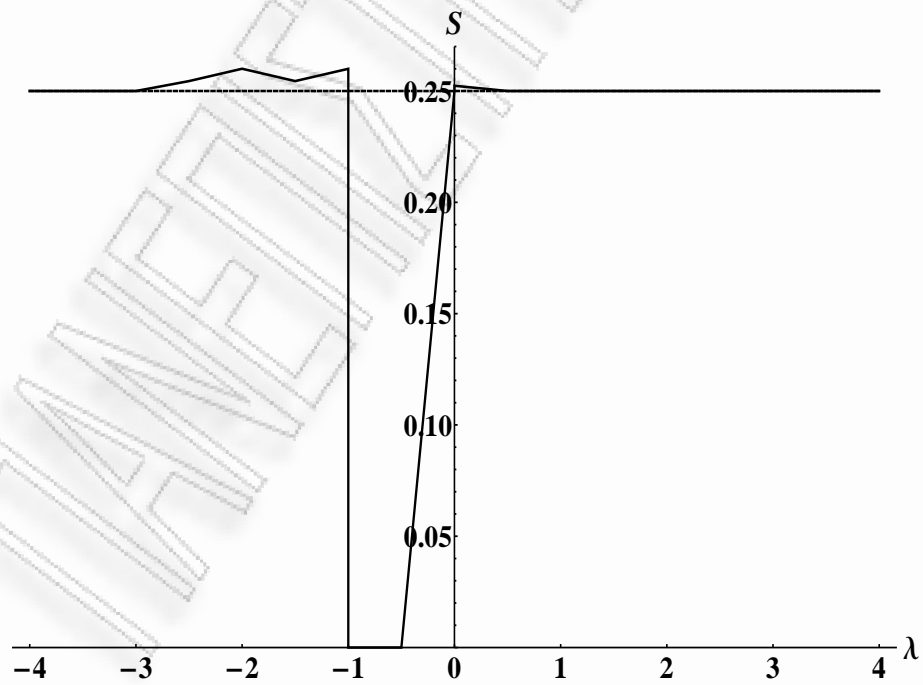
$$\chi^2 = \sum_{i=1}^n \frac{(d_x - l_x v_x)^2}{l_x v_x (1 - v_x)}$$

Από τον Πίνακα 6.4 παρατηρούμε ότι σχεδόν όλες οι εξομαλύνσεις μέσω της απόκλισης δύναμης των Cressie - Read, δίνουν την ίδια τιμή για το μέτρο της ομαλότητας  $S$ , η οποία είναι η τιμή του  $M$  στον περιορισμό της ομαλότητας, ενώ η τιμή του μέτρου καλής προσαρμογής  $F$  μειώνεται καθώς αυξάνεται το  $\lambda$  (για τα σύνολα δεδομένων HK01M και HK01F). Συνοψίζοντας, μπορούμε να πούμε ότι για το σύνολο δεδομένων L85 μπορούμε να χρησιμοποιήσουμε τις τιμές  $\lambda = 2/3$  και  $\lambda = 1$  που δίνουν τα ίδια αποτελέσματα ενώ η τιμή  $\lambda = 2$  είναι η καλύτερη για εξομάλυνση μέσω της απόκλισης δύναμης των Cressie - Read για τα σύνολα HK01M και HK01F.

Περαιτέρω αριθμητική διερεύνηση αποκάλυψε σημαντικά στοιχεία για την «καλύτερη» τιμή της παραμέτρου  $\lambda$ . Στα Σχήματα 6.4 - 6.6, έχουμε απεικονίσει την τιμή του μέτρου ομαλότητας  $S$  έναντι της τιμής της παραμέτρου  $\lambda$ , με το  $S$  απεικονισμένο στον άξονα των  $y$  και το  $\lambda$  στον άξονα των  $x$  για τα σύνολα δεδομένων L85, HK01M και HK01F, αντίστοιχα.

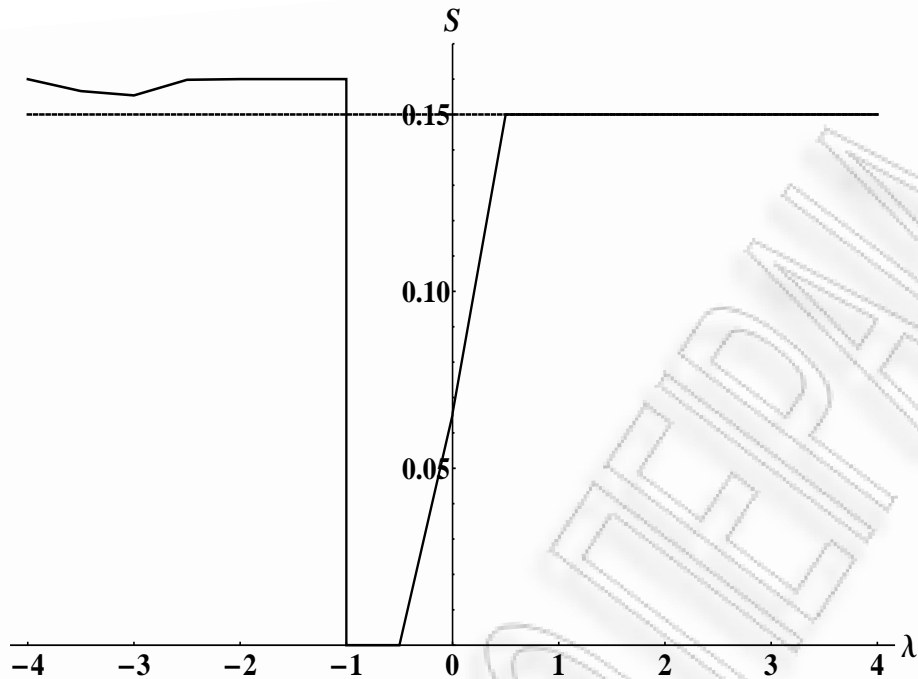


Σχήμα 6.4: Ομαλότητα  $S$  έναντι  $\lambda$  (L85)



Σχήμα 6.5: Ομαλότητα  $S$  έναντι  $\lambda$  (HK01M)



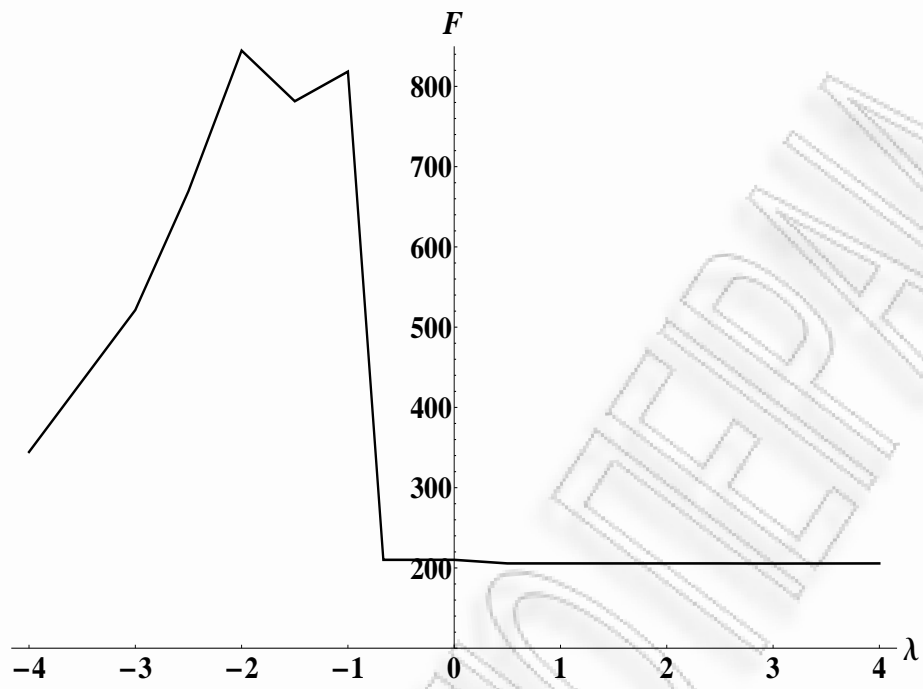


Σχήμα 6.6: Ομαλότητα  $S$  έναντι  $\lambda$  (HK01F)

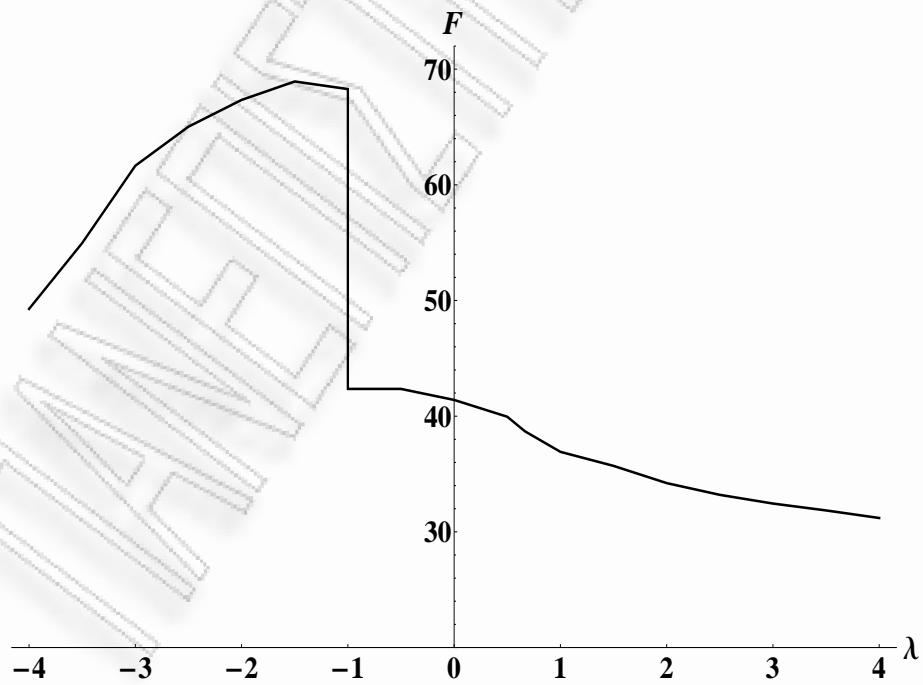
Η διακεκομμένη γραμμή σε κάθε διάγραμμα απεικονίζει την τιμή της παραμέτρου  $M$  στον περιορισμό της ομαλότητας. Παρατηρούμε ότι τα τρία διαγράμματα παρουσιάζουν σχεδόν το ίδιο πρότυπο. Όταν ισχύει  $-\infty < \lambda < -1$ , το μέτρο  $S$  παίρνει μια τιμή κοντά στην τιμή του  $M$ . Στη συνέχεια, όταν  $-1 < \lambda < -0.5$ , το  $S$  παίρνει μια πολύ μικρή τιμή σχεδόν ίση με το μηδέν και ύστερα για τις υπόλοιπες τιμές του  $\lambda$ , ξαναπαίρνει μια τιμή κοντά στην τιμή του  $M$ . Συνεπώς, για τιμές του  $\lambda$  μεταξύ των  $-1$  και  $-0.5$ , η μέθοδος υπερεξομαλύνει τα δεδομένα.

Στα Σχήματα 6.7 - 6.9, παρουσιάζουμε τα αντίστοιχα διαγράμματα που αφορούν το μέτρο προσαρμογής  $F$ , με το  $F$  να απεικονίζεται στον άξονα των  $y$  και το  $\lambda$  στον άξονα των  $x$  για τα σύνολα δεδομένων L85, HK01M και HK01F, αντίστοιχα. Παρατηρούμε και πάλι ότι τα τρία διαγράμματα παρουσιάζουν σχεδόν το ίδιο πρότυπο. Για τιμές του  $\lambda$  μικρότερες του  $-1$ , το μέτρο της προσαρμογής αυξάνεται μέχρι τη μέγιστή του τιμή. Αυτό σημαίνει ότι η εξομάλυνση δεν είναι αποδεκτή καθώς οι εξομαλυμένες τιμές απομακρύνονται αρκετά από τις αδρές παρατηρήσεις. Όταν το  $\lambda$  παίρνει τιμή σχεδόν ίση με  $-1$ , το μέτρο  $F$  μειώνεται και σταθεροποιείται για τις υπόλοιπες τιμές του  $\lambda$ .

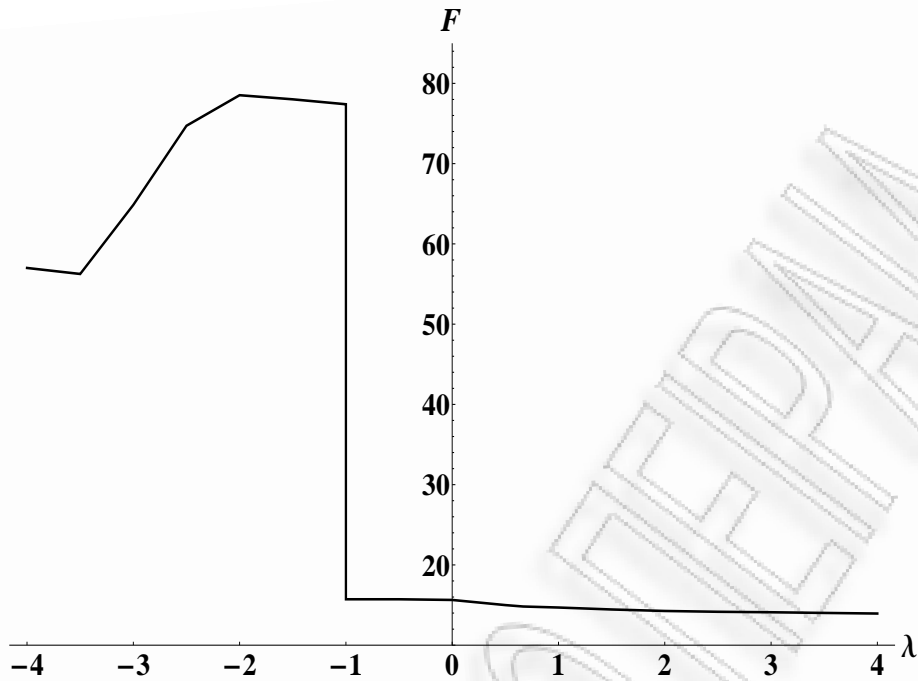
Συγκρίνοντας την εξομάλυνση μέσω της ελαχιστοποίησης της απόκλισης δύναμης των Cressie and Read και της εξομάλυνσης μέσω της ελαχιστοποίησης της απόκλισης των Kullback-Leibler, η οποία είναι απόκλιση δύναμης με  $\lambda \rightarrow 0$ , μπορούμε να πούμε ότι από πλευράς καλής προσαρμογής παίρνουμε σχεδόν τα ίδια αποτελέσματα για τιμές του  $\lambda > -1$ . Όμως όσον αφορά την ομαλότητα, στο σύνολο δεδομένων L85 η εξομάλυνση μέσω της από-



Σχήμα 6.7: Καλή προσαρμογή  $F$  έναντι  $\lambda$  (L85)



Σχήμα 6.8: Καλή προσαρμογή  $F$  έναντι  $\lambda$  (HK01M)



Σχήμα 6.9: Καλή προσαρμογή  $F$  έναντι  $\lambda$  (HK01F)

κλισης των Kullback-Leibler δίνει μια πολύ μικρή τιμή για το μέτρο ομαλότητας  $S$ , το οποίο σημαίνει ότι η μέθοδος υπερεξομαλύνει τα δεδομένα. Το ίδιο αποτέλεσμα προκύπτει επίσης από την ελαχιστοποίηση της απόκλισης δύναμης με  $\lambda \in (-1, 0)$ . Στο σύνολο δεδομένων HK01M η ελαχιστοποίηση της απόκλισης των Kullback-Leibler δίνει το ίδιο αποτέλεσμα με την ελαχιστοποίηση της απόκλισης δύναμης με  $\lambda < -1$  και  $\lambda > 0$ . Τέλος, στο σύνολο HK01F η εξομάλυνση μέσω της απόκλισης των Kullback-Leibler επίσης υπερεξομαλύνει τα δεδομένα, κάτι το οποίο επίσης συμβαίνει χρησιμοποιώντας την απόκλιση δύναμης με  $-1 < \lambda < 0.5$ . Το τελικό μας συμπέρασμα είναι ότι η επιλογή της τιμής  $\lambda = 2/3$  που προτάθηκε από τους Cressie and Read (1984) από πλευράς στατιστικής ισχύος είναι επίσης μια καλή επιλογή για εξομάλυνση.

### 6.3 Ελαχιστοποίηση απόκλισης C - R υπό 6 περιορισμούς

Στη συνέχεια εξομαλύνουμε τα ίδια δεδομένα χρησιμοποιώντας εκτός από τους περιορισμούς (i) - (v) της Ενότητας 4.6 τη σχέση  $\sum_{x=1}^n u_x = \sum_{x=1}^n v_x$ , η οποία όπως είδαμε είναι η ελάχιστη απαίτηση για ένα μέτρο χωρίς μέτρα πιθανότητας να είναι μέτρο απόκλισης. Τα αποτελέσματα για  $\lambda = 0, 2/3, 1, 2$  και τη μέθοδο των Whittaker - Henderson δίνονται στους Πίνακες 6.5, 6.6 και 6.7 για τα σύνολα δεδομένων L85, HK01M και HK01F, αντίστοιχα. Στα Σχήματα 6.10, 6.11 και 6.12 δίνουμε τα αποτελέσματα κάθε εξομάλυνσης σε λογα-

Πίνακας 6.5: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων L85 (6 περιορισμοί)

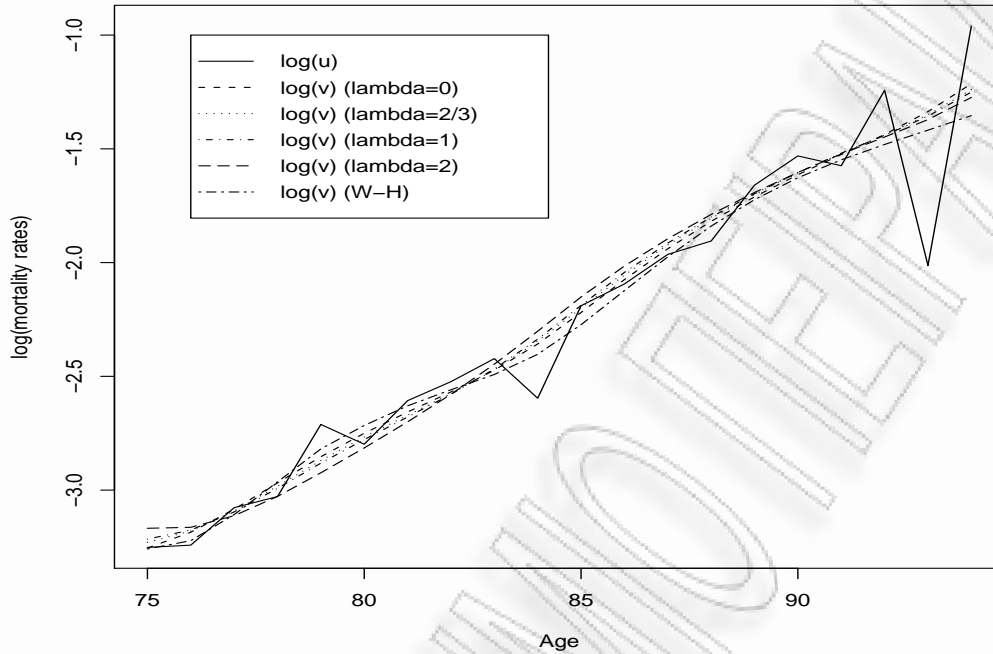
$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
75	0.03870	0.03859	0.03958	0.04014	0.04210	0.03844
76	0.03909	0.04137	0.04159	0.04177	0.04228	0.03987
77	0.04609	0.04577	0.04541	0.04523	0.04449	0.04470
78	0.04839	0.05139	0.05065	0.05018	0.04842	0.05161
79	0.06644	0.05765	0.05671	0.05605	0.05372	0.05967
80	0.06091	0.06391	0.06286	0.06217	0.05993	0.06615
81	0.07376	0.07017	0.06919	0.06864	0.06721	0.07210
82	0.08019	0.07664	0.07607	0.07586	0.07586	0.07739
83	0.08874	0.08432	0.08450	0.08480	0.08658	0.08276
84	0.07455	0.09464	0.09580	0.09667	0.10009	0.09044
85	0.11195	0.10869	0.11077	0.11212	0.11636	0.10302
86	0.12320	0.12557	0.12814	0.12962	0.13332	0.11965
87	0.14031	0.14383	0.14626	0.14747	0.15027	0.13879
88	0.14879	0.16226	0.16439	0.16531	0.16723	0.15886
89	0.19020	0.18070	0.18252	0.18316	0.18418	0.17847
90	0.21628	0.19913	0.20065	0.20101	0.20114	0.19650
91	0.20718	0.21757	0.21877	0.21885	0.21809	0.21261
92	0.28866	0.23764	0.23720	0.23670	0.23505	0.22747
93	0.13362	0.26281	0.25927	0.25765	0.25372	0.24224
94	0.38265	0.29705	0.28938	0.28630	0.27965	0.25817

ριθμική κλίμακα. Οι κώδικες που γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, για την εύρεση των εξομαλυμένων τιμών, παρατίθενται στο Παράρτημα Γ.2.

Από τον Πίνακα 6.8 παρατηρούμε ότι σχεδόν όλες οι εξομαλύνσεις μέσω της απόκλισης δύναμης των Cressie - Read δίνουν την ίδια τιμή για το μέτρο ομαλότητας  $S$ , η οποία είναι πάλι η τιμή της παραμέτρου  $M$  στον περιορισμό της ομαλότητας, ενώ η τιμή του μέτρου καλής προσαρμογής  $F$  αυξάνεται καθώς αυξάνεται το  $\lambda$ . Συνοψίζοντας, μπορούμε να πούμε ότι για το σύνολο δεδομένων L85 μπορούμε να χρησιμοποιήσουμε τις τιμές  $\lambda = 2/3$  και  $\lambda = 1$  που δίνουν τα ίδια αποτελέσματα ενώ η τιμή  $\lambda = 2$  είναι η καλύτερη για εξομάλυνση μέσω της απόκλισης δύναμης για τα σύνολα HK01M και HK01F. Τα αποτελέσματα είναι τώρα σχεδόν ισοδύναμα, από πλευράς ομαλότητας και καλής προσαρμογής, με αυτά που προκύπτουν με τη μέθοδο των Whittaker - Henderson.

Όσον αφορά την ομαλότητα, κάθε επιλογή του  $\lambda$  δίνει σχεδόν την ίδια τιμή για το μέτρο  $S$ . Πιο συγκεκριμένα, έχουμε ότι  $S \approx 4 \times 10^{-5}$  για το σύνολο δεδομένων L85,  $S \approx 3 \times 10^{-5}$  για το σύνολο HK01M και  $S \approx 15 \times 10^{-6}$  για το σύνολο HK01F. Συγκρίνοντας τα αποτελέσματα μεταξύ των εξομαλύνσεων υπό 5 και 6 περιορισμούς, από πλευράς ομαλότητας, παρατηρούμε ότι η χρήση του επιπλέον περιορισμού ( $v_i$ ) βελτιώνει τα αποτελέσματα, καθώς στην περίπτωση αυτή δεν έχουμε το πρόβλημα της υπερεξομάλυνσης.

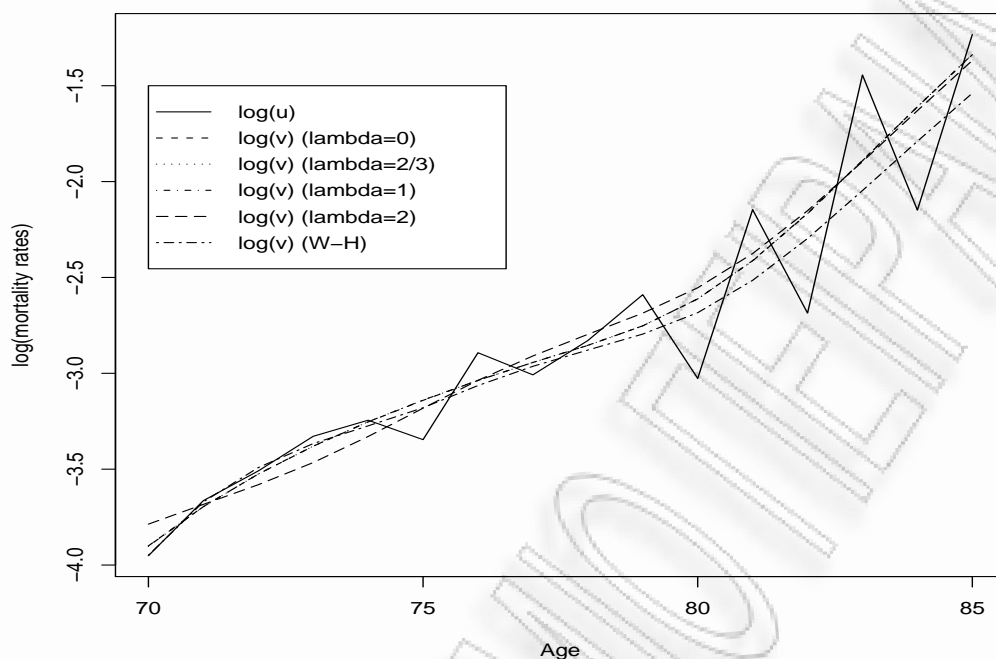
Σχήμα 6.10: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (L85 - 6 περιορισμοί)



Πίνακας 6.6: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων ΗΚ01Μ (6 περιορισμοί)

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
70	0.01923	0.02025	0.02023	0.02022	0.02267	0.01928
71	0.02563	0.02484	0.02483	0.02483	0.02513	0.02546
72	0.02992	0.02944	0.02944	0.02944	0.02785	0.03045
73	0.03585	0.03404	0.03405	0.03406	0.03126	0.03451
74	0.03899	0.03863	0.03866	0.03867	0.03581	0.03785
75	0.03523	0.04324	0.04329	0.04331	0.04150	0.04160
76	0.05543	0.04797	0.04799	0.04800	0.04800	0.04671
77	0.04939	0.05270	0.05269	0.05269	0.05451	0.05164
78	0.05906	0.05761	0.05760	0.05759	0.06101	0.05621
79	0.07503	0.06377	0.06378	0.06378	0.06815	0.06110
80	0.04848	0.07338	0.07341	0.07342	0.07777	0.06830
81	0.11692	0.08960	0.08949	0.08947	0.09297	0.08085
82	0.06816	0.11474	0.11452	0.11447	0.11625	0.10057
83	0.23598	0.15117	0.15093	0.15087	0.15019	0.12914
84	0.11659	0.19989	0.19993	0.19995	0.19591	0.16700
85	0.29152	0.26231	0.26272	0.26281	0.25458	0.21458

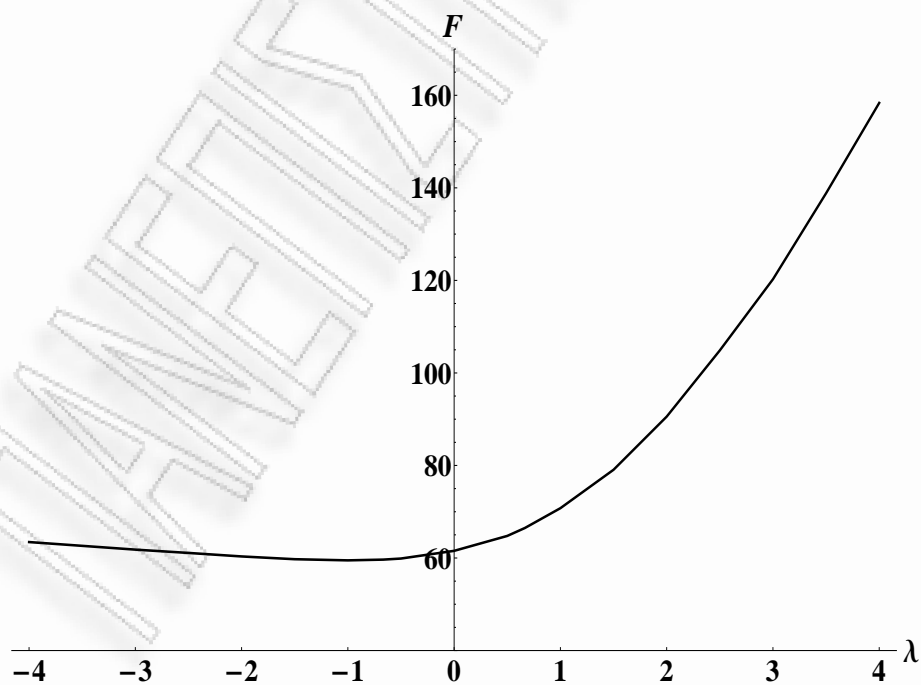
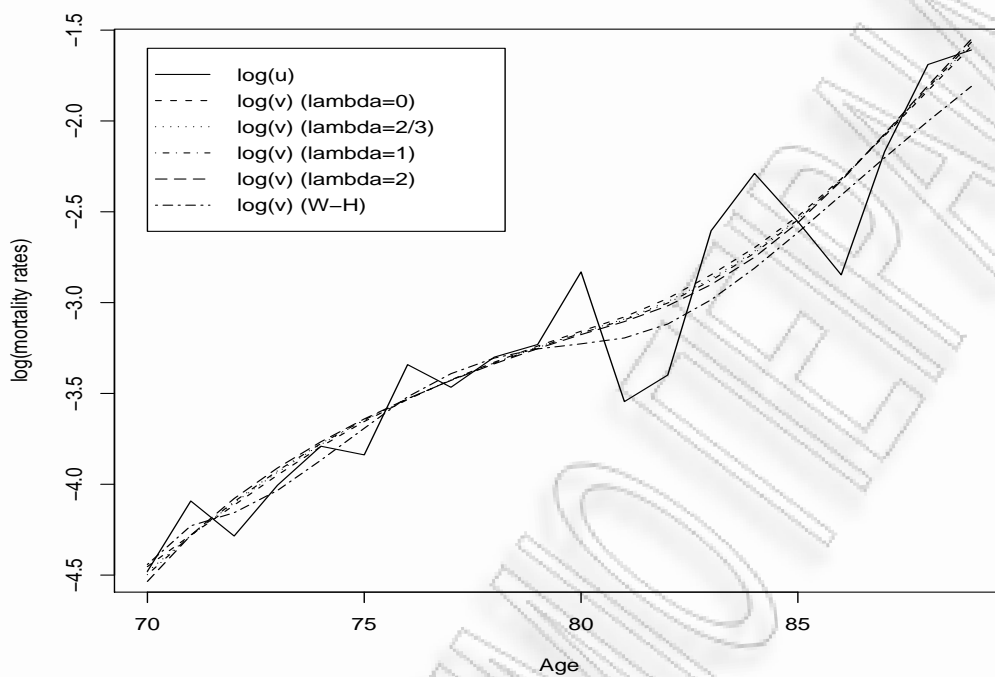
Σχήμα 6.11: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (HK01M - 6 περιορισμοί)



Πίνακας 6.7: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01F (6 περιορισμοί)

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
70	0.01134	0.01164	0.01132	0.01114	0.01073	0.01178
71	0.01670	0.01385	0.01381	0.01381	0.01383	0.01457
72	0.01378	0.01633	0.01652	0.01664	0.01693	0.01566
73	0.01822	0.01926	0.01957	0.01972	0.02003	0.01771
74	0.02258	0.02254	0.02280	0.02291	0.02314	0.02091
75	0.02153	0.02587	0.02604	0.02611	0.02624	0.02492
76	0.03539	0.02920	0.02928	0.02930	0.02934	0.02964
77	0.03124	0.03253	0.03251	0.03249	0.03244	0.03366
78	0.03688	0.03587	0.03575	0.03569	0.03555	0.03669
79	0.03954	0.03920	0.03899	0.03888	0.03865	0.03859
80	0.05896	0.04253	0.04222	0.04207	0.04175	0.03968
81	0.02887	0.04602	0.04550	0.04526	0.04485	0.04098
82	0.03342	0.05094	0.05007	0.04968	0.04892	0.04426
83	0.07395	0.05803	0.05680	0.05622	0.05498	0.05057
84	0.10138	0.06740	0.06605	0.06540	0.06393	0.06018
85	0.07776	0.08010	0.07903	0.07851	0.07730	0.07324
86	0.05800	0.09845	0.09803	0.09783	0.09732	0.08995
87	0.11429	0.12476	0.12529	0.12554	0.12606	0.11056
88	0.18465	0.15994	0.16167	0.16250	0.16431	0.13513
89	0.20000	0.20403	0.20726	0.20878	0.21217	0.16371

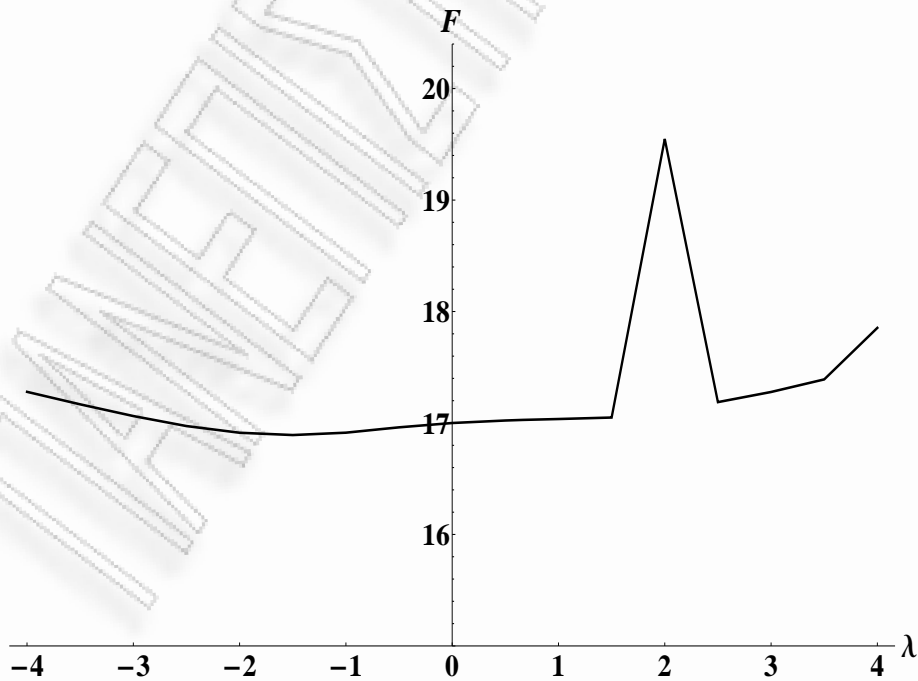
Σχήμα 6.12: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (HK01F - 6 περιορισμοί)



Σχήμα 6.13: Καλή προσαρμογή  $F$  έναντι  $\lambda$  (L85- 6 περιορισμοί)

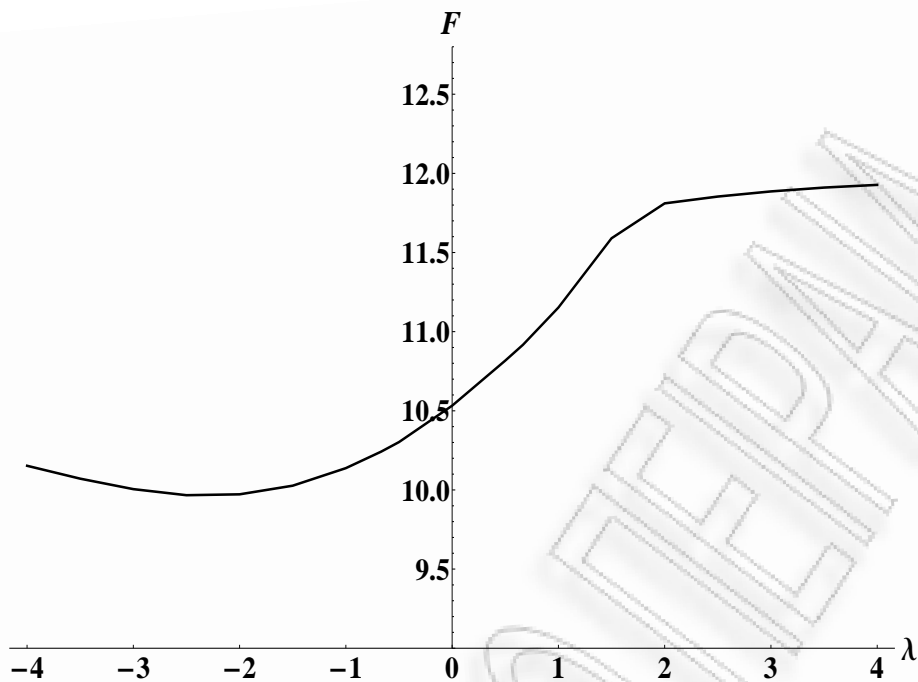
Πίνακας 6.8: Τιμές των μέτρων ομαλότητας και προσαρμογής (6 περιορισμοί)

	<i>L85</i>				
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
<i>S</i>	0.000039	0.00004	0.00004	0.000039	0.000039
<i>F</i>	61.55	66.49	70.77	90.55	61.122
Deviance	64.35	68.91	72.99	92.19	61.75
log-likelihood	-17871.46	-17873.74	-17875.78	-17885.38	-17870.16
$\chi^2$	61.56	66.50	70.78	90.56	61.13
	<i>HK01M</i>				
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
<i>S</i>	0.000029	0.000029	0.00003	0.00003	0.000025
<i>F</i>	17.00	17.03	17.04	19.55	19.964
Deviance	17.11	17.13	17.13	19.83	18.29
log-likelihood	-2081.28	-2081.29	-2081.30	-2082.64	-2081.87
$\chi^2$	16.77	16.80	16.80	19.33	19.73
	<i>HK01F</i>				
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
<i>S</i>	0.000015	0.000015	0.000015	0.000015	0.000015
<i>F</i>	10.53	10.92	11.15	11.81	9.571
Deviance	10.71	11.09	11.33	11.99	9.24
log-likelihood	-1794.12	-1794.31	-1794.43	-1794.76	-1793.38
$\chi^2$	10.81	11.22	11.48	12.17	9.82



Σχήμα 6.14: Καλή προσαρμογή *F* έναντι  $\lambda$  (HK01M- 6 περιορισμοί)



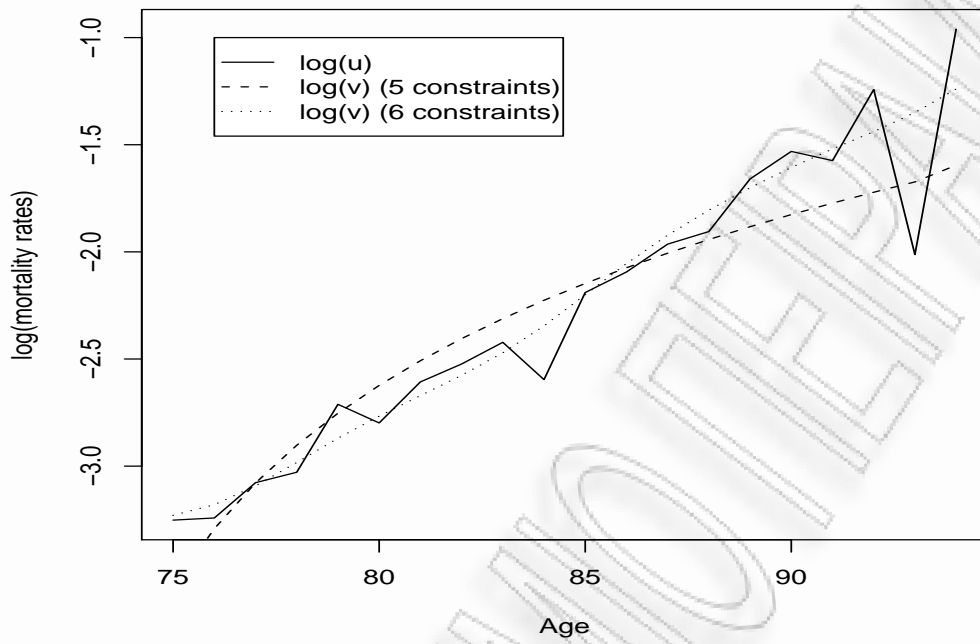


Σχήμα 6.15: Καλή προσαρμογή  $F$  έναντι  $\lambda$  (HK01F- 6 περιορισμοί)

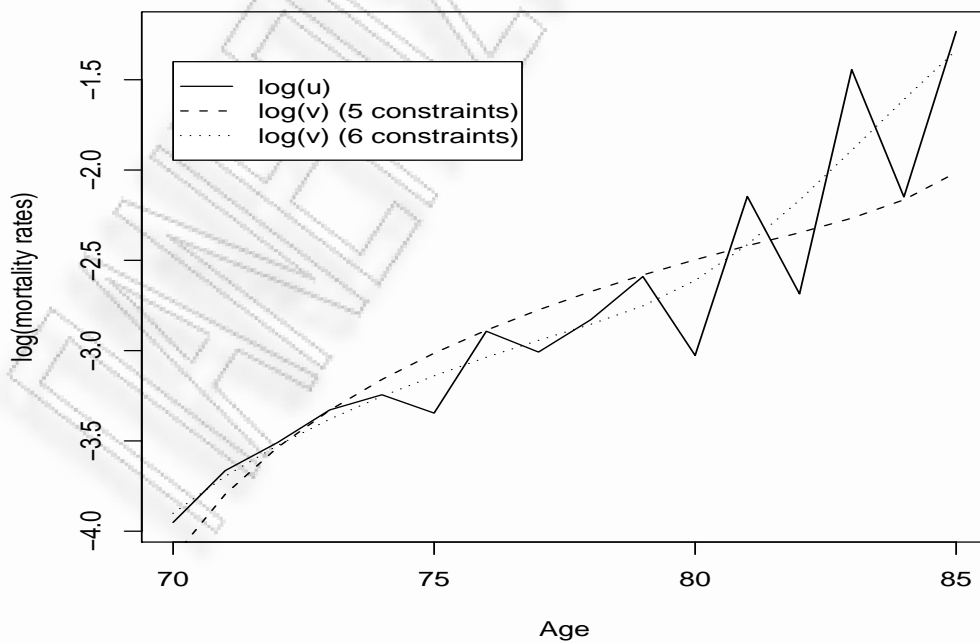
Στο Σχήμα 6.13, παρουσιάζουμε το διάγραμμα που αφορά το μέτρο προσαρμογής  $F$ , με το  $F$  να απεικονίζεται στον άξονα των  $y$  και το  $\lambda$  στον άξονα των  $x$ , για το σύνολο δεδομένων L85. Παρατηρούμε ότι η τιμή του  $F$  ισούται περίπου με 60 για  $\lambda \leq 0$  ενώ για τις θετικές τιμές του  $\lambda$  το μέτρο καλής προσαρμογής αυξάνεται. Για το σύνολο HK01M, παρατηρούμε ότι το μέτρο  $F$  ισούται περίπου με 17 για όλες τις τιμές του  $\lambda \neq 2$  ενώ για  $\lambda = 2$  έχουμε  $F \approx 19.5$  (Σχήμα 6.14). Τέλος, από το Σχήμα 6.15 που αναφέρεται στο σύνολο HK01F, παρατηρούμε ότι δεν υπάρχουν μεγάλες διαφορές καθώς το  $F \in (9.9, 12)$ . Συγκρίνοντας πάλι τα αποτελέσματα μεταξύ των εξομαλύνσεων υπό 5 και 6 περιορισμούς, από πλευράς καλής προσαρμογής, ο επιπλέον περιορισμός βελτιώνει τα αποτελέσματα καθώς οι εξομαλυμένες τιμές παρουσιάζουν μια σαφώς καλύτερη προσαρμογή.

Στη συνέχεια στα γραφήματα 6.16, 6.17 και 6.18 απεικονίζονται οι αδρές  $u_x$  και οι εξομαλυμένες  $v_x$  τιμές, σε λογαριθμική κλίμακα, που επιτυγχάνονται από την εξομάλυνση μέσω του μέτρου των Cressie and Read με τιμή της παραμέτρου  $\lambda = 2/3$  και τη χρήση των 5 και 6 περιορισμών για τα σύνολα δεδομένων L85, HK01M και HK01F, αντίστοιχα. Παρατηρούμε ότι και στις τρεις περιπτώσεις η εξομάλυνση μέσω των 6 περιορισμών είναι καλύτερη όσον αφορά την προσαρμογή των εξομαλυμένων τιμών. Σημειώνουμε ότι το παραπάνω συμπέρασμα ισχύει για όλες τις τιμές του  $\lambda$  και όχι μόνο για την τιμή  $\lambda = 2/3$  που για λόγους χώρου παρουσιάζουμε.

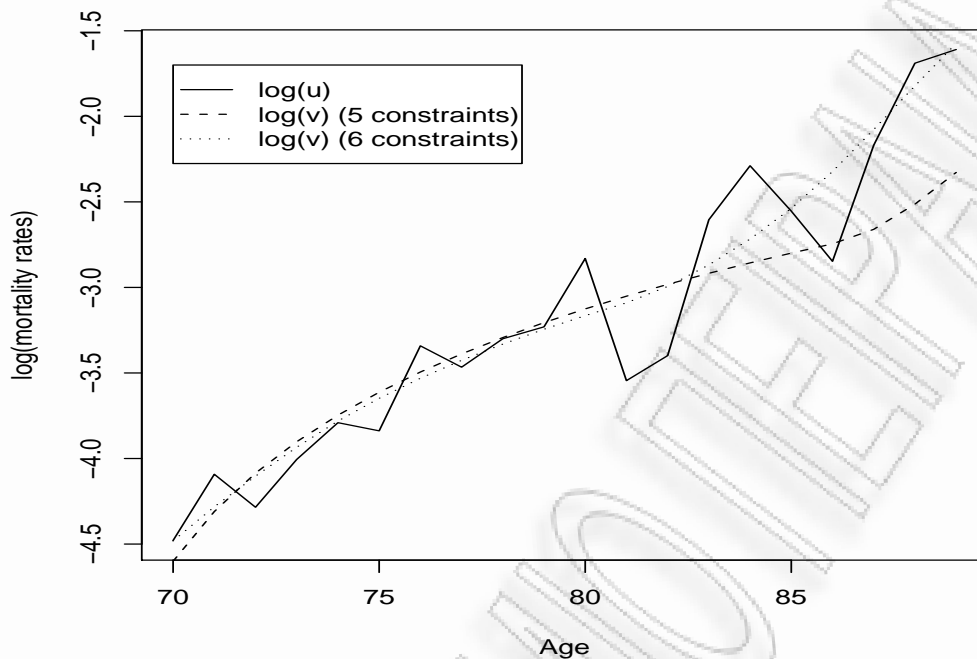
Σχήμα 6.16: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (L85 - 5 και 6 περιορισμοί και  $\lambda = 2/3$ )



Σχήμα 6.17: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (HK01M - 5 και 6 περιορισμοί και  $\lambda = 2/3$ )



Σχήμα 6.18: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (HK01F - 5 και 6 περιορισμοί και  $\lambda = 2/3$ )



#### 6.4 Ελαχιστοποίηση απόκλισης C - R υπό 4 περιορισμούς

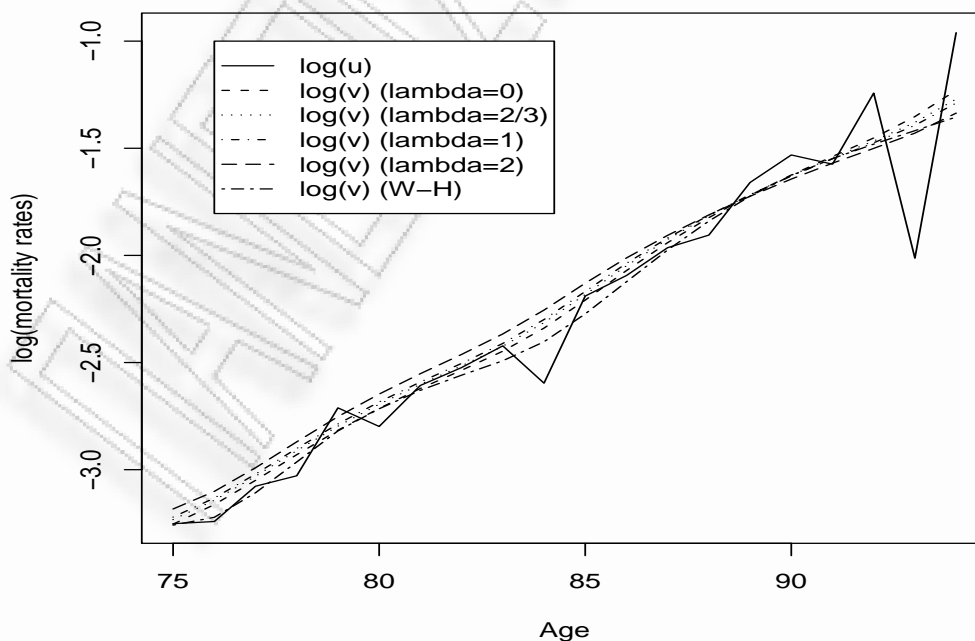
Στην ενότητα αυτή παρουσιάζουμε την εξομάλυνση των ίδιων δεδομένων χρησιμοποιώντας τους περιορισμούς (i), (ii), (iii) και (vi). Ο λόγος για τον οποίο διεξάγουμε τις εξομαλύνσεις με αυτούς μόνο τους περιορισμούς είναι για να δούμε εάν οι αναλογιστικοί περιορισμοί (iv) και (v) που προτείνει ο Brockett (1991) είναι απαραίτητοι. Τα αποτελέσματα για  $\lambda = 0, 2/3, 1, 2$  και τη μέθοδο των Whittaker - Henderson δίνονται στους Πίνακες 6.9, 6.10 και 6.11 για τα σύνολα L85, HK01M και HK01F, αντίστοιχα. Στα Σχήματα 6.19, 6.20 και 6.21 δίνουμε τα αποτελέσματα κάθε εξομάλυνσης σε λογαριθμική κλίμακα. Οι κώδικες που γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, για την εύρεση των εξομαλυμένων τιμών, παρατίθενται στο Παράρτημα Γ.3.

Παρατηρώντας, τόσο τους τρεις αυτούς πίνακες όσο και τον Πίνακα 6.12, αντιλαμβανόμαστε ότι τα αποτελέσματα είναι παρόμοια με αυτά του Πίνακα 6.8. Η μόνη διαφορά που παρατηρούμε είναι μια μικρή αύξηση στην τιμή των μέτρων καλής προσαρμογής. Από την άποψη των διαφορών στην προσαρμογή, θεωρούμε ότι οι δυο αναλογιστικοί περιορισμοί δεν χρειάζονται κατά την ελαχιστοποίηση των αποκλίσεων δύναμης. Για να είμαστε όμως περισσότερο ανταγωνιστικοί με τη μέθοδο των Whittaker - Henderson, οι περιορισμοί (iv) και (v) καλό είναι να χρησιμοποιούνται. Άλλωστε η επιβάρυνσή τους στον υπολογιστικό χρόνο της ελαχιστοποίησης είναι μηδαμινή.

Πίνακας 6.9: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων L85 (4 περιορισμοί)

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
75	0.03870	0.03861	0.03944	0.03990	0.04144	0.03844
76	0.03909	0.04230	0.04316	0.04361	0.04499	0.03987
77	0.04609	0.04737	0.04829	0.04877	0.05023	0.04470
78	0.04839	0.05340	0.05442	0.05497	0.05670	0.05161
79	0.06644	0.05984	0.06099	0.06163	0.06378	0.05967
80	0.06091	0.06628	0.06757	0.06829	0.07085	0.06615
81	0.07376	0.07273	0.07415	0.07496	0.07793	0.07210
82	0.08019	0.07917	0.08082	0.08177	0.08525	0.07739
83	0.08874	0.08656	0.08858	0.08974	0.09381	0.08276
84	0.07455	0.09640	0.09889	0.10033	0.10486	0.09044
85	0.11195	0.10985	0.11274	0.11436	0.11886	0.10302
86	0.12320	0.12607	0.12892	0.13042	0.13374	0.11965
87	0.14031	0.14364	0.14584	0.14682	0.14862	0.13879
88	0.14879	0.16141	0.16276	0.16323	0.16350	0.15886
89	0.19020	0.17918	0.17969	0.17964	0.17837	0.17847
90	0.21628	0.19696	0.19661	0.19605	0.19325	0.19650
91	0.20718	0.21473	0.21354	0.21245	0.20813	0.21261
92	0.28866	0.23415	0.23079	0.22886	0.22301	0.22747
93	0.13362	0.25870	0.25174	0.24836	0.23949	0.24224
94	0.38265	0.29237	0.28076	0.27553	0.26291	0.25817

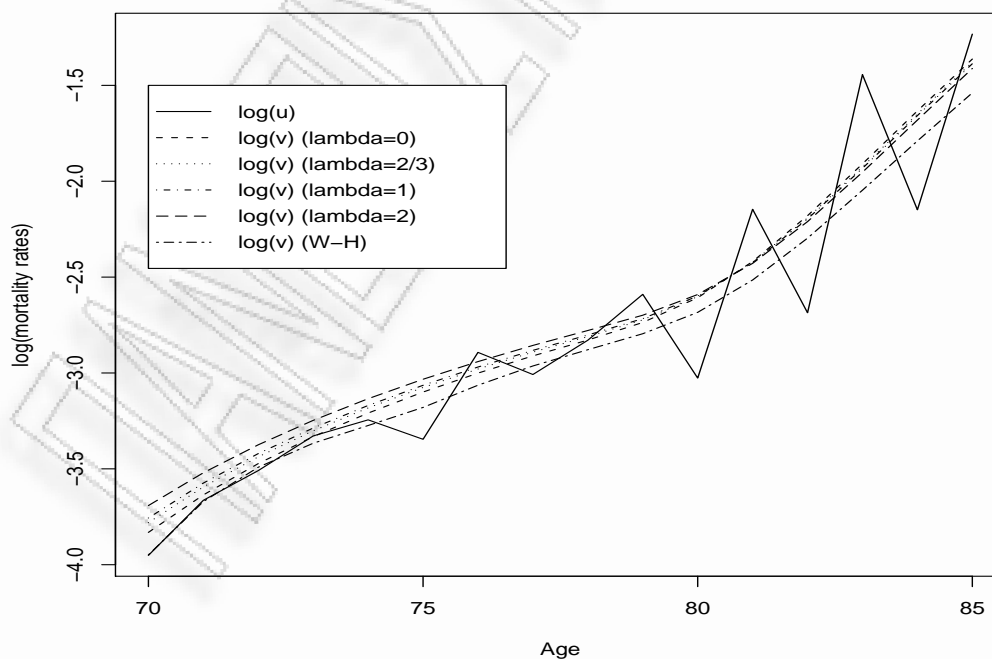
Σχήμα 6.19: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (L85 - 4 περιορισμοί)



Πίνακας 6.10: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων ΗΚ01Μ (4 περιορισμοί)

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
70	0.01923	0.02168	0.02281	0.02337	0.02493	0.01928
71	0.02563	0.02636	0.02748	0.02803	0.02958	0.02546
72	0.02992	0.03104	0.03215	0.03270	0.03423	0.03045
73	0.03585	0.03571	0.03682	0.03736	0.03888	0.03451
74	0.03899	0.04039	0.04149	0.04203	0.04353	0.03785
75	0.03523	0.04507	0.04616	0.04669	0.04818	0.04160
76	0.05543	0.04975	0.05083	0.05136	0.05283	0.04671
77	0.04939	0.05443	0.05550	0.05602	0.05748	0.05164
78	0.05906	0.05910	0.06017	0.06069	0.06213	0.05621
79	0.07503	0.06479	0.06567	0.06610	0.06729	0.06110
80	0.04848	0.07371	0.07415	0.07437	0.07496	0.06830
81	0.11692	0.08904	0.08868	0.08854	0.08818	0.08085
82	0.06816	0.11307	0.11176	0.11118	0.10963	0.10057
83	0.23598	0.14817	0.14585	0.14477	0.14185	0.12914
84	0.11659	0.19531	0.19215	0.19059	0.18619	0.16700
85	0.29152	0.25595	0.25190	0.24978	0.24368	0.21458

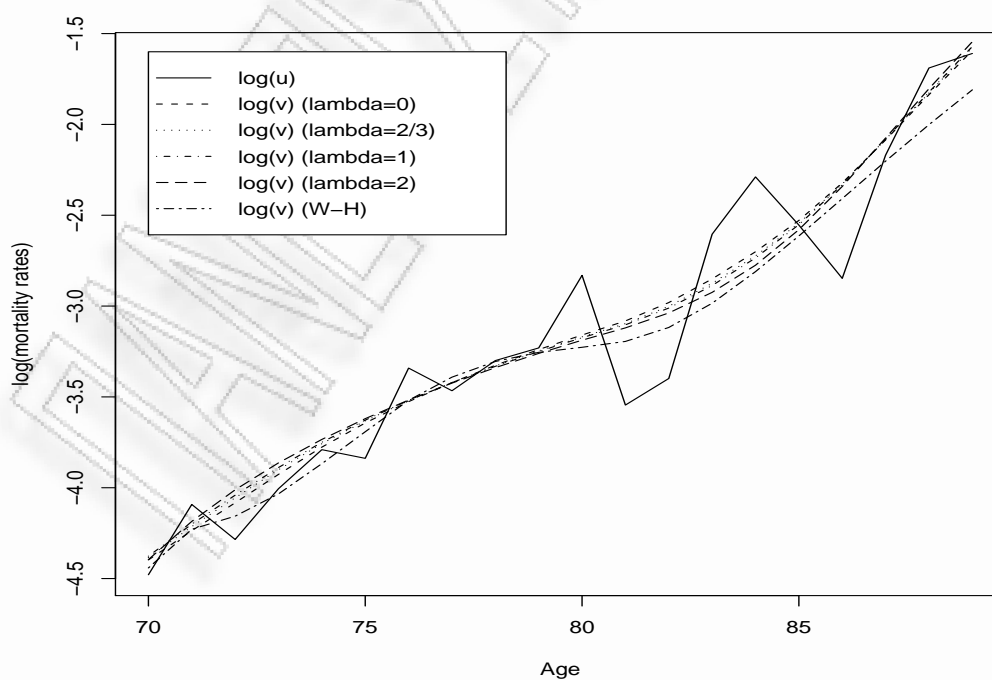
Σχήμα 6.20: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (ΗΚ01Μ - 4 περιορισμοί)



Πίνακας 6.11: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων ΗΚ01F (4 περιορισμοί)

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
70	0.01134	0.01231	0.01247	0.01257	0.01233	0.01178
71	0.01670	0.01445	0.01479	0.01499	0.01522	0.01457
72	0.01378	0.01686	0.01733	0.01758	0.01811	0.01566
73	0.01822	0.01971	0.02023	0.02047	0.02100	0.01771
74	0.02258	0.02290	0.02334	0.02352	0.02390	0.02091
75	0.02153	0.02616	0.02645	0.02657	0.02679	0.02492
76	0.03539	0.02942	0.02957	0.02962	0.02968	0.02964
77	0.03124	0.03268	0.03268	0.03268	0.03258	0.03366
78	0.03688	0.03593	0.03580	0.03573	0.03547	0.03669
79	0.03954	0.03919	0.03891	0.03878	0.03836	0.03859
80	0.05896	0.04245	0.04202	0.04183	0.04126	0.03968
81	0.02887	0.04587	0.04521	0.04490	0.04415	0.04098
82	0.03342	0.05076	0.04973	0.04925	0.04794	0.04426
83	0.07395	0.05781	0.05643	0.05575	0.05369	0.05057
84	0.10138	0.06714	0.06564	0.06490	0.06234	0.06018
85	0.07776	0.07978	0.07855	0.07794	0.07581	0.07324
86	0.05800	0.09806	0.09746	0.09718	0.09617	0.08995
87	0.11429	0.12428	0.12462	0.12479	0.12556	0.11056
88	0.18465	0.15936	0.16089	0.16164	0.16472	0.13513
89	0.20000	0.20335	0.20635	0.20780	0.21339	0.16371

Σχήμα 6.21: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές (ΗΚ01F - 4 περιορισμοί)



Πίνακας 6.12: Τιμές των μέτρων ομαλότητας και προσαρμογής (4 περιορισμοί)

	<i>L85</i>				
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
<i>S</i>	0.00004	0.000039	0.000039	0.000039	0.000039
<i>F</i>	63.24	71.34	77.52	104.38	61.122
Deviance	66.44	74.61	80.91	108.64	61.75
log-likelihood	-17872.50	-17876.59	-17879.73	-17893.60	-17870.16
$\chi^2$	63.25	71.35	77.53	104.39	61.13
<i>HK01M</i>					
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
	<i>S</i>	0.00003	0.00003	0.00003	0.00003
<i>F</i>	18.15	20.19	21.53	26.30	19.964
Deviance	18.24	20.31	21.70	26.72	18.29
log-likelihood	-2081.85	-2082.89	-2083.58	-2086.09	-2081.87
$\chi^2$	17.91	19.94	21.28	26.04	19.73
<i>HK01F</i>					
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
	<i>S</i>	0.000015	0.000015	0.000015	0.000015
<i>F</i>	10.63	11.23	11.62	12.55	9.571
Deviance	10.83	11.44	11.82	12.75	9.24
log-likelihood	-1794.18	-1794.48	-1794.67	-1795.14	-1793.38
$\chi^2$	10.80	11.37	11.73	12.66	9.82

## 6.5 Ελαχιστοποίηση του εκτεταμένου μέτρου απόκλισης των Cressie and Read

Στην παράγραφο αυτή, θα παρουσιάσουμε τα αποτελέσματα της ελαχιστοποίησης του εκτεταμένου μέτρου απόκλισης δύναμης των Cressie and Read

$$I_{ext}^{CR}(\mathbf{v}, \mathbf{u}) = \frac{1}{\lambda(\lambda + 1)} \sum_x \left\{ v_x \left[ \left( \frac{v_x}{u_x} \right)^\lambda - 1 \right] - v_x + u_x \right\},$$

για τα τρία σύνολα δεδομένων. Η ελαχιστοποίηση του μέτρου έγινε υπό τους περιορισμούς (i) - (v). Τα αποτελέσματα για  $\lambda = 0, 2/3, 1, 2$  και τη μέθοδο των Whittaker - Henderson δίνονται στους Πίνακες 6.13, 6.14 και 6.15 για τα σύνολα L85, HK01M και HK01F, αντίστοιχα. Στα Σχήματα 6.22, 6.23 και 6.24 δίνουμε τα αποτελέσματα κάθε εξομάλυνσης σε λογαριθμική κλίμακα. Οι κώδικες που γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, για την εύρεση των εξομαλυμένων τιμών, παρατίθενται στο Παράρτημα Γ.4.

Τα αποτελέσματα που αφορούν την ομαλότητα και την προσαρμογή των δεδομένων παρουσιάζονται στον Πίνακα 6.16. Οι τιμές του μέτρου ομαλότητας  $S$  είναι ίσες ή σχεδόν ίσες για όλες τις επιλεγμένες τιμές του  $\lambda$  στους Πίνακες 6.4 και 6.8. Εάν συγκρίνουμε την προσαρμογή με την αντίστοιχη της ελαχιστοποίησης του μέτρου των Cressie and Read υπό τους πέντε περιορισμούς (Πίνακας 6.4), παρατηρούμε μια σημαντικά καλύτερη προσαρμογή. Παρόλα αυτά, η προσαρμογή που προέκυψε από την ελαχιστοποίηση του μέτρου των Cressie and Read υπό τους έξι περιορισμούς που προτείνουμε (Πίνακας 6.8) είναι καλύτερη για κάθε επιλεγμένη τιμή της παραμέτρου  $\lambda$ . Ως συμπέρασμα μπορούμε να πούμε ότι το εκτεταμένο μέτρο απόκλισης δύναμης των Cressie and Read συμπεριφέρεται συγκριτικά με παρόμοιο τρόπο, αλλά όχι καλύτερα, με το μέτρο των Cressie and Read με 6 περιορισμούς.

## 6.6 Ελαχιστοποίηση της διαφοράς του Jensen

Στην παράγραφο αυτή, θα παρουσιάσουμε τα αποτελέσματα της ελαχιστοποίησης της διαφοράς του Jensen. Τα δεδομένα που θα χρησιμοποιήσουμε είναι, για λόγους σύγκρισης, τα τρία σύνολα δεδομένων που χρησιμοποιήσαμε προηγουμένως ενώ η ελαχιστοποίηση γίνεται πάλι υπό τους περιορισμούς (i) - (v) και/ή (vi). Οι κώδικες που γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, για την εύρεση των εξομαλυμένων τιμών, παρατίθενται στο Παράρτημα Γ.5.



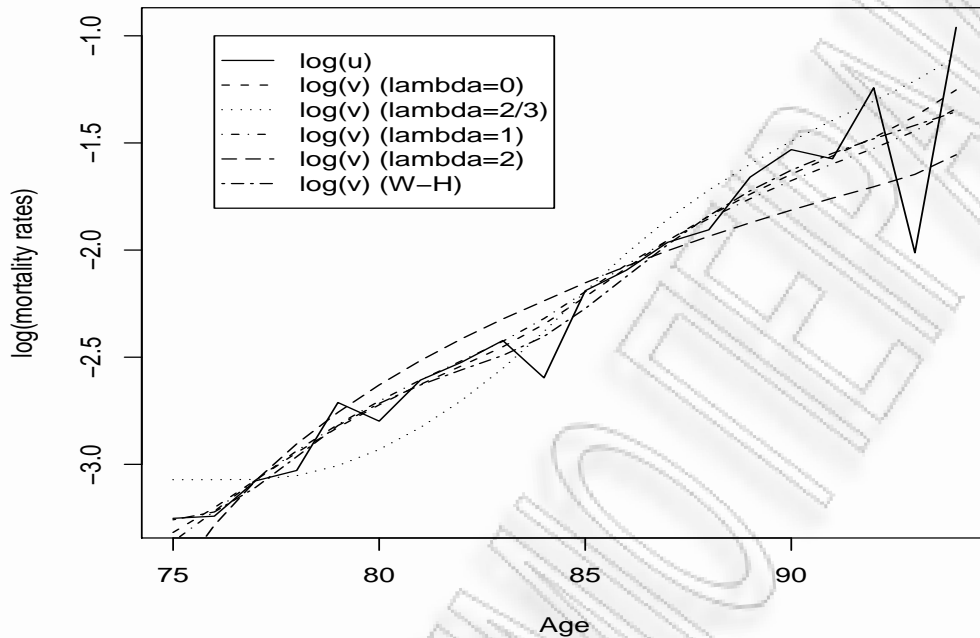
Πίνακας 6.13: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων L85 με το εκτεταμένο μέτρο CR και 5 περιορισμούς

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
75	0.03870	0.03623	0.04634	0.03472	0.02880	0.03844
76	0.03909	0.04067	0.04634	0.03999	0.03746	0.03987
77	0.04609	0.04626	0.04641	0.04623	0.04613	0.04470
78	0.04839	0.05259	0.04723	0.05304	0.05479	0.05161
79	0.06644	0.05915	0.04949	0.05996	0.06345	0.05967
80	0.06091	0.06572	0.05335	0.06688	0.07212	0.06615
81	0.07376	0.07229	0.05913	0.07381	0.08078	0.07210
82	0.08019	0.07885	0.06709	0.08073	0.08944	0.07739
83	0.08874	0.08618	0.07791	0.08841	0.09811	0.08276
84	0.07455	0.09575	0.09253	0.09832	0.10677	0.09044
85	0.11195	0.10878	0.11147	0.11143	0.11616	0.10302
86	0.12320	0.12445	0.13330	0.12637	0.12554	0.11965
87	0.14031	0.14142	0.15622	0.14156	0.13493	0.13879
88	0.14879	0.15855	0.17914	0.15675	0.14432	0.15886
89	0.19020	0.17568	0.20205	0.17194	0.15370	0.17847
90	0.21628	0.19282	0.22497	0.18714	0.16309	0.19650
91	0.20718	0.20995	0.24788	0.20233	0.17247	0.21261
92	0.28866	0.22871	0.27116	0.21752	0.18186	0.22747
93	0.13362	0.25263	0.29818	0.23583	0.19271	0.24224
94	0.38265	0.28575	0.33332	0.26193	0.21110	0.25817

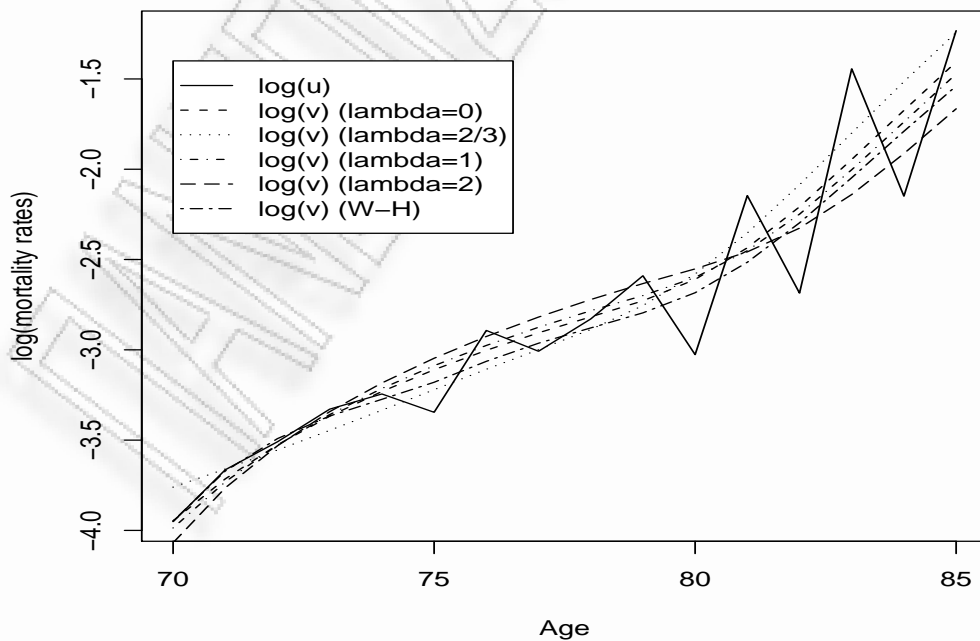
Πίνακας 6.14: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων HK01M με το εκτεταμένο μέτρο CR και 5 περιορισμούς

$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
70	0.01923	0.01931	0.02328	0.01857	0.01710	0.01928
71	0.02563	0.02437	0.02574	0.02398	0.02319	0.02546
72	0.02992	0.02942	0.02859	0.02938	0.02927	0.03045
73	0.03585	0.03448	0.03187	0.03479	0.03535	0.03451
74	0.03899	0.03953	0.03564	0.04019	0.04143	0.03785
75	0.03523	0.04459	0.03992	0.04560	0.04752	0.04160
76	0.05543	0.04964	0.04470	0.05101	0.05360	0.04671
77	0.04939	0.05470	0.04982	0.05641	0.05968	0.05164
78	0.05906	0.05975	0.05571	0.06182	0.06577	0.05621
79	0.07503	0.06531	0.06362	0.06722	0.07185	0.06110
80	0.04848	0.07351	0.07575	0.07431	0.07793	0.06830
81	0.11692	0.08760	0.09510	0.08619	0.08565	0.08085
82	0.06816	0.10991	0.12418	0.10553	0.09766	0.10057
83	0.23598	0.14287	0.16537	0.13497	0.11766	0.12914
84	0.11659	0.18746	0.21983	0.17585	0.14758	0.16700
85	0.29152	0.24520	0.28861	0.22943	0.18912	0.21458

Σχήμα 6.22: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές με το εκτεταμένο μέτρο CR (L85 - 5 περιορισμοί)



Σχήμα 6.23: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές με το εκτεταμένο μέτρο CR (HK01M - 5 περιορισμοί)



Πίνακας 6.15: Διάφορες εξομαλύνσεις για το σύνολο δεδομένων ΗΚ01F με το εκτεταμένο μέτρο CR και 5 περιορισμούς

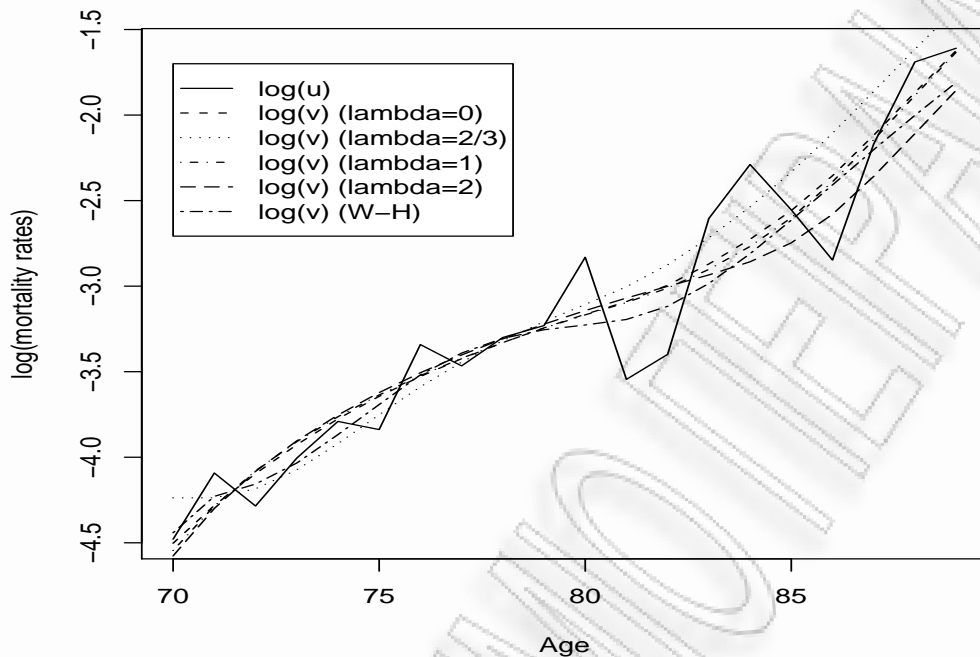
$x$	$u_x$	$v_x (\lambda = 0)$	$v_x (\lambda = 2/3)$	$v_x (\lambda = 1)$	$v_x (\lambda = 2)$	$v_x (W - H)$
70	0.01134	0.01108	0.01445	0.01062	0.01029	0.01178
71	0.01670	0.01382	0.01445	0.01377	0.01358	0.01457
72	0.01378	0.01663	0.01523	0.01691	0.01686	0.01566
73	0.01822	0.01970	0.01702	0.02006	0.02015	0.01771
74	0.02258	0.02292	0.01983	0.02321	0.02344	0.02091
75	0.02153	0.02614	0.02348	0.02636	0.02672	0.02492
76	0.03539	0.02936	0.02765	0.02951	0.03001	0.02964
77	0.03124	0.03258	0.03192	0.03265	0.03330	0.03366
78	0.03688	0.03580	0.03619	0.03580	0.03659	0.03669
79	0.03954	0.03902	0.04047	0.03895	0.03987	0.03859
80	0.05896	0.04224	0.04474	0.04210	0.04316	0.03968
81	0.02887	0.04549	0.04962	0.04524	0.04645	0.04098
82	0.03342	0.05006	0.05642	0.04923	0.04973	0.04426
83	0.07395	0.05668	0.06605	0.05491	0.05316	0.05057
84	0.10138	0.06546	0.07913	0.06280	0.05741	0.06018
85	0.07776	0.07743	0.09694	0.07424	0.06413	0.07324
86	0.05800	0.09494	0.12168	0.09159	0.07577	0.08995
87	0.11429	0.12029	0.15540	0.11713	0.09457	0.11056
88	0.18465	0.15439	0.19891	0.15171	0.12141	0.13513
89	0.20000	0.19729	0.25233	0.19543	0.15641	0.16371

Πίνακας 6.16: Τιμές των μέτρων ομαλότητας και προσαρμογής για την εξομάλυνση με (i) το εκτεταμένο CR μέτρο απόκλισης και 5 περιορισμούς και (ii) τη μέθοδο W-H (χωρίς περιορισμούς) για τα σύνολα δεδομένων L85, HK01M, HK01F

	L85				
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
$S$	0.00004	0.00004	0.00004	0.00004	0.000039
$F$	63.61	173.37	75.73	187.56	61.12
Deviance	65.95	174.38	76.66	178.29	61.75
log-likelihood	-17872.26	-17926.47	-17877.61	-17928.43	-17870.16
$\chi^2$	63.63	173.38	75.75	187.58	61.13
$I_{ext}^{CR}(\mathbf{v}, \mathbf{u})$	0.068	0.058	0.075	0.042	0.070*
HK01M					
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
$S$	0.00003	0.00003	0.00003	0.00003	0.000025
$F$	17.50	19.80	19.12	25.94	19.96
Deviance	17.46	20.85	18.68	24.27	18.29
log-likelihood	-2081.46	-2083.16	-2082.07	-2084.86	-2081.87
$\chi^2$	17.50	19.80	19.12	25.94	19.73
$I_{ext}^{CR}(\mathbf{v}, \mathbf{u})$	0.065	0.066	0.067	0.047	0.069*
HK01F					
	$\lambda = 0$	$\lambda = 2/3$	$\lambda = 1$	$\lambda = 2$	$W - H$
$S$	0.000015	0.000015	0.000015	0.000015	0.000015
$F$	11.08	12.66	11.94	13.08	9.57
Deviance	11.29	13.04	12.15	13.26	9.24
log-likelihood	-1794.41	-1795.28	-1794.84	-1795.39	-1793.38
$\chi^2$	11.42	12.66	12.33	13.57	9.82
$I_{ext}^{CR}(\mathbf{v}, \mathbf{u})$	0.034	0.014	0.035	0.017	0.041*

\* για  $\lambda = 0$

Σχήμα 6.24: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές με το εκτεταμένο μέτρο CR (HK01F - 5 περιορισμοί)



Τα αποτελέσματα της ελαχιστοποίησης της διαφοράς του Jensen

$$J(\mathbf{u}, \mathbf{v}) = - \sum_{i=1}^n \frac{1}{2} (u_i + v_i) \ln \left( \frac{1}{2} (u_i + v_i) \right) + \frac{1}{2} \left[ \sum_{i=1}^n v_i \ln v_i + \sum_{i=1}^n u_i \ln u_i \right]$$

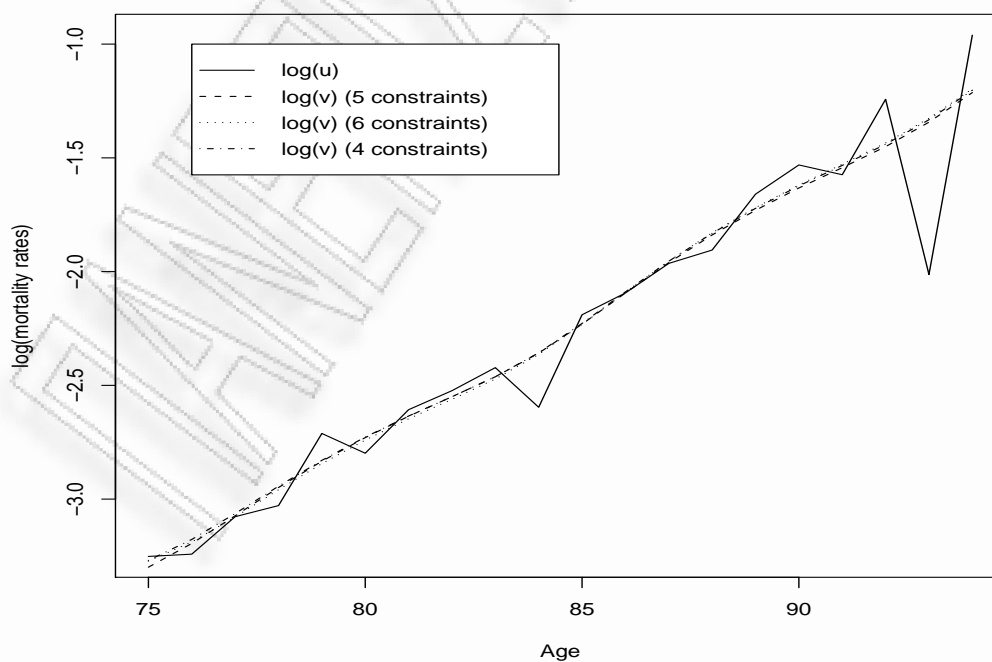
μεταξύ των αδρών και των εξομαλυμένων ποσοστών θνησιμότητας  $\mathbf{u}$  και  $\mathbf{v}$  εμφανίζονται στους Πίνακες 6.17, 6.18 και 6.19. Συγκεκριμένα, στον Πίνακα 6.17 παρουσιάζονται τα αποτελέσματα που αφορούν το σύνολο δεδομένων L85, στον Πίνακα 6.18 τα αποτελέσματα που αφορούν το σύνολο δεδομένων HK01M ενώ στον Πίνακα 6.19 τα αποτελέσματα που αφορούν το σύνολο δεδομένων HK01F. Στα Σχήματα 6.25, 6.26 και 6.27 δίνουμε τα αποτελέσματα κάθε εξομάλυνσης σε λογαριθμική κλίμακα.

Παρατηρούμε ότι στο σύνολο δεδομένων HK01M η εξομάλυνση με 4 περιορισμούς δεν έδωσε ικανοποιητικά αποτελέσματα. Παρατηρώντας τους πίνακες αυτούς καθώς επίσης και τον Πίνακα 6.20, ο οποίος παρουσιάζει τις τιμές των μέτρων ομαλότητας και προσαρμογής, βλέπουμε ότι τα αποτελέσματα είναι αντίστοιχα με αυτά που παρουσιάζονται στον Πίνακα 6.8. Η χρήση του επιπλέον περιορισμού ( $v_i$ ) βελτίωσε λίγο την προσαρμογή των δεδομένων.

Πίνακας 6.17: Εξομάλυνση μέσω Jensen difference για το σύνολο δεδομένων L85

$x$	$u_x$	$v_x$ (5 περιορισμοί)	$v_x$ (6 περιορισμοί)	$v_x$ (4 περιορισμοί)
75	0.03870	0.03691	0.03792	0.03803
76	0.03909	0.04097	0.04124	0.04170
77	0.04609	0.04626	0.04603	0.04673
78	0.04839	0.05239	0.05187	0.05269
79	0.06644	0.05881	0.05820	0.05904
80	0.06091	0.06524	0.06453	0.06539
81	0.07376	0.07166	0.07086	0.07174
82	0.08019	0.07809	0.07719	0.07809
83	0.08874	0.08524	0.08442	0.08525
84	0.07455	0.09459	0.09404	0.09473
85	0.11195	0.10743	0.10730	0.10777
86	0.12320	0.12320	0.12358	0.12380
87	0.14031	0.14076	0.14171	0.14166
88	0.14879	0.15899	0.16052	0.16019
89	0.19020	0.17722	0.17932	0.17873
90	0.21628	0.19544	0.19812	0.19726
91	0.20718	0.21370	0.21698	0.21585
92	0.28866	0.23439	0.23826	0.23687
93	0.13362	0.26101	0.26544	0.26382
94	0.38265	0.29726	0.30218	0.30038

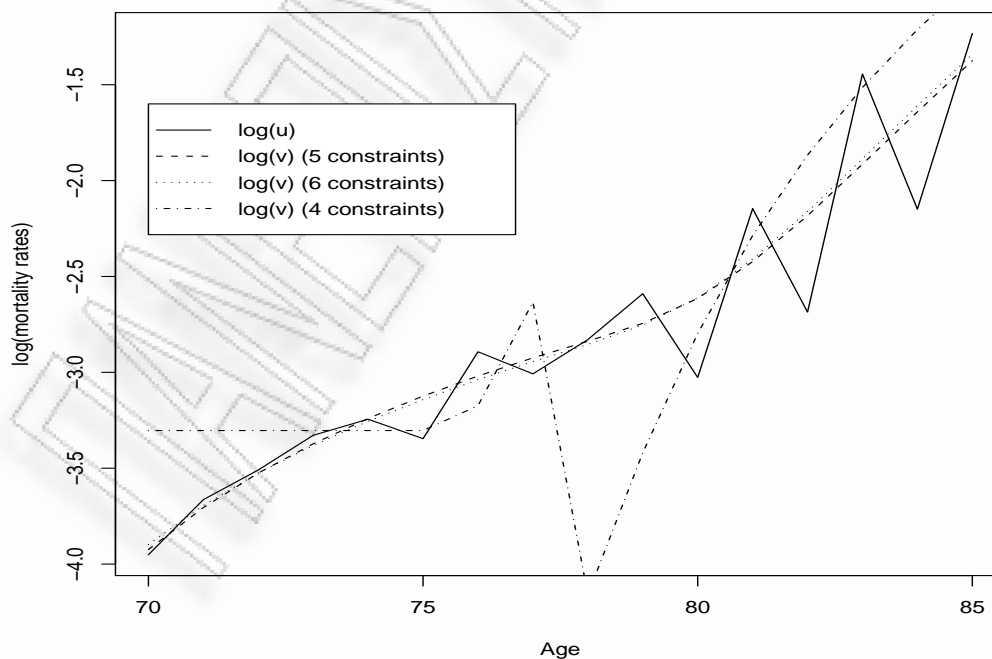
Σχήμα 6.25: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές μέσω Jensen difference (L85)



Πίνακας 6.18: Εξομάλυνση μέσω Jensen difference για το σύνολο δεδομένων ΗΚ01Μ

$x$	$u_x$	$v_x$ (5 περιορισμοί)	$v_x$ (6 περιορισμοί)	$v_x$ (4 περιορισμοί)
70	0.01923	0.01971	0.02027	0.02092
71	0.02563	0.02458	0.02485	0.02556
72	0.02992	0.02944	0.02942	0.03020
73	0.03585	0.03430	0.03400	0.03484
74	0.03899	0.03916	0.03857	0.03948
75	0.03523	0.04402	0.04314	0.04412
76	0.05543	0.04888	0.04794	0.04884
77	0.04939	0.05374	0.05273	0.05357
78	0.05906	0.05860	0.05768	0.05836
79	0.07503	0.06434	0.06384	0.06426
80	0.04848	0.07325	0.07347	0.07356
81	0.11692	0.08870	0.08993	0.08962
82	0.06816	0.11259	0.11514	0.11433
83	0.23598	0.14714	0.15134	0.14991
84	0.11659	0.19281	0.19902	0.19684
85	0.29152	0.25158	0.26006	0.25703

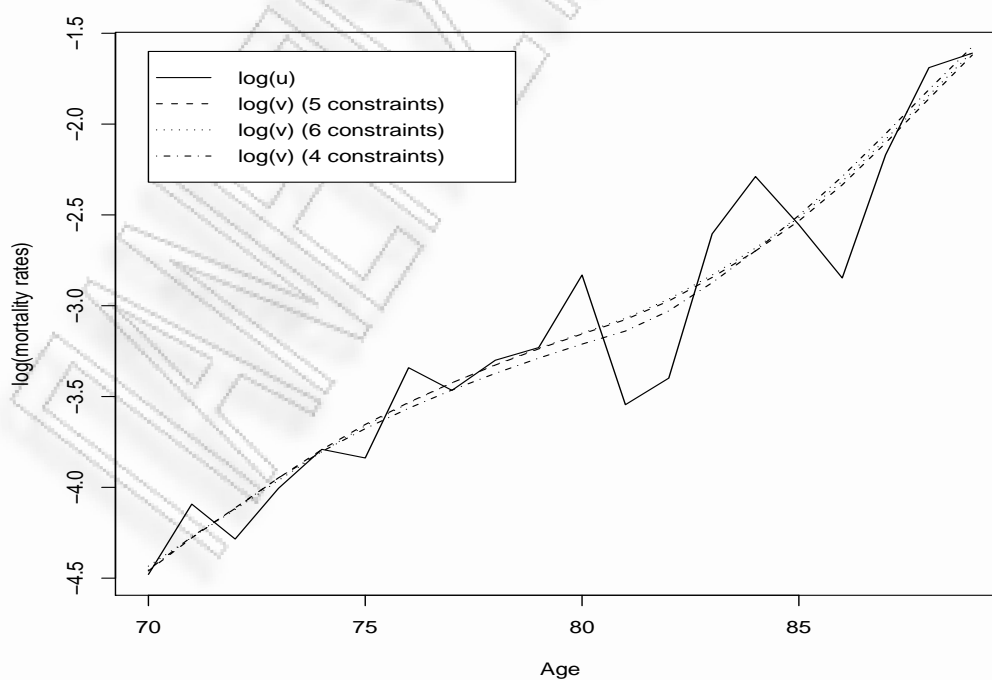
Σχήμα 6.26: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές μέσω Jensen difference (ΗΚ01Μ)



Πίνακας 6.19: Εξομάλυνση μέσω Jensen difference για το σύνολο δεδομένων HK01F

$x$	$u_x$	$v_x$ (5 περιορισμοί)	$v_x$ (6 περιορισμοί)	$v_x$ (4 περιορισμοί)
70	0.01134	0.01156	0.01187	0.01162
71	0.01670	0.01386	0.01390	0.01396
72	0.01378	0.01636	0.01621	0.01630
73	0.01822	0.01928	0.01903	0.01930
74	0.02258	0.02255	0.02229	0.02230
75	0.02153	0.02589	0.02571	0.02530
76	0.03539	0.02923	0.02913	0.02830
77	0.03124	0.03258	0.03255	0.03130
78	0.03688	0.03592	0.03597	0.03430
79	0.03954	0.03926	0.03940	0.03730
80	0.05896	0.04260	0.04282	0.04030
81	0.02887	0.04612	0.04646	0.04330
82	0.03342	0.05110	0.05163	0.04839
83	0.07395	0.05818	0.05896	0.05651
84	0.10138	0.06731	0.06840	0.06736
85	0.07776	0.07940	0.08088	0.08186
86	0.05800	0.09683	0.09875	0.10143
87	0.11429	0.12193	0.12436	0.12831
88	0.18465	0.15563	0.15861	0.16368
89	0.20000	0.19797	0.20156	0.20734

Σχήμα 6.27: Αδρές  $u_x$  και εξομαλυμένες  $v_x$  τιμές μέσω Jensen difference (HK01F)





Πίνακας 6.20: Τιμές των μέτρων ομαλότητας και προσαρμογής (Jensen difference)

	<i>L85</i>		
	5 περιορισμοί	6 περιορισμοί	4 περιορισμοί
<i>S</i>	0.00004	0.00004	0.000039
<i>F</i>	60.65	59.80	60.23
Deviance	63.52	62.85	63.44
log-likelihood	-17871.04	-17870.71	-17870.99
$\chi^2$	60.66	59.81	60.25
<i>HK01M</i>			
	5 περιορισμοί	6 περιορισμοί	4 περιορισμοί
<i>S</i>	0.00003	0.00003	0.00003
<i>F</i>	17.23	16.96	17.25
Deviance	17.15	17.08	17.36
log-likelihood	-2081.31	-2081.27	-2081.41
$\chi^2$	16.99	16.73	17.01
<i>HK01F</i>			
	5 περιορισμοί	6 περιορισμοί	4 περιορισμοί
<i>S</i>	0.000015	0.000015	0.000016
<i>F</i>	10.51	10.30	10.96
Deviance	10.71	10.49	10.88
log-likelihood	-1794.12	-1794.00	-1794.21
$\chi^2$	10.80	10.55	11.23

## 6.7 Χρήση άλλων μέτρων απόκλισης στην εξομάλυνση

Όπως έχουμε αναφέρει αρκετές φορές έως τώρα, μια από τις ομολογουμένως καλύτερες μεθόδους εξομάλυνσης, είναι αυτή που πρότειναν οι Whittaker and Henderson. Υπενθυμίζουμε ότι στη μέθοδο των Whittaker - Henderson ελαχιστοποιούμε τη συνάρτηση  $M_1 = F + hS = \sum_{x=1}^n w_x(u_x - v_x)^2 + h \sum_{x=1}^{n-3} (\Delta^3 v_x)^2$  ως προς τις εξομαλυμένες τιμές  $v_x$ , όπου  $w_x$  είναι κατάλληλα επιλεγμένα βάρη και συνήθως η παράμετρος  $h$  θεωρείται ως η μέση τιμή των  $w_x$ . Παραγωγίζοντας ως προς τα  $v_x$  οδηγούμαστε σε ένα γραμμικό σύστημα το οποίο μπορεί εύκολα να επιλυθεί.

Μια ιδέα που πήραμε από τη μέθοδο αυτή, είναι να αντικαταστήσουμε τη στατιστική συνάρτηση  $X^2$  με την απόκλιση των Kullback - Leibler. Άλλωστε, η απόκλιση των Kullback - Leibler είναι ένα ευρέως αποδεκτό μέτρο καλής προσαρμογής. Προσπαθήσαμε να ελαχιστοποιήσουμε τη συνάρτηση  $M_2 = F_1 + hS = \sum_{x=1}^n w_x u_x \ln \frac{u_x}{v_x} + h \sum_{x=1}^{n-3} (\Delta^3 v_x)^2$  πάλι παραγωγίζοντας ως προς τα  $v_x$  και χρησιμοποιώντας την προσέγγιση  $(\ln v_x)/v_x \approx (\ln u_x)/u_x$ , όπως προτείνεται από τον London (1985, σελ.11), αλλά χωρίς ικανοποιητικά αποτελέσματα. Συγκεκριμένα, η μέθοδος αυτή είτε δίνει εξομαλυμένες τιμές μεγαλύτερες από τις αδρές, το οποίο σημαίνει ότι τις υπερεκτιμά, είτε δεν δίνει την επιθυμητή μορφή της καμπύλης. Κατ' επέκταση, η προσαρμογή των δεδομένων δεν είναι ικανοποιητική. Αντίθετα, όσον αφορά την ομαλότητα, η μετατροπή του μοντέλου δίνει καλά αποτελέσματα.

## 6.8 Συμπεράσματα Κεφαλαίου

Στο κεφάλαιο αυτό, διερευνήσαμε μέσω ενός αριθμητικού παραδείγματος με τρία σύνολα δεδομένων, τη χρήση της γενικής οικογένειας των μέτρων απόκλισης δύναμης των Cressie - Read στο πρόβλημα της εξομάλυνσης ποσοστών θνησιμότητας καθώς και τη συμπεριφορά στο ίδιο πρόβλημα της διαφοράς του Jensen. Η αριθμητική διερεύνηση έδειξε ότι, η ελαχιστοποίηση των αποκλίσεων δύναμης για διάφορες τιμές της παραμέτρου  $\lambda$ , υπό τους περιορισμούς (i) - (v) και/ή (vi), έδωσε ισοδύναμα αποτελέσματα, από πλευράς ομαλότητας, με αυτά άλλων μεθόδων εξομάλυνσης, όπως είναι η ευρέως χρησιμοποιούμενη μέθοδος των Whittaker - Henderson.

Μπορούμε να λάβουμε διαφορετική τιμή για το μέτρο της ομαλότητας  $S$  αλλάζοντας την τιμή της παραμέτρου  $M$  στον πρώτο περιορισμό. Βέβαια, δεν μπορούμε ευθέως να πούμε ποια τιμή της παραμέτρου  $\lambda$  είναι η καλύτερη για εξομάλυνση. Για τις εξομαλύνσεις υπό τους περιορισμούς (i) - (v) έχουμε τα εξής: Τιμές του  $\lambda < -1$  δίνουν μη ικανοποιητικά αποτελέσματα ως προς την καλή προσαρμογή και για το λόγο αυτό θα πρέπει να αποφεύγονται. Όσον αφορά την ομαλότητα  $S$ , τιμές του  $\lambda \in (-1, 0.5)$ , υπερεξομαλύνουν τα δεδομένα. Για  $\lambda \geq 0$  οι διάφορες αποκλίσεις δίνουν παρόμοια αποτελέσματα ως προς την ομαλότητα και

την προσαρμογή, συνεπώς η τιμή  $\lambda = 2/3$  που πρότειναν οι Cressie and Read (1984) από πλευράς στατιστικής ισχύος, είναι μια καλή επιλογή για εξομάλυνση.

Η χρήση του περιορισμού (vi)  $\sum_{x=1}^n u_x = \sum_{x=1}^n v_x$  που προτείνουμε, θεωρείται απαραίτητη καθώς αυτή είναι η ελάχιστη απαίτηση για ένα μέτρο χωρίς διανύσματα πιθανότητας να είναι μέτρο απόκλισης. Τα αριθμητικά αποτελέσματα υποστηρίζουν τη χρήση της επιπλέον συνθήκης και έδειξαν αξιοσημείωτη βελτίωση στην προσαρμογή των εξομαλυμένων τιμών. Επιπλέον, στην περίπτωση αυτή, η τιμή της παραμέτρου  $\lambda$  φαίνεται πλέον να μην επηρεάζει την ομαλότητα  $S$ .

Η ομοιότητα των αποτελεσμάτων μεταξύ της μεθόδου των Whittaker - Henderson και της ελαχιστοποίησης των αποκλίσεων δύναμης υπό τους συγκεκριμένους περιορισμούς, μας δίνει τη δυνατότητα να υποστηρίξουμε ότι οι δυο μέθοδοι εξομάλυνσης είναι περίπου ισοδύναμες. Η άποψη αυτή δεν υποστηρίζεται μόνο από την αριθμητική διερεύνηση του θέματος αλλά και από το γεγονός ότι στη μέθοδο των Whittaker - Henderson ελαχιστοποιούμε τη Λαγκρανζιανή συνάρτηση  $F + hS$ , ενώ στη μέθοδο των αποκλίσεων δύναμης ελαχιστοποιούμε την  $F$ , μεταξύ άλλων, υπό έναν περιορισμό  $S$ , η οποία τελικά οδηγεί σε μια παρόμοια Λαγκρανζιανή συνάρτηση.

Όσον αφορά την ελαχιστοποίηση της διαφοράς του Jensen, τα αποτελέσματα είναι συγκρίσιμα με τις εξομαλύνσεις που γίνονται μέσω της απόκλισης των Kullback - Leibler, των Cressie - Read και της μεθόδου των Whittaker - Henderson.

Μια σημαντική επέκταση της μεθόδου μας θα ήταν η πρόβλεψη μελλοντικών τιμών. Δυστυχώς τα μοντέλα μας δεν έχουν προβλεπτικό χαρακτήρα. Μια σκέψη είναι να χρησιμοποιήσουμε ένα παραμετρικό μοντέλο που θα προσαρμόσουμε σε ένα μέρος των εξομαλυμένων τιμών και στη συνέχεια να χρησιμοποιήσουμε τις υπόλοιπες εξομαλυμένες τιμές για προβλεπτικούς σκοπούς (time based training-test split). Συγκεκριμένα, μπορούμε να διαιρέσουμε τα σύνολα δεδομένων σε δυο ίσα διαστήματα, να εξομαλύνουμε το πρώτο διάστημα μέσω της ελαχιστοποίησης της απόκλισης των Cressie - Read, να προσαρμόσουμε το παραμετρικό μοντέλο και στη συνέχεια να συγκρίνουμε τις προβλεπόμενες τιμές, που παράγονται από το μοντέλο, με τις αδρές τιμές του δεύτερου διαστήματος. Η προσέγγιση αυτή όμως έχει ένα σημαντικό μειονέκτημα. Τα αποτελέσματα εξαρτώνται από το επιλεγμένο παραμετρικό μοντέλο. Μια καλύτερη προσέγγιση είναι η ελαχιστοποίηση της απόκλισης των Cressie - Read  $D^{CR}(f(\mathbf{x}), \mathbf{u})$  ως προς συναρτήσεις  $f$ , όπου το διάνυσμα  $\mathbf{v}$  έχει αντικατασταθεί από μια άγνωστη συνάρτηση  $f(\mathbf{x})$  των ηλικιών  $\mathbf{x}$ , υπό περιορισμούς ανάλογους με τους περιορισμούς (i) - (vi). Αυτό είναι ένα πρόβλημα λογισμού μεταβολών και έχουμε την υπόνοια ότι η λύση του θα είναι μια συνάρτηση spline. Το θέμα αυτό θα αποτελέσει αντικείμενο μελλοντικής έρευνας.

# РАНЕЕЗНАМО ПЕРПАА

## ΚΕΦΑΛΑΙΟ 7

# ΥΠΟΛΟΓΙΣΜΟΣ ΕΝΤΡΟΠΙΑΣ ΓΙΑ ΚΑΤΑΝΟΜΕΣ ΑΠΩΛΕΙΑΣ

### 7.1 Εισαγωγή

Τα παραμετρικά μαθηματικά μοντέλα αποτελούν ένα χρήσιμο εργαλείο στη Στατιστική και κυρίως στην Αναλογιστική Επιστήμη για τη μοντελοποίηση διάφορων τυχαίων φαινομένων (Hogg and Kluggman, 1984). Είναι γνωστά ως *Μοντέλα Απώλειας* (Loss Models) όταν σχετίζονται με ασφαλίσεις περιουσίας και ευθύνης ενώ ως *Μοντέλα Επιβίωσης* (Survival Models) όταν σχετίζονται με ασφαλίσεις ζωής και αναπηρίας. Τα ίδια μοντέλα έχουν χρησιμοποιηθεί επίσης και σε άλλους τομείς, όπως είναι η Θεωρία Αξιοπιστίας, τα οικονομικά και τα χρηματοοικονομικά.

Οι πιο γνωστές κατανομές απώλειας είναι η *εκθετική κατανομή* με συνάρτηση πυκνότητας πιθανότητας (σ.π.π.)

$$f(x) = \lambda e^{-\lambda x},$$

όπου  $\lambda > 0$  και  $x \geq 0$ , η *κατανομή Pareto* με σ.π.π.

$$f(x) = \alpha \lambda^\alpha (\lambda + x)^{-(\alpha+1)},$$

όπου  $\alpha > 0$ ,  $\lambda > 0$  και  $x > 0$ , η *κατανομή Weibull* με σ.π.π.

$$f(x) = c\tau x^{\tau-1} e^{-cx^\tau},$$

όπου  $c > 0$ ,  $\tau > 0$  και  $x > 0$ , η *κατανομή Gamma* με σ.π.π.

$$f(x) = \lambda^\alpha x^{\alpha-1} e^{-\lambda x} / \Gamma(\alpha),$$

όπου  $\alpha > 0$ ,  $\lambda > 0$  και  $x > 0$ , (για  $\alpha$  ακέραιο προκύπτει η κατανομή Erlang), η *Transformed*

Gamma κατανομή με σ.π.π.

$$f(x) = \lambda^\alpha \tau x^{\tau\alpha-1} e^{-(\lambda x)^\tau} / \Gamma(\alpha),$$

όπου  $\alpha > 0$ ,  $\lambda > 0$ ,  $\tau > 0$  και  $x > 0$ , η κατανομή Loggamma με σ.π.π.

$$f(x) = \lambda^\alpha (\ln x)^{\alpha-1} / (x^{\lambda+1} \Gamma(\alpha)),$$

όπου  $\alpha > 0$ ,  $\lambda > 0$  και  $x > 1$ , η Lognormal κατανομή με σ.π.π.

$$f(x) = \exp \left\{ -\frac{1}{2} \left( \frac{\ln x - \mu}{\sigma} \right)^2 \right\} / (x \sigma \sqrt{2\pi}),$$

όπου  $-\infty < \mu < \infty$ ,  $\sigma > 0$  και  $x > 0$ , η κατανομή Burr με σ.π.π.

$$f(x) = \alpha \tau \lambda^\alpha x^{\tau-1} (\lambda + x^\tau)^{-(\alpha+1)},$$

όπου  $\alpha > 0$ ,  $\lambda > 0$ ,  $\tau > 0$  και  $x > 0$ , και η κατανομή Generalized Pareto με κατανομή

$$f(x) = \Gamma(\alpha + k) \lambda^\alpha x^{k-1} / (\Gamma(\alpha) \Gamma(k) (\lambda + x)^{k+\alpha}),$$

όπου  $\alpha > 0$ ,  $\lambda > 0$ ,  $k > 0$  και  $x > 0$ .

Όπως έχουμε ήδη αναφέρει, η κυρίαρχη έννοια της Στατιστικής Θεωρίας Πληροφοριών είναι η εντροπία του Shannon, η οποία δίνεται από τη σχέση

$$H(X) = - \sum_x p(x) \ln p(x)$$

ή

$$H(X) = - \int f(x) \ln f(x) dx$$

ανάλογα με το εάν η τυχαία μεταβλητή  $X$  είναι διακριτή ή συνεχής, αντίστοιχα (Shannon, 1948). Στη δεύτερη περίπτωση, τη συνεχή, η  $H(X)$  ονομάζεται επίσης και διαφορική εντροπία (differential entropy). Στο παρόν κεφάλαιο θα ασχοληθούμε με τη διαφορική εντροπία καθώς οι υπό μελέτη κατανομές είναι συνεχείς. Η εντροπία ποσοτικοποιεί την αναμενόμενη αβεβαιότητα που σχετίζεται με ένα πείραμα. Με άλλα λόγια παρέχει πληροφορία για την προβλεψιμότητα του αποτελέσματος μιας τυχαίας μεταβλητής  $X$ . Όσο μεγαλύτερη είναι η εντροπία τόσο λιγότερο συγκεντρωμένη είναι η κατανομή της  $X$  και συνεπώς μια παρατήρηση της  $X$  παρέχει λίγη πληροφορία. Η εντροπία του Shannon, μαζί με άλλα μέτρα εντροπίας όπως η εντροπία του Rényi, μπορεί να θεωρηθεί ως περιγραφική ποσότητα της αντίστοιχης συνάρτησης πυκνότητας πιθανότητας. Η εντροπία θεωρείται ως μέτρο μεταβλητότητας για

Πίνακας 7.1: Εκφράσεις εντροπίας

Κατανομή	Εντροπία
Εκθετική	$1 - \ln \lambda$
Pareto	$1 + \frac{1}{\alpha} + \ln \lambda - \ln \alpha$
Weibull	$1 - \ln c - \ln \tau + (\tau - 1) \frac{\gamma + \ln c}{\tau}$
Gamma	$\alpha + \ln \Gamma(\alpha) - (\alpha - 1)\psi(\alpha) - \ln \lambda$
Tr. Gamma	$\alpha - \ln \tau - \ln \lambda + \ln \Gamma(\alpha) + (\frac{1}{\tau} - \alpha)\psi(\alpha)$
Loggamma	$\alpha (1 + \frac{1}{\lambda}) + \ln \Gamma(\alpha) - (\alpha - 1)\psi(\alpha) - \ln \lambda$
Lognormal	$\frac{1}{2} + \mu + \ln(\sigma\sqrt{2\pi})$
Burr	$1 + \frac{1}{\alpha} + \frac{\ln \lambda}{\tau} - \ln \lambda + \ln \tau - (\frac{1}{\tau-1})(\gamma + \psi(\alpha))$
Gen. Pareto	$\ln \frac{\lambda \Gamma(\alpha+k)}{\Gamma(\alpha)\Gamma(k)} - (\alpha - 1)\psi(\alpha) - (k - 1)\psi(k) + (\alpha + k)\psi(\alpha + k)$

συνεχείς μεταβλητές ή ως μέτρο variation or diversity για τις πιθανές τιμές μιας διακριτής μεταβλητής (Harris, 1982; Nadarajah and Zografos, 2003; Zografos, 2008). Άλλα μέτρα εντροπίας και γενικεύσεις αυτών έχουν επίσης προταθεί. Για περισσότερες πληροφορίες ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο Κεφάλαιο 2 αυτής της διατριβής καθώς επίσης και στους Arndt (2001) και Pardo (2006). Η εξαγωγή ακριβών εκφράσεων για την εντροπία του Shannon αποτελεί αντικείμενο μελέτης σε πολλές περιπτώσεις. Διάφοροι συγγραφείς έχουν ασχοληθεί με το θέμα, μεταξύ των οποίων οι Ahmed and Gokhale (1989), Guerrero - Cususmano (1996) και πιο πρόσφατα οι Darbellay and Vajda (2000), Nadarajah and Zografos (2003) και Zografos and Nadarajah (2005).

Ο Shannon (1948) χρησιμοποίησε τις διαφορές των εντροπιών για τη σύγκριση των κατανομών. Οι Kullback and Leibler (1951) εισήγαγαν τη συνάρτηση διαχωριστικής πληροφορίας μεταξύ δυο κατανομών. Από τότε οι στατιστικοί επιστήμονες και οι οικονομέτρους έχουν αναπτύξει μέτρα ή δείκτες πληροφορίας για κατηγορικά δεδομένα, παλινδρόμηση, συσχέτιση, ταυτόχρονες εξισώσεις, χρονολογικές σειρές, έλεγχο υποθέσεων, περιγραφή ανισοτήτων εισοδήματος κ.α.

Η εντροπία, τόσο η διακριτή όσο και η διαφορική, έχει ευρύτατα διαδεδομένες εφαρμογές. Πέραν αυτού ο κυρίαρχος ρόλος που έχει αυτή στη Θεωρία Επικοινωνιών, στο σχηματισμό της capacity of a channel και άλλων εννοιών, στη Στατιστική και ειδικότερα στην Αναλογιστική Επιστήμη χρησιμοποιείται υπό το πλαίσιο της Αρχής της Μέγιστης Εντροπίας. Οι εκφράσεις της εντροπίας του Shannon για τις πιο πάνω αναφερθείσες κατανομές παρουσιάζονται στον Πίνακα 7.1. Σημειώνουμε ότι οι παραπάνω εκφράσεις έχουν παρουσιαστεί, υπό διαφορετικό πλαίσιο, μεταξύ άλλων στους Lazo and Rathie (1978) και στους Nadarajah and Zografos (2003).

Υψηλές αβέβαιες ασφαλίσεις είναι λιγότερο αξιόπιστες. Η αβεβαιότητα για τις απώλειες των ασφαλίσεων μπορεί να ποσοτικοποιηθεί μέσω των αντίστοιχων κατανομών απώλειας. Παρόλα αυτά, συχνά στην αναλογιστική πρακτική, ο αναλογιστής έχει στα χέρια του μετασχηματισμένα δεδομένα ως συνέπεια περιπτώσεων όπως ο πληθωρισμός και η αποκοπή.

Έτσι είναι εύκολα κατανοητό ότι η εντροπία δεν μπορεί να υπολογιστεί από τις εκφράσεις που παρουσιάστηκαν στον Πίνακα 7.1 όταν υπεισέρχονται ποσότητες όπως ο ρυθμός πληθωρισμού, τα όρια φερεγγυότητας και τα αφαιρετέα ποσά. Στην Ενότητα 7.2 μελετάμε την επίδραση του πληθωρισμού στην εντροπία ενώ η επίδραση της αποκοπής από κάτω μελετάται στην Ενότητα 7.3. Ο τρόπος με τον οποίο η αποκοπή από πάνω επηρεάζει την εντροπία μελετάται στην Ενότητα 7.4. Παρουσιάζονται αναλυτικές εκφράσεις σε κάθε περίπτωση.

## 7.2 Επίδραση του πληθωρισμού

Η βασική επίδραση που έχει ο πληθωρισμός είναι η αύξηση στις απώλειες. Το μοντέλο θα πρέπει να τροποποιηθεί έτσι ώστε να εκφράζει το τρέχον επίπεδο της κατάστασης των απωλειών καθώς, γενικά, το μοντέλο έχει εκτιμηθεί από παρατηρήσεις που έχουν συλλεχθεί στο παρελθόν. Επιπλέον, μπορεί να επιθυμούμε μια προβολή (projection) η οποία θα αντανακλά τις ζημιές σε μελλοντικές χρονικές περιόδους.

Στη συνέχεια θα διερευνήσουμε την επίδραση του πληθωρισμού στη διαφορική εντροπία (Sachlas and Papaioannou, 2009b). Έστω  $X$  μια τυχαία μεταβλητή, η οποία μοντελοποιεί τις απώλειες. Η τυχαία μεταβλητή που μοντελοποιεί τις απώλειες υπό πληθωρισμό είναι η  $Z(r) = (1+r)X$ , όπου  $r$ ,  $0 < r < 1$ , είναι ο ετήσιος ρυθμός του πληθωρισμού. Με άλλα λόγια εάν  $X$  είναι η τιμή των απωλειών και  $r$  ο ετήσιος ρυθμός πληθωρισμού, μετά από ένα έτος η απώλεια θα έχει τιμή  $Z(r) = (1+r)X$ . Η συνάρτηση κατανομής της  $Z(r)$  που σχετίζεται με αυτή της  $X$  είναι

$$F_{Z(r)}(z) = Pr(Z(r) \leq z) = Pr\left(\frac{Z(r)}{1+r} \leq \frac{z}{1+r}\right) = F_X\left(\frac{z}{1+r}\right),$$

όπου  $F_X(\cdot)$  είναι η συνάρτηση κατανομής της  $X$ . Οι δυο συναρτήσεις πυκνότητας πιθανότητας,  $F_{Z(r)}(z)$  και  $F_X(x)$ , σχετίζονται μέσω της σχέσης

$$f_{Z(r)}(z) = \frac{\partial}{\partial z} F_{Z(r)}(z) = \frac{1}{1+r} f_X\left(\frac{z}{1+r}\right).$$

Η σχέση μεταξύ των εντροπιών της  $X$  και της  $Z(r) = (1+r)X$ , δίνεται στο ακόλουθο λήμμα.

**Λήμμα 7.1.** Η διαφορική εντροπία της τυχαίας μεταβλητής  $Z(r) = (1+r)X$ , όπου  $r$  είναι ο ρυθμός πληθωρισμού,  $r \in (0, 1)$ , δίνεται από τη σχέση

$$H(Z(r), r) = H(X) + \ln(1+r),$$

όπου  $H(X)$  είναι η εντροπία της  $X$ .



*Απόδειξη.* Το παραπάνω λήμμα αποτελεί εφαρμογή του Θεωρήματος 9.6.4 των Cover and Thomas (1991, σελ. 233) και συνεπώς η απόδειξη παραλείπεται. ■

Όπως γίνεται αντιληπτό από το παραπάνω λήμμα, οι εκφράσεις αυτές μπορούν να βρεθούν προσθέτοντας στις εκφράσεις του Πίνακα 7.1 τη σταθερά  $\ln(1+r)$ . Στη συνέχεια ακολουθούν δυο χρήσιμες προτάσεις.

**Πρόταση 7.1.** *Η εντροπία της  $Z(r)$  είναι πάντα μεγαλύτερη από αυτή της  $X$  και είναι μια αύξουσα συνάρτηση ως προς  $r$ .*

*Απόδειξη.* Η απόδειξη είναι προφανής δοθέντος ότι ισχύει  $\ln(1+r) > 0$  για  $r \in (0, 1)$ . ■

Όπως έχουμε ήδη αναφέρει, η εντροπία ποσοτικοποιεί την αβεβαιότητα που σχετίζεται με το αποτέλεσμα ενός πειράματος. Συνεπώς καθώς αυξάνει η εντροπία αυξάνει και η αβεβαιότητα. Άρα η παραπάνω πρόταση μας λέει ότι στην περίπτωση των κατανομών απώλειας υπό πληθωρισμό, θα έχουμε πάντα μεγαλύτερη αβεβαιότητα εν συγκρίσει με τις απώλειες χωρίς πληθωρισμό. Επιπρόσθετα, η αβεβαιότητα για τις απώλειες αυξάνει όταν αυξάνεται ο ρυθμός του πληθωρισμού. Πιο συγκεκριμένα, καθώς η διαφορική εντροπία θεωρείται ως ένα περιγραφικό μέτρο της μεταβλητότητας μιας κατανομής, όσο περισσότερο αυξάνει ο πληθωρισμός τόσο αυξάνει και η μεταβλητότητα των απωλειών ή του κινδύνου των επιστροφών.

### 7.3 Επίδραση της αποκοπής από κάτω

Έστω ότι οι απώλειες δεν καταγράφονται όταν είναι μικρότερες από ένα προκαθορισμένο ποσό. Στην περίπτωση αυτή τα δεδομένα είναι αποκομμένα από κάτω (truncated from below) ή αριστερά αποκομμένα (left truncated). Ο πιο συνήθης λόγος αριστερής αποκοπής είναι η χρήση ενός αφαιρετέου ποσού (deductible). Έστω  $X$  η τυχαία μεταβλητή των απωλειών και  $d$  η τιμή του αφαιρετέου ποσού. Τότε η τυχαία μεταβλητή των παρατηρούμενων τιμών δίνεται από τη σχέση

$$Y(d) = \begin{cases} X, & X > d \\ \text{δεν ορίζεται,} & X \leq d. \end{cases}$$

Ας διερευνήσουμε τώρα την επίδραση που έχει στην εντροπία η αποκοπή από κάτω (Sachlas and Papaioannou, 2009b). Η διαφορική εντροπία υπό αυτή την περίπτωση είναι μια χρήσιμη ποσότητα καθώς μετρά την αβεβαιότητα σχετικά με την απώλεια που καλύπτει η ασφαλιστική εταιρεία. Η σχέση που συνδέει τις εντροπίες των τυχαίων μεταβλητών  $X$  και  $Y(d)$  δίνεται στο παρακάτω λήμμα.

**Λήμμα 7.2.** Η εντροπία της τυχαίας μεταβλητής  $Y(d)$  των παρατηρούμενων αποκομμένων από κάτω δεδομένων, δίνεται από τη σχέση

$$H_{X|X>d} = H_{Y(d)} = \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_0^d f_X(x) \ln f_X(x) dx + \bar{F}_X(d) \ln \bar{F}_X(d) \right],$$

όπου  $H_X$  είναι η εντροπία της κατανομής  $X$ ,  $\bar{F}_X(x) = 1 - F_X(x)$ , όπου  $F_X(x)$  είναι η συνάρτηση κατανομής της  $X$  και  $d$  είναι το αφαιρετέο ποσό.

Απόδειξη. Είναι γνωστό ότι η συνάρτηση κατανομής της  $Y(d)$  δίνεται από τη σχέση

$$f_{Y(d)}(x) = \begin{cases} \frac{f_X(x)}{1-F_X(d)} = \frac{\bar{F}_X(d)\lambda_X(x)}{F_X(d)}, & X > d \\ 0, & X \leq d, \end{cases}$$

όπου  $\lambda_X(x)$  είναι η συνάρτηση κινδύνου (hazard function) της  $X$ . Συνεπώς η εντροπία της  $Y(d)$  δίνεται από τη σχέση

$$\begin{aligned} H_{Y(d)} &= - \int_d^\infty \frac{f_X(x)}{\bar{F}_X(d)} \ln \frac{f_X(x)}{\bar{F}_X(d)} dx \\ &= - \frac{1}{\bar{F}_X(d)} \left[ \int_d^\infty f_X(x) \ln f_X(x) dx - \int_d^\infty f_X(x) \ln \bar{F}_X(d) dx \right] \\ &= \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_0^d f_X(x) \ln f_X(x) dx + \bar{F}_X(d) \ln \bar{F}_X(d) \right], \end{aligned}$$

όπου  $\bar{F}_X(d) = 1 - F_X(d)$  είναι η συνάρτηση επιβίωσης (survival function) στο σημείο του αφαιρετέου ποσού  $d$ . ■

Μια έκφραση της  $H_{Y(d)}$  σε σχέση με τη συνάρτηση κινδύνου  $\lambda_X(x)$  της  $X$  παρουσιάστηκε από τον Ebrahimi (1996):

$$H_{Y(d)} = 1 - \frac{1}{\bar{F}_X(d)} \int_d^\infty (\log \lambda_X(x)) f_X(x) dx.$$

Η παραπάνω έκφραση παρουσιάστηκε από τον Ebrahimi (1996) ως μέτρο της αβεβαιότητας της κατανομής του υπολοιπούμενου χρόνου ζωής (residual lifetime distribution), δηλαδή δοθέντος ότι ένα στοιχείο (item) έχει επιζήσει έως τη χρονική στιγμή  $t$  και είναι μια δυναμική έκδοση της κλασικής εντροπίας. Ανάλογα μέτρα πληροφορίας για πολυμεταβλητές κατανομές με προοδευτικά αποκομμένο στήριγμα έχουν προταθεί από τους Ebrahimi et al. (2007). Όμως, αναλυτικές εκφράσεις δεν έχουν παρουσιαστεί στη βιβλιογραφία, εκτός από τις περιπτώσεις της εκθετικής κατανομής και της κατανομής Pareto (Belzunce et al., 2004).

Οι αναλυτικές εκφράσεις για την εντροπία του Shannon όταν τα δεδομένα είναι αποκομμένα από κάτω για τα πιο γνωστά μοντέλα απώλειας παρουσιάζονται στον Πίνακα 7.2. Τα

περισσότερα από τα ολοκληρώματα εδώ καθώς και σε ολόκληρο το κεφάλαιο υπολογίστηκαν αναλυτικά και μέσω του μαθηματικού προγράμματος Mathematica. Η αποκοπή από επάνω δεν επηρεάζει την εντροπία της εκθετικής κατανομής. Παρατηρούμε ότι η έκφραση για κάποια από τα μοντέλα απώλειας είναι αρκετά περίπλοκη, καθώς εμφανίζονται σε αυτή συναρτήσεις όπως η συνάρτηση Meijer

$$G_{m,n}^{p,q} \left( z \left| \begin{matrix} \mathbf{w} \\ \mathbf{x} \end{matrix} \right. \right) = \frac{1}{2\pi i} \int_L \frac{\prod_{j=1}^m \Gamma(x_j - s) \prod_{j=m+1}^q \Gamma(1 - w_j + s)}{\prod_{j=1}^n \Gamma(1 - x_j + s) \prod_{n=1}^p \Gamma(w_j - s)} z^s ds,$$

$\mathbf{w} = (w_1, \dots, w_q)^T$ ,  $\mathbf{x} = (x_1, \dots, x_q)^T$ , η συνάρτηση σφάλματος (error function)  $\text{Erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$  και η συμπληρωματική συνάρτηση σφάλματος (complementary error function)  $\text{Erfc}(z) = 1 - \text{Erf}(z)$  (Abramowitz and Stegun, 1972). Για κάποια μοντέλα, όπως η κατανομή Burr και η γενικευμένη κατανομή Pareto, η εντροπία του Shannon των παρατηρούμενων αποκομμένων από κάτω δεδομένων δεν υπολογίζεται σε κλειστή μορφή, συνεπώς για τον υπολογισμό της πρέπει να χρησιμοποιηθεί μια αριθμητική μέθοδος. Αναλυτικοί υπολογισμοί παρουσιάζονται στο Παράρτημα Β.1.

Η πολυπλοκότητα των εκφράσεων δεν μας επιτρέπει να δώσουμε μια γενική και ευθεία ερμηνεία της επίδρασης της αποκοπής από κάτω στην εντροπία. Παρόλα αυτά, μπορούμε να δούμε ότι η αποκοπή από κάτω δεν επηρεάζει την εντροπία της εκθετικής κατανομής.

Η εντροπία της  $Y(d)$  έχει μια επιπρόσθετη ιδιότητα: Εάν οι απώλειες  $X$  είναι μεγαλύτερες από το αφαιρετέο ποσό  $d$ , η  $H_{Y(d)}$  μετρά την αναμενόμενη αβεβαιότητα των υπολοίπων απωλειών  $X - d$  δοθέντος ότι  $X > d$ . Έστω τώρα ότι  $\delta(d)$  είναι η μέση υπολοίπιμη απώλεια δοθέντος ότι  $X > d$ , δηλαδή

$$\delta(d) = E(X - d | X > d) = \begin{cases} \frac{\int_d^\infty \bar{F}_X(x) dx}{F_X(d)}, & d < d^* \\ 0, & \text{διαφορετικά,} \end{cases}$$

όπου  $d^* = \sup \{x : \bar{F}(x) > 0\}$ . Για τα μοντέλα που μελετάμε στο παρόν κεφάλαιο ισχύει ότι  $d^* = \infty$ . Σημειώνουμε ότι η  $\delta(d)$  μπορεί επίσης να γραφεί ως

$$\delta(d) = \int_d^\infty \exp\left(-\int_d^y \lambda_X(z) dz\right) dy = E(Y(d) - d).$$

Τα επόμενα δυο αποτελέσματα, αρχικά παρουσιασμένα από τον Ebrahimi (1996), στο πλαίσιο των υπολοίπων χρόνων ζωής, παρέχουν επιπλέον insight στην εντροπία των υπολοίπων απωλειών.

**Θεώρημα 7.1.** Έστω  $\delta(d) < \infty$ , τότε

$$H(Y, d) \leq 1 + \ln \delta(d).$$

Πίνακας 7.2: Εντροπία υπό αποκοπή από κάτω

Κατανομή	Εντροπία
Εκθετική	$1 - \ln \lambda$
Pareto	$1 + \frac{1}{\alpha} - \ln \alpha + (1 + 2\alpha) \ln(\lambda + d)$
Weibull	$1 - \ln c - \ln \tau - (\tau - 1) \ln d - \frac{(\tau-1)e^{cd^\tau} \Gamma(0, cd^\tau)}{\tau}$
Gamma	$\frac{1}{\varphi(\alpha, \lambda d)} \left\{ \Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda d) + (\alpha - 1) \Gamma(\alpha, \lambda d) \ln d + (\alpha - 1) G_{2,0}^{3,0} \left( \lambda d \left  \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix} \right. \right) - \Gamma(\alpha) \left[ \alpha + (\alpha - 2) \ln \lambda + \ln \xi(\alpha, \lambda d) + \Gamma(\alpha, \lambda d) \ln \frac{\lambda^\alpha}{\xi(\alpha, \lambda d)} \right] \right\}$
Tr. Gamma	$\frac{1}{\varphi(\alpha, (\lambda d)^\tau)} \left\{ \frac{\tau\alpha-1}{\tau} G_{2,0}^{3,0} \left( (\lambda d)^\tau \left  \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix} \right. \right) + \Gamma(\alpha + 1) - \Gamma(\alpha + 1, (\lambda d)^\tau) + (\tau\alpha - 1) \Gamma(\alpha, (\lambda d)^\tau) \ln d - \Gamma(\alpha) \left[ \alpha + \ln \frac{\lambda^{\tau\alpha\tau}}{\xi(\alpha, (\lambda d)^\tau)} + \Gamma(\alpha, (\lambda d)^\tau) \ln \frac{\lambda^{\tau\alpha\tau}}{\xi(\alpha, (\lambda d)^\tau)} \right] \right\}$
Loggama	$\frac{1}{\lambda\varphi(\alpha, \lambda \ln d)} \left\{ \lambda(\alpha - 1) G_{2,0}^{3,0} \left( \lambda \ln d \left  \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix} \right. \right) + (\alpha - 1) \lambda \Gamma(\alpha, \lambda \ln d) \ln(\ln d) - (\lambda + 1) \Gamma(\alpha + 1, \lambda \ln d) + \lambda(1 - \Gamma(\alpha, \lambda \ln d)) [\Gamma(\alpha + 1) \ln \lambda - \Gamma(\alpha) \ln \xi(\alpha, \lambda \ln d)] \right\}$
Lognormal	$\frac{1}{1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)} \left\{ \frac{1}{2} + \mu + \ln(\sigma\sqrt{2\pi}) - \left( \ln(\sigma\sqrt{2\pi}) + \frac{\mu^2}{\sigma^2} \right) \Phi\left(\frac{\ln d - \mu}{\sigma}\right) + (1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)) \ln\left(1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)\right) - \frac{1}{2\sigma^2} \left[ \frac{\mu^2 + \sigma^2}{2} \operatorname{Erfc}\left(\frac{\mu - \ln d}{\sqrt{2}\sigma}\right) - \frac{\sigma(\mu + \ln d)}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln d)^2}{2\sigma^2}} \right] + \left(\frac{\mu}{\sigma^2} - 1\right) \left[ -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln d)^2}{2\sigma^2}} + \frac{\mu}{2} - \frac{\mu}{2} \operatorname{Erf}\left(\frac{\mu - \ln d}{\sqrt{2}\sigma}\right) \right] \right\}$
Burr	όχι σε κλειστή μορφή
Gen. Pareto	όχι σε κλειστή μορφή

$$\varphi(\alpha, x) = \Gamma(\alpha) (\Gamma(\alpha, x) - 1) \quad \mathbf{a} = (1, 1)$$

$$\xi(\alpha, x) = \Gamma(\alpha) (1 - \Gamma(\alpha, x)) \quad \mathbf{b} = (0, 0, \alpha)$$

Πίνακας 7.3: Άνω φράγμα της εντροπίας υπό αποκοπή από κάτω

Κατανομή	Φράγμα
Εκθετική	$1 - \ln \lambda$
Pareto	$1 - (\alpha - 1) \ln(\lambda + d) - \ln(\alpha - 1)$
Weibull	$1 + cd^\tau + \ln \left[ \frac{e^{-\frac{1}{\tau}}}{\tau} \Gamma \left( \frac{1}{\tau}, cd^\tau \right) \right]$
Gamma	$1 + \ln \left( \frac{d(\lambda d)^\alpha}{\alpha} \right) + \ln [e^{-\lambda d} + (\alpha - \lambda d) E_{-\alpha}(\lambda d)]$
Tr. Gamma	όχι σε κλειστή μορφή
Loggamma	όχι σε κλειστή μορφή
Lognormal	όχι σε κλειστή μορφή
Burr	$1 + \ln \left[ \frac{d^{-(\tau\alpha-1)}}{\tau\alpha-1} {}_2F_1 \left( \alpha, \alpha - \frac{1}{\tau}; 1 + \alpha - \frac{1}{\tau}, -\lambda d^{-\tau} \right) \right]$
Gen. Pareto	όχι σε κλειστή μορφή

Η εντροπία των υπολοιπόμενων απωλειών είναι φραγμένη από επάνω από τη μέση υπολοιπόμενη απώλεια.

**Πρόταση 7.2.** Η εντροπία  $H(Y, d)$  των αποκομμένων από κάτω απωλειών (i) είναι αύξουσα συνάρτηση του  $d$  εάν η  $\lambda_X(x)$  είναι φθίνουσα ως προς  $x$  και (ii) είναι φθίνουσα ως προς  $d$  εάν η  $\lambda_X(x)$  είναι αύξουσα ως προς  $x$ .

Απόδειξη. Βλέπε Ebrahimi (1996). ■

Απόρροια των παραπάνω είναι ότι εάν η συνάρτηση κινδύνου των απωλειών αυξάνεται όταν αυξάνεται το  $x$  τότε η αποκομμένη από κάτω εντροπία μειώνεται.

Στον Πίνακα 7.3 παρουσιάζουμε το άνω φράγμα για κάθε κατανομή απώλειας. Το φράγμα για την κατανομή Gamma δίνεται συναρτήσει της exponential integral function  $E_n(z) = \int_1^\infty e^{-zt}/t^n dt$ . Παρατηρούμε ότι παρόλο που δεν μπορούμε να υπολογίσουμε αλγεβρικά την εντροπία για την αποκομμένη από κάτω κατανομή Burr, μπορούμε να βρούμε ένα κάτω όριο αυτής, στην οποία εμφανίζονται η υπεργεωμετρική συνάρτηση (hypergeometric function)  ${}_2F_1(a, b; c; z) = \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}$ , όπου  $(x)_n = x(x+1)\dots(x+n-1)$ . Για τις κατανομές Transformed Gamma, Loggamma, Lognormal και Generalized Pareto το άνω φράγμα δεν μπορεί να υπολογιστεί σε κλειστή μορφή. Πιο συγκεκριμένα, για τις κατανομές Transformed Gamma και Loggamma ένα ολοκλήρωμα της μορφής

$$\int_d^\infty \frac{\Gamma(a, x(d)) - 1}{\Gamma(a)[\Gamma(a, x(y)) - 1]} dy$$

πρέπει να υπολογιστεί.

## 7.4 Επίδραση της αποκοπής από επάνω

Σκοπός της ενότητας αυτής είναι να μελετήσουμε πώς επιδρά στην εντροπία η αποκοπή από επάνω (Sachlas and Papaioannou, 2009b). Πριν όμως από αυτό, ας ορίσουμε την έννοια της αποκοπής από επάνω.

Έστω ότι οι απώλειες  $X$  δεν καταγράφονται όταν είναι ίσες ή ξεπερνούν ένα συγκεκριμένο όριο φερεγγυότητας (liability limit)  $u$ . Στην περίπτωση αυτή τα δεδομένα  $S$  είναι αποκομμένα από επάνω ή, διαφορετικά, δεξιά αποκομμένα. Πιο συγκεκριμένα, η τυχαία μεταβλητή που περιγράφει τις αποκομμένες από επάνω απώλειες είναι η  $S = X$  εάν  $X \leq u$  ενώ η  $S$  δεν ορίζεται εάν ισχύει  $X > u$ . Ισοδύναμα έχουμε ότι  $S(u) = X|X \leq u$ . Για παράδειγμα, ένα ασφαλιστήριο συμβόλαιο αυτοκινήτου καλύπτει απώλειες μέχρι ενός ορίου  $u$ , ενώ μεγαλύτερες απώλειες καλύπτονται από τον κάτοχο του αυτοκινήτου. Εάν η ζημιά είναι  $X$  τότε η απώλεια για την ασφαλιστική εταιρεία είναι  $X|X \leq u$ . Σημειώνουμε ότι η αποκομμένη από επάνω απώλεια είναι διαφορετική από την λογοκριμένη από επάνω απώλεια, η οποία ορίζεται ως η τυχαία μεταβλητή  $S_{cens}(u) = \min\{X, u\}$ . Στην περίπτωση αυτή, εάν η απώλεια είναι  $X > u$ , η ασφαλιστική εταιρεία πληρώνει ένα ποσό  $u$ .

Η σχέση μεταξύ των εντροπιών της  $X$  και της  $S(u)$  δίνεται στο παρακάτω λήμμα.

**Λήμμα 7.3.** Η εντροπία της τυχαίας μεταβλητής  $S(u)$  των παρατηρούμενων αποκομμένων από επάνω δεδομένων, δίνεται από τη σχέση

$$H_{S(u)} = \frac{1}{F_X(u)} \left[ H_X + \int_u^\infty f_X(x) \ln f_X(x) dx + F_X(u) \ln F_X(u) \right],$$

όπου  $H_X$  είναι η εντροπία της  $X$ ,  $F_X(x)$  είναι η συνάρτηση κατανομής της  $X$  και  $u$  είναι το σημείο της αποκοπής (όριο φερεγγυότητας).

Απόδειξη. Είναι γνωστό ότι η συνάρτηση πυκνότητας πιθανότητας της  $S(u)$  δίνεται από τη σχέση

$$f_{S(u)}(x) = \frac{f_X(x)}{F_X(u)}.$$

Συνεπώς η εντροπία της μεταβλητής  $S(u)$  δίνεται από τη σχέση

$$\begin{aligned} H_{S(u)} &= - \int_0^u \frac{f_X(x)}{F_X(u)} \ln \frac{f_X(x)}{F_X(u)} dx \\ &= - \frac{1}{F_X(u)} \left[ \int_0^u f_X(x) \ln f_X(x) dx - \ln F_X(u) \int_0^u f_X(x) dx \right] \\ &= \frac{1}{F_X(u)} \left[ H_X + \int_u^\infty f_X(x) \ln f_X(x) dx + F_X(u) \ln F_X(u) \right], \end{aligned}$$

όπου  $H_X$  είναι η εντροπία της μεταβλητής  $X$ .

Ακολουθούμε ανάλογη διαδικασία για τη διακριτή περίπτωση. ■

Η παραπάνω έκφραση της εντροπίας υπό αποκοπή από επάνω είναι ίδια με αυτή που παρουσίασαν οι Di Crescenzo and Longobardi (2002), ως μέτρο της αβεβαιότητας για past lifetime distributions, και ονομάζεται past entropy. Αναλυτικές εκφράσεις δεν έχουν παρουσιαστεί στη βιβλιογραφία.

Οι σχετικές αναλυτικές εκφράσεις της εντροπίας του Shannon για τα πιο γνωστά μοντέλα απώλειας, παρουσιάζονται στον Πίνακα 7.4. Επιπλέον, για όλες τις παραπάνω κατανομές, η συνάρτηση κατανομής  $F_X(\cdot)$  είναι αναλυτικά γνωστή. Παρατηρούμε ότι η έκφραση για τα περισσότερα μοντέλα απώλειας είναι αρκετά περίπλοκη, καθώς εμπλέκονται συναρτήσεις όπως η συνάρτηση Meijer  $G$ , η συνάρτηση σφάλματος Erf και η συμπληρωματική συνάρτηση σφάλματος Erfc. Αυτό κάνει δύσκολη τη γενική και ευθεία ερμηνεία της επίδρασης στην εντροπία της αποκοπής από επάνω.

Η επόμενη πρόταση δείχνει τη σχέση μεταξύ των εντροπιών  $H_X$ ,  $H_{Y(d)}$  και  $H_{S(d)}$ .

**Πρόταση 7.3.** Για όλα τα  $d > 0$ , ισχύει

$$H_X = H [F_X(d), \bar{F}_X(d)] + \bar{F}_X(d)H_{Y(d)} + F_X(d)H_{S(d)}$$

όπου  $H [p, \bar{p}] = -p \ln p - \bar{p} \ln \bar{p}$  για  $\bar{p} = 1 - p$  είναι η εντροπία της κατανομής Bernoulli.

Απόδειξη. Από τα Λήμματα 7.2 και 7.3 έχουμε ότι

$$\begin{aligned} \bar{F}_X(d)H_{Y(d)} + F_X(d)H_{S(d)} &= H_X + \int_0^d f_X(x) \ln f_X(x) dx + \bar{F}_X(d) \ln \bar{F}_X(d) \\ &\quad + H_X + \int_d^\infty f_X(x) \ln f_X(x) dx + F_X(d) \ln F_X(d) \\ &= 2H_X + \int_0^\infty f_X(x) \ln f_X(x) dx - H [F_X(d), \bar{F}_X(d)] \\ &= 2H_X - H_X - H [F_X(d), \bar{F}_X(d)] \\ &= H_X - H [F_X(d), \bar{F}_X(d)]. \end{aligned}$$

Συνεπώς η τελευταία ισότητα μπορεί να γραφεί ως

$$H_X = H [F_X(d), \bar{F}_X(d)] + \bar{F}_X(d)H_{Y(d)} + F_X(d)H_{S(d)}.$$

■

Μια ερμηνεία της παραπάνω σχέσης είναι ότι η αβεβαιότητα για μια απώλεια διασπάται στην (i) αβεβαιότητα σχετικά με το εάν οι απώλειες έχουν ξεπεράσει ή όχι την τιμή  $d$ , (ii) αβεβαιότητα σχετικά με το μέγεθος της απώλειας που έχει ξεπεράσει την τιμή του αφαιρετέου

ποσού  $d$ , δηλαδή τη ζημιά που καλύπτεται από την ασφαλιστική εταιρεία και (iii) αβεβαιότητα σχετικά με το μέγεθος της απώλειας που καλύπτει η ασφαλιστική εταιρεία δοθέντος ότι η απώλεια είναι μικρότερη του ορίου φερεγγυότητας  $d$ .

**Σημείωση 7.1.** Η εντροπία υπό αποκοπή από επάνω μπορεί επίσης να υπολογιστεί μέσω της έκφρασης

$$H_{S(u)} = \ln F_X(u) - \frac{1}{F_X(u)} \int_0^u f_X(x) \ln f_X(x) dx.$$

Δυο άλλες ενδιαφέρουσες ποσότητες που σχετίζονται με τις κατανομές απώλειες και τα ασφαλιστικά μαθηματικά είναι η  $W_1(u) = u - S(j)$ , η οποία περιγράφει το «υπόλοιπο» της κάλυψης και η  $W_2(u) = X - u$ , η οποία περιγράφει το «υπόλοιπο» της απαίτησης. Επειδή η διαφορική εντροπία είναι αναλλοίωτη υπό μετατοπίσεις έχουμε ότι:

$$H_{W_1(u)} = H_{S(u)} \quad \text{και} \quad H_{W_2(u)} = H_X$$

## 7.5 Συμπεράσματα

Ο σκοπός του κεφαλαίου αυτού είναι διπλός. Αρχικά διερευνήσαμε τη χρήση της εντροπίας ως ένα μέτρο αβεβαιότητας για τις κατανομές απώλειες και γενικότερα στα αναλογιστικά μαθηματικά. Στη συνέχεια, μελετήσαμε την επίδραση που έχουν στον υπολογισμό της εντροπίας ο πληθωρισμός, η αποκοπή από κάτω και η αποκοπή από επάνω. Οι περιπτώσεις αυτές εμφανίζονται συχνά στην πράξη. Συνεπώς η γενική, αλλά και η ειδική για κάθε κατανομή απώλειας αναλυτική έκφραση δίνονται σε σχέση με την εντροπία των αρχικών δεδομένων. Αποδείξαμε ότι ο πληθωρισμός αυξάνει την εντροπία, το οποίο σημαίνει ότι υπό πληθωρισμό έχουμε περισσότερη αβεβαιότητα σχετικά με την αρχική κατάσταση. Εάν η αρχική εντροπία είναι αρνητική τότε η αποκοπή από επάνω μειώνει την εντροπία.

Η αποκοπή από κάτω είναι η πιο συχνά εμφανιζόμενη στην πράξη περίπτωση. Παρόλα αυτά, η επίδρασή της στην εντροπία δεν είναι εύκολα ερμηνεύσιμη καθώς οι εκφράσεις της εντροπίας είναι αρκετά περίπλοκες και διαφέρουν από μοντέλο σε μοντέλο. Σε ορισμένες περιπτώσεις δεν είναι δυνατόν να βρούμε αναλυτική έκφραση. Η φυσική ερμηνεία της επίδρασης της αποκοπής από επάνω δεν είναι επίσης εύκολη. Συναρτήσεις όπως η συνάρτηση Meijer  $G$ , η συνάρτηση σφάλματος Erf και η συμπληρωματική συνάρτηση σφάλματος Erfc εμφανίζονται στους τύπους υπολογισμού.

Η εντροπία, ως μέτρο αβεβαιότητας και πληροφορίας, είναι χρήσιμη για τη μελέτη και αξιολόγηση αναλογιστικών μοντέλων. Μια μέθοδος εκτίμησης και κατασκευής μοντέλων είναι η μέθοδος της μέγιστης εντροπίας. Σύμφωνα με αυτή, ξεκινώντας αρχικά από κάποιες ροπές, οι οποίες αποτελούν τη μόνη γνωστή πληροφορία για το μοντέλο, ως καταλληλότερο μοντέλο επιλέγεται αυτό με τη μεγαλύτερη εντροπία. Η μέθοδος αυτή, χρησιμοποιείται



Πίνακας 7.4: Εντροπία υπό αποκοπή από επάνω

Κατανομή	Εντροπία
Εκθετική	$\ln \frac{1-e^{-\lambda u}}{\lambda} + 1 - \frac{\lambda u e^{-\lambda u}}{1-e^{-\lambda u}}$
Pareto	$\ln \frac{1-\left(\frac{\lambda}{\lambda+u}\right)^\alpha}{\alpha \lambda^\alpha} + \frac{\alpha+1}{\alpha \left[1-\left(\frac{\lambda}{\lambda+u}\right)^\alpha\right]} \left[1 + \alpha \ln \lambda - \frac{\lambda^\alpha(1+\alpha \ln(\lambda+u))}{(\lambda+u)^\alpha}\right]$
Weibull	$\ln \frac{1-e^{-cu^\tau}}{c\tau} + \frac{1}{1-e^{-cu^\tau}} \left[1 + \frac{\tau-1}{\tau} (\gamma + \Gamma(0, cu^\tau) + \ln c + e^{-cu^\tau} \tau \ln u) - (1 - cu^\tau)e^{-cu^\tau}\right]$
Gamma	$\ln \frac{k(\alpha, \lambda u)}{\lambda^\alpha} + \frac{1}{k(\alpha, \lambda u)} \left[ \Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda u) \right. \\ \left. + (\alpha - 1) \left( \Gamma(\alpha, \lambda u) \ln u + G_{2,0}^{3,0} \left( \lambda u \mid \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix} \right) - \Gamma(\alpha) (\psi(\alpha) + \ln \lambda) \right) \right]$
Tr. Gamma	$\ln \frac{k(\alpha, (\lambda u)^\tau)}{\tau \lambda^{\tau \alpha}} + \frac{1}{k(\alpha, (\lambda u)^\tau)} \left\{ \Gamma(\alpha + 1) - \Gamma(\alpha + 1, (\lambda u)^\tau) \right. \\ \left. + \frac{\tau \alpha - 1}{\tau} \left[ \tau \ln u \Gamma(\alpha, (\lambda u)^\tau) + G_{2,0}^{3,0} \left( (\lambda u)^\tau \mid \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix} \right) + \Gamma(\alpha) (\tau \ln u - \psi(\alpha)) \right] \right\}$
Loggama	$\ln \frac{k(\alpha, \lambda \ln u)}{\lambda^\alpha} + \frac{1}{k(\alpha, \lambda \ln u)} \left[ (\alpha - 1) \left( \Gamma(\alpha) \Gamma(\lambda) + \Gamma(\alpha, \lambda \ln u) \ln(\ln u) + G_{2,0}^{3,0} \left( \lambda \ln u \mid \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix} \right) \right) \right. \\ \left. + \frac{\lambda+1}{\lambda} (\Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda \ln u)) + (\alpha - 1) \psi(\alpha) \Gamma(\alpha) \right]$
Lognormal	$\ln \Phi \left( \frac{\ln u - \mu}{\sigma} \right) + \ln (\sigma \sqrt{2\pi}) + \frac{\mu^2}{\sigma^2} + \frac{1}{2\sigma^2 \Phi \left( \frac{\ln u - \mu}{\sigma} \right)} \left[ \frac{\mu^2 + \sigma^2}{2} \operatorname{Erfc} \left( \frac{\mu - \ln u}{\sqrt{2}\sigma} \right) - \frac{\sigma(\mu + \ln u)}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln u)^2}{2\sigma^2}} \right] \\ - \frac{1}{\Phi \left( \frac{\ln u - \mu}{\sigma} \right)} \left( \frac{\mu}{\sigma^2} - 1 \right) \left[ -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln u)^2}{2\sigma^2}} + \frac{\mu}{2} - \frac{\mu}{2} \operatorname{Erf} \left( \frac{\mu - \ln u}{\sqrt{2}\sigma} \right) \right]$
Burr	όχι σε κλειστή μορφή
Gen. Pareto	όχι σε κλειστή μορφή

$$k(\alpha, x) = \Gamma(\alpha) \Gamma(\alpha, x) \quad \mathbf{a} = (1, 1) \quad \mathbf{b} = (0, 0, \alpha)$$

ευρύτατα σε διάφορους τομείς και επιστήμες όπως τα Οικονομικά, η Λογιστική, η Βιολογία, η Ιατρική, η Οικολογία, η Αναλογιστική Επιστήμη κλπ. (Karim, 1989). Στην Αναλογιστική Επιστήμη, η αρχή της μέγιστης εντροπίας μπορεί να χρησιμοποιηθεί για την εύρεση της κατανομής του αριθμού των απαιτήσεων μιας ασφαλιστικής εταιρείας σε ένα συγκεκριμένο χρονικό διάστημα, την εύρεση της κατανομής που περιγράφει ένα καταστροφικό γεγονός κλπ. Συνεπώς η εντροπία για τις περιπτώσεις μου μελετήθηκαν σε αυτή την παράγραφο είναι ενδιαφέρουσα και μπορεί να χρησιμοποιηθεί σε ένα πλαίσιο μέγιστης εντροπίας. Συνεπώς οι αναλυτικές εκφράσεις της εντροπίας στις περιπτώσεις που μελετήθηκαν σε αυτό το κεφάλαιο είναι ενδιαφέρουσες.

## ΚΕΦΑΛΑΙΟ 8

### ΣΥΜΠΕΡΑΣΜΑΤΑ ΔΙΑΤΡΙΒΗΣ

#### 8.1 Συμπεράσματα

Η ενασχόλησή μας με το θέμα που διαπραγματεύεται η διατριβή ξεκίνησε μελετώντας το άρθρο του Brockett (1991) με τίτλο «Information Theoretic Approach to Actuarial Science: A Unification and Extension of Relevant Theory and Applications». Πρόκειται για ένα εξαιρετικό άρθρο, στο οποίο ο συγγραφέας παρουσιάζει διάφορες χρήσεις της Στατιστικής Θεωρίας Πληροφοριών στην Αναλογιστική Επιστήμη. Πιο συγκεκριμένα επιλύει ορισμένα αναλογιστικά προβλήματα με τη χρήση του απόκλισης των Kullback - Leibler. Ένα θέμα με το οποίο ασχολήθηκε ήταν το πρόβλημα της εξομάλυνσης ποσοστών θνησιμότητας, στο οποίο όμως, όπως αναφέραμε στο Κεφάλαιο 3, τα ποσοστά θνησιμότητας δεν αθροίζουν στη μονάδα αλλά την ξεπερνούν. Αυτή η λεπτομέρεια μας έδωσε το ερέθισμα να μελετήσουμε τις ιδιότητες ορισμένων μέτρων απόκλισης όταν σε αυτά εμπλέκονται μη πιθανοτικά διανύσματα.

Η θεωρητική διερεύνηση του μέτρου απόκλισης των Kullback - Leibler  $D^{KL}(\mathbf{p}, \mathbf{q})$  με μη πιθανοτικά διανύσματα, στο Κεφάλαιο 3, έδειξε ότι το μέτρο αυτό ικανοποιεί ορισμένες από τις βασικές ιδιότητες των μέτρων απόκλισης και στατιστικής πληροφορίας, τις οποίες ικανοποιεί και το μέτρο απόκλισης των Kullback - Leibler  $I^{KL}(\mathbf{p}^*, \mathbf{q}^*)$ . Υπό ορισμένες συνθήκες, το μέτρο  $D^{KL}(\mathbf{p}, \mathbf{q})$  είναι μη αρνητικό, προσθετικό και αναλλοίωτο υπό επαρκείς περιορισμούς. Επιπλέον, το μέτρο αυτό ικανοποιεί την ιδιότητα της μέγιστης πληροφορίας. Συνεπώς, το μέτρο  $D^{KL}(\mathbf{p}, \mathbf{q})$  μπορεί να θεωρηθεί ως μέτρο απόκλισης και πληροφορίας.

Στη συνέχεια μελετήσαμε το μέτρο απόκλισης δύναμης των Cressie and Read με μη πιθανοτικά διανύσματα. Το μέτρο αυτό επίσης ικανοποιεί ορισμένες από τις ιδιότητες που το μέτρο με διανύσματα πιθανότητας ικανοποιεί. Πιο συγκεκριμένα, υπό ορισμένες συνθήκες είναι μη αρνητικό, προσθετικό, μεγαλύτερο από το μέτρο  $I^{CR}(\mathbf{p}^*, \mathbf{q}^*)$  και αναλλοίωτο υπό επαρκείς περιορισμούς. Επιπλέον, ικανοποιείται η ιδιότητα της μέγιστης πληροφορίας.

Στη συνέχεια ασχοληθήκαμε με μια επέκταση των μέτρων απόκλισης και ιδιαίτερα του μέτρου των Kullback - Leibler για μη αρνητικές συναρτήσεις ή διανύσματα που έγινε από τον

Csiszar (1991), ο οποίος προσέθεσε την ποσότητα  $\sum_i q_i - \sum_i p_i$  στην έκφραση του τυπικού μέτρου. Η θεωρητική διερεύνηση της εκτεταμένης απόκλισης των Kullback - Leibler έδειξε ότι αυτή ικανοποιεί την ελάχιστη απαίτηση για μια συνάρτηση να θεωρείται μέτρο απόκλισης. Εφαρμόσαμε την ίδια τροποποίηση και στο μέτρο απόκλισης δύναμης των Cressie - Read. Όμως αυτό το εκτεταμένο μέτρο ικανοποιεί την ελάχιστη απαίτηση μόνο για  $\lambda = 0$ ,  $\lambda = 1$  και  $\lambda = -1$ . Όταν ισχύει η ισότητα  $\sum_i p_i = \sum_i q_i$  τα δυο προαναφερθέντα εκτεταμένα μέτρα ταυτίζονται με τα μέτρα απόκλισης των Kullback - Leibler και Cressie - Read για μη πιθανοτικά διανύσματα, αντίστοιχα, όπως αυτά ορίστηκαν στους Ορισμούς 3.1 και 3.4, του Κεφαλαίου 3, αντίστοιχα.

Παρόμοια ήταν τα αποτελέσματα της διερεύνησης και για τη διαφορά του Jensen που πρότειναν οι Burbea and Rao (1982). Και αυτό το μέτρο ικανοποιεί τις παραπάνω βασικές ιδιότητες των μέτρων απόκλισης και στατιστικής πληροφορίας.

Το σημαντικότερο, κατά τη γνώμη μας, αποτέλεσμα της διερεύνησης και ανάλυσης που διεξάγαμε ήταν η αποκάλυψη ενός περιορισμού ή συνθήκης για την ισχύ των ιδιοτήτων που αναφέραμε παραπάνω. Συγκεκριμένα, για να χαρακτηριστεί ένα μέτρο απόκλισης μεταξύ μη πιθανοτικών διανυσμάτων ως μέτρο απόκλισης και στατιστικής πληροφορίας, θα πρέπει να ισχύει η ισότητα

$$\sum_i p_i = \sum_i q_i.$$

Η συνθήκη αυτή είναι μια γενική συνθήκη που περιλαμβάνει και την ισότητα  $\sum_i p_i^* = \sum_i q_i^* = 1$ , η οποία ισχύει στην περίπτωση των διανυσμάτων πιθανότητας.

Αφού μελετήσαμε τις ιδιότητες των μέτρων απόκλισης, στη συνέχεια, στο Κεφάλαιο 5, παρουσιάσαμε τρόπους ελαχιστοποίησης αυτών υπό διάφορους περιορισμούς. Στην περίπτωση που έχουμε γραμμικούς και τετραγωνικούς περιορισμούς ανισότητας, όπως συμβαίνει στο πρόβλημα της εξομάλυνσης, η επίλυση του προβλήματος ελαχιστοποίησης των μέτρων απόκλισης είναι δύσκολη, λόγω του μεγάλου όγκου των περιορισμών. Βέβαια στη σημερινή εποχή η τεχνολογία έχει εξελιχθεί σε σημαντικό βαθμό οπότε είναι εύκολο μέσω του ηλεκτρονικού υπολογιστή να επιλύσουμε σε σύντομο χρονικό διάστημα προβλήματα τέτοιου είδους. Παρόλα αυτά, από μαθηματικής άποψης κρίθηκε σκόπιμο να μελετηθούν τα μέτρα απόκλισης από την πλευρά της Λαγκρανζιανής δυϊκότητας. Στο τέλος του κεφαλαίου δώσαμε ιδιαίτερη βάση στην εξαγωγή ορισμένων δυϊκών αποτελεσμάτων για το πρόβλημα της ελαχιστοποίησης της απόκλισης των Cressie - Read τάξης  $r$  και της διαφοράς του Jensen υπό τετραγωνικούς και γραμμικούς περιορισμούς ανισότητας. Το σημαντικότερο εύρημα της μελέτης μας είναι, ότι πέραν της εύρεσης του δυϊκού προβλήματος, το δυϊκό πρόβλημα και στις δυο περιπτώσεις διατυπώνεται χωρίς, δηλαδή δεν έχει, περιορισμούς και συνεπώς είναι πολύ πιο εύκολο να εφαρμοσθεί.

Στη συνέχεια και στο Κεφάλαιο 6, διερευνήσαμε μέσω ενός αριθμητικού παραδείγματος με τρία σύνολα δεδομένων, τη χρήση της γενικής οικογένειας των μέτρων απόκλισης δύναμης

των Cressie - Read στο πρόβλημα της εξομάλυνσης ποσοστών θνησιμότητας καθώς και τη συμπεριφορά στο ίδιο πρόβλημα της διαφοράς του Jensen. Η αριθμητική διερεύνηση έδειξε ότι, η ελαχιστοποίηση των αποκλίσεων δύναμης για διάφορες τιμές της παραμέτρου  $\lambda$ , υπό τους περιορισμούς (i) - (v) και/ή (vi), έδωσε ισοδύναμα αποτελέσματα, από πλευράς ομαλότητας, με αυτά άλλων μεθόδων εξομάλυνσης, όπως είναι η ευρέως χρησιμοποιούμενη μέθοδος των Whittaker - Henderson.

Σίγουρα η τιμή της παραμέτρου  $\lambda$  επηρεάζει τα αποτελέσματα της εξομάλυνσης. Από την αριθμητική διερεύνηση του θέματος προέκυψε ότι δεν μπορούμε ευθέως να θεωρήσουμε κάποια τιμή της παραμέτρου  $\lambda$  ως την καλύτερη για εξομάλυνση. Παρόλα αυτά για τις εξομαλύνσεις υπό τους περιορισμούς (i) - (v) προέκυψαν τα εξής γενικά συμπεράσματα: Τιμές του  $\lambda < -1$  δίνουν μη ικανοποιητικά αποτελέσματα ως προς την καλή προσαρμογή και για το λόγο αυτό θα πρέπει να αποφεύγονται. Όσον αφορά την ομαλότητα  $S$ , τιμές του  $\lambda \in (-1, 0.5)$ , υπερεξομαλύνουν τα δεδομένα. Για  $\lambda \geq 0$  οι διάφορες αποκλίσεις δίνουν παρόμοια αποτελέσματα ως προς την ομαλότητα και την προσαρμογή, συνεπώς η τιμή  $\lambda = 2/3$  που πρότειναν οι Cressie and Read (1984) από πλευράς στατιστικής ισχύος, είναι μια καλή επιλογή για εξομάλυνση.

Η χρήση του περιορισμού (vi)  $\sum_{x=1}^n u_x = \sum_{x=1}^n v_x$  που προτείνουμε, θεωρείται απαραίτητη καθώς αυτή είναι η ελάχιστη απαίτηση για ένα μέτρο χωρίς διανύσματα πιθανότητας να είναι μέτρο απόκλισης. Πέραν αυτού και τα αριθμητικά αποτελέσματα υποστηρίζουν τη χρήση της επιπλέον συνθήκης και έδειξαν αξιοσημείωτη βελτίωση στην προσαρμογή των εξομαλυμένων τιμών. Σημαντικό εύρημα της ανάλυσης αποτελεί το γεγονός ότι, στην περίπτωση αυτή, η τιμή της παραμέτρου  $\lambda$  φαίνεται πλέον να μην επηρεάζει την ομαλότητα  $S$ .

Όσον αφορά τα αριθμητικά αποτελέσματα που προκύπτουν από την ελαχιστοποίηση της διαφοράς του Jensen, είναι συγκρίσιμα με τις εξομαλύνσεις που γίνονται μέσω της απόκλισης των Kullback - Leibler, των Cressie - Read και της μεθόδου των Whittaker - Henderson.

Τέλος στο Κεφάλαιο 7 διερευνήσαμε τη χρήση της εντροπίας του Shannon ως μέτρο της αβεβαιότητας της απώλειας, δηλαδή για κατανομές απώλειας (loss models or distributions) και γενικότερα στα αναλογιστικά μαθηματικά. Πιο συγκεκριμένα, μελετήσαμε την επίδραση που έχουν στον υπολογισμό της εντροπίας της απώλειας ο πληθωρισμός (inflation), η αποκοπή από επάνω (truncation from above)  $x|x \leq u$  (που αντιστοιχεί στις απώλειες με κάποιο όριο αξιοπιστίας (liability) και η αποκοπή από κάτω (truncation from below)  $x|x \geq d$  (που αντιστοιχεί στις απώλειες με αφαιρετέο ποσό (deductible)). Η γενική, αλλά και η ειδική για κάθε κατανομή απώλειας, αναλυτική έκφραση δίνονται σε σχέση με την εντροπία των αρχικών δεδομένων. Αποδεικνύεται ότι ο πληθωρισμός αυξάνει την εντροπία, το οποίο σημαίνει ότι υπό πληθωρισμό έχουμε περισσότερη αβεβαιότητα σχετικά με την αρχική κατάσταση.

Η αποκοπή από κάτω είναι η πιο συχνά εμφανιζόμενη στην πράξη περίπτωση. Δυστυχώς δεν μπορούμε να δώσουμε μια σαφή ερμηνεία για την επίδρασή της στην εντροπία. Αυτό οφείλεται στο γεγονός ότι οι εκφράσεις της εντροπίας είναι αρκετά περίπλοκες και διαφέρουν

από μοντέλο σε μοντέλο. Σε ορισμένες περιπτώσεις, μάλιστα, δεν είναι δυνατόν να βρούμε αναλυτική έκφραση. Για τον ίδιο λόγο δεν είναι εύκολο να δώσουμε τη φυσική ερμηνεία της επίδρασης της αποκοπής από επάνω. Πολύπλοκες συναρτήσεις όπως η συνάρτηση Meijer  $G$ , η συνάρτηση σφάλματος Erfi και η συμπληρωματική συνάρτηση σφάλματος Erfc εμφανίζονται στους τύπους υπολογισμού.

Οι απώλειες με όριο αξιοπιστίας ισοδυναμούν με την έννοια του διανυθέντος χρόνου ζωής (past lifetime) στη Θεωρία Αξιοπιστίας και η παρατήρηση αυτή μας οδήγησε σε ενδιαφέρουσες σχέσεις και ερμηνείες που αφορούν τις αβεβαιότητες των απωλειών με όριο αξιοπιστίας, με deductible, το υπόλοιπο κάλυψης και το υπόλοιπο της απαίτησης (claim).

Συνδέσαμε την απώλεια με αφαιρετέο ποσό με την έννοια του residual lifetime και τον κίνδυνο απώλειας με αφαιρετέο ποσό και αποδείξαμε ότι η αβεβαιότητα της απώλειας με αφαιρετέο ποσό αυξάνεται ή μειώνεται ανάλογα με το αν ο κίνδυνος μειώνεται ή αυξάνεται.

Στην επόμενη ενότητα θα παρουσιάσουμε κάποιες σκέψεις για μελλοντική έρευνα που προέκυψαν από την μέχρι τώρα ενασχόλησή μας με το θέμα και την έρευνά μας.

## 8.2 Μελλοντική Έρευνα

Από τη στιγμή που τα μέτρα πληροφoρίας και απόκλισης με μη πιθανοτικά διανύσματα μπορούν να θεωρηθούν ως κανονικά μέτρα πληροφoρίας και απόκλισης, κρίνεται σκόπιμο να μελετήσουμε και την ασυμπτωτική τους κατανομή. Για το λόγο αυτό κρίνεται σκόπιμο να εκτιμήσουμε και να μελετήσουμε την ασυμπτωτική συμπεριφορά των εκτιμώμενων μέτρων απόκλισης με μη πιθανοτικά διανύσματα. Στη συνέχεια τα αποτελέσματα αυτά θα αποτελέσουν τη βάση για την τροποποίηση των τυπικών ελέγχων καλής προσαρμογής, που παρουσιάστηκαν στην Ενότητα 4.4 και χρησιμοποιούνται για την αξιολόγηση των εξομαλύνσεων.

Όσον αφορά τη χρήση των μέτρων απόκλισης στο πρόβλημα της εξομαλύνσης ποσοστών θνησιμότητας, μια σημαντική επέκταση της μεθόδου μας θα ήταν η πρόβλεψη μελλοντικών τιμών. Δυστυχώς τα μοντέλα μας δεν έχουν προβλεπτικό χαρακτήρα. Μια σκέψη είναι να χρησιμοποιήσουμε ένα παραμετρικό μοντέλο που θα προσαρμόσουμε στις εξομαλυμένες τιμές και στη συνέχεια να χρησιμοποιήσουμε ένα time based training-test split. Συγκεκριμένα, μπορούμε να διαιρέσουμε τα σύνολα δεδομένων σε δυο ίσα διαστήματα, να εξομαλύνουμε το πρώτο διάστημα μέσω της ελαχιστοποίησης του μέτρου των Cressie - Read, να προσαρμόσουμε το παραμετρικό μοντέλο και στη συνέχεια να συγκρίνουμε τις προβλεπόμενες τιμές, που παράγονται από το μοντέλο, με τις αδρές τιμές του δεύτερου διαστήματος. Η προσέγγιση αυτή όμως έχει ένα σημαντικό μειονέκτημα. Τα αποτελέσματα εξαρτώνται από το επιλεγμένο παραμετρικό μοντέλο. Μια καλύτερη προσέγγιση είναι η ελαχιστοποίηση της απόκλισης των Cressie - Read  $D^{CR}(f(\mathbf{x}), \mathbf{u})$  ως προς συναρτήσεις  $f$ , όπου το διάνυσμα  $\mathbf{v}$

έχει αντικατασταθεί από μια άγνωστη συνάρτηση  $f(\mathbf{x})$  των ηλικιών  $\mathbf{x}$ , υπό περιορισμούς ανάλογους με τους περιορισμούς (i) - (vi). Αυτό είναι ένα πρόβλημα λογισμού μεταβολών και έχουμε την υπόνοια ότι η λύση του θα είναι μια συνάρτηση spline. Το θέμα αυτό σίγουρα θα αποτελέσει αντικείμενο μελλοντικής έρευνας.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΡΑΙΑ

# РАНЕЕЗНАМО ПЕРПАА



## ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Παπαϊωάννου, Τ. (2000). *Θεωρία Πιθανοτήτων και Στατιστικής*, Εκδόσεις Αθ. Σταμούλης, Αθήνα.
- [2] Παπαϊωάννου, Τ. και Φερεντίνος, Κ. (2000). *Μαθηματική Στατιστική*, 2η Έκδοση, Εκδόσεις Αθ. Σταμούλης, Αθήνα.
- [3] Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series, No. 55.
- [4] Agresti, A. (2002). *Categorical Data Analysis*, Second Edition, Wiley - Interscience, New Jersey.
- [5] Ahmed, N.A. and Gokhale, D.V. (1989). Entropy expressions and their estimators for multivariate distributions, *IEEE Transactions on Information Theory*, IT-35, 688-692.
- [6] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle, In B.N. Petrov and F. Csaki, (Eds.) *2nd International Symposium on Information Theory*, Budapest: Akademiai Kiado, 267 - 281.
- [7] Ali, S.M. and Silvey, S.D. (1966). A general class of coefficients of divergence of one distribution from another, *Journal of the Royal Statistical Society Series B*, 28, 131 - 142.
- [8] Alwan, L.C., Ebrahimi, N. and Soofi, E.S. (1998). Information theoretic framework for process control, *European Journal of Operations Research*, 111, 526 - 542.
- [9] Arndt, C. (2001). *Information Measures - Information and its Description in Science and Engineering*, Springer - Verlag, Berlin.
- [10] Basu, A., Harris, I.R., Hjort, N.L., Jones, M.C. (1998). Robust and Efficient Estimation by Minimising a Density Power Divergence, *Biometrika*, 85, 3, 549 - 559.

- [11] Bhattacharyya, A. (1943). On a Measure of Divergence Between Two Statistical Populations Defined by their Probability Distributions, *Bull. Calcutta Math. Soc.*, 35, 99 - 110.
- [12] Belzunce, F., Navarro, J., Ruiz, J. and Aguila, Y. (2004). Some results on residual entropy function, *Metrika*, 59, 147-161.
- [13] Benish, W.A. (1999). Relative Entropy as a Measure of Diagnostic Information, *Medical Decision Making*, 19, 2, doi: 10.1177/0272989X9901900211.
- [14] Benjamin, P., Pollard, J.H. (1980). *The Analysis of Mortality and Other Actuarial Statistics*, Heinemann, London.
- [15] Ben-Tal, A. and Charnes, A. (1979). A dual optimization framework for some problems in information theory and statistics, *Problems of Control and Information Theory*, 8, 387-401.
- [16] Berliner and Lev (1978). A dual optimization framework for some problems in information theory and statistics, *Problems of Control and Information Theory*, 8, 387-401.
- [17] Boekee, D. (1979). The  $D_f$ -information of order  $s$ , in *Transactions of the 8th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Prague, 55 - 66.
- [18] Boyd, S. and Vandenberghe, L. (2006). *Convex Optimization*, Cambridge University Press, Cambridge, United Kingdom.
- [19] Brazauskas, V. (2003). Information Matrix for Pareto (IV), Burr and Related Distributions, *Communications in Statistics*, 32 (2), xx-xx.
- [20] Brockett, P.L. (1991). Information Theoretic Approach to Actuarial Science: A Unification and Extension of Relevant Theory and Applications, *Transactions of the Society of Actuaries*, 43, 73 - 114.
- [21] Brockett, P.L. and Cox, S. (1984). Statistical Adjustment of Mortality Tables to Reflect Known Information, *TSA*, 36, 63 - 71.
- [22] Brockett, P.L. and Zhang, J. (1986). Information Theoretical Mortality Graduation, *Scandinavian Actuarial Journal*, vol. 3 - 4, 131 - 140.
- [23] Brockett, P.L., Charnes, A. and Cooper, W.W. (1980). MDI Estimation via Unconstrained Convex Programming, *Communications in Statistics B*, 9, no. 3, 223 - 234.

- [24] Brockett, P.L., Cox, S., Golang, B., Phillips, F. and Song, Y. (1995). Actuarial Usage of Grouped Data: An Approach to Incorporating Secondary Data, *TSA*, 47, 89 • 113.
- [25] Broffitt, J.D. (1996). On Smoothness in Multidimensional Whittaker Graduation, *Insurance: Mathematics and Economics*, 18, 13 - 27.
- [26] Brunel, V. (2004). Minimal models for credit risk: an information theory approach, *working paper*
- [27] Burbea, J., Rao, C.R. (1982a). On the Convexity of Some Divergence Measures Based on Entropy Functions, *IEEE Transactions on Information Theory*, 28, 489 - 495.
- [28] Burbea, J., Rao, C.R. (1982b). Entropy Differential Metric, Distance and Divergence Measures in Probability Spaces: A Unified Approach, *Journal of Multivariate Analysis*, 12, 575 - 596.
- [29] Burbea, J., Rao, C.R. (1982c). On the convexity of higher order Jensen difference based on entropy functions, *IEEE Transactions on Information Theory*, 28, 961 - 963.
- [30] Burbea, J., Rao, C.R. (1984). Differential metrics in probability spaces, *Probability and Mathematical Statistics*, 3, 241 - 258
- [31] Casella, G. and Berger, R.L. (2001). *Statistical Inference*, 2nd edn, Duxbury, Pacific Grove, CA.
- [32] Cover, T.M. and Thomas, J.A. (1991). *Elements of Information Theory*, Wiley, New York.
- [33] Cressie, N.A.C., Read, T.R.C. (1984). Multinomial Goodness-of-Fit Tests, *Journal of Royal Statistical Society, B*, Vol. 46, No. 3, 440 - 464.
- [34] Csiszar, I. (1963). Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten, *Publications of the Mathematical Institute of the Hungarian Academy of Science*, 8, 85 - 108.
- [35] Csiszar, I. (1991). Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems, *Annals of Statistics*, 19, 4, 2032 - 2066.
- [36] Cutler, A., Cordero - Brana, O.I. (1996). Minimum Hellinger distance estimation for finite mixture models, *Journal of American Statistical Association*, 91, pp. 1716 - 1723.

- [37] Darbellay, G.A. and Vajda, I. (2000). Entropy expressions for multivariate continuous distributions, *IEEE Transactions on Information Theory*, IT-46, 709-712.
- [38] Darooneh, A.H. (2004). Non-life insurance pricing: multiagent model, *Insurance: Mathematics and Economics* 24, 23 - 29.
- [39] Debon, A., Montes, F., Sala, R. (2005). A Comparison of Parametric Models for Mortality Graduation. Application to Mortality Data for the Valencia Region (Spain), *SORT*, 29, (2), 269 - 288.
- [40] Debon, A., Montes, F., Sala, R. (2006). A Comparison of Nonparametric Methods in the Graduation of Mortality: Application to Data from the Valencia Region (Spain), *International Statistical Review*, 74, no. 2, 215 - 233.
- [41] Di Crescenzo, A. and Longobardi, M. (2002). Entropy-based measure of uncertainty in past lifetime distributions, *Journal of Applied Probability*, 39, 434-440.
- [42] Ebrahimi, N. (1996). How to measure uncertainty in the residual life distributions, *Sankhya A*, 58, 48-57.
- [43] Ebrahimi, N., Kirmani, S.N.U.A., Soofi, E.C. (2007). Multivariate dynamic information, *Journal of Multivariate Analysis*, 98, 328 - 349.
- [44] Ferentinos, K., Papaioannou, T. (1981). New parametric measures of information, *Information and Control*, Vol. 51, 193 - 208.
- [45] Fisher, R.A. (1921). On the probable error of a coefficient of correlation deduced from a small sample, *Metron* 1, 3 - 32.
- [46] Fisher, R.A (1925). Theory of statistical estimation, *Cambridge Philosophical Society Proceedings*, 22, 7 - 25.
- [47] Gabor, D. (1946). Theory of communication, *J. Inst. Electr. Eng.*, 93, 429 - 457.
- [48] Gini, C. (1912). *Variabilita e mutabilita*. Bologna
- [49] Greville, T.N.E. (1983). Graduation, in *Encyclopedia of Statistical Sciences*, (Eds., S. Kotz and N.L. Johnson), Wiley, New York, 3, 463 - 469.
- [50] Guerreo-Cusumano, J. (1996). An asymptotic test of independence for multivariate t and Cauchy random variables with applications, *Information Sciences*, 92, 33-45.
- [51] Haberman, S. (1983). Decrement tables and the measurement of morbidity I, *Journal of the Institute of Actuaries*, 110, 361 - 381.

- [52] Haberman, S. (1984). Decrement tables and the measurement of morbidity II, *Journal of the Institute of Actuaries*, 111, 73 - 86.
- [53] Haberman, S. (1998). Actuarial Methods, *Encyclopedia of Biostatistics*, 1, (Eds., P. Armitage and Th. Colton), 37 - 49, John Wiley & Sons, New York.
- [54] Haberman, S., Renshaw, A.E. (1999). A Simple Graphical Method for the Comparison of two Mortality Experiences, *Applied Stochastic Models in Business and Industry*, 15, 333 - 352.
- [55] Harris, B. (1982). Entropy, in: S. Kotz, N.L. Johnson (Eds.), *Encyclopedia of Statistical Sciences*, vol. 2, Wiley, New York, pp. 512 - 516.
- [56] Hartley, R.V.L. (1928). Transmission of information, *Bell System Technical Journal*, 7, 535 - 563.
- [57] Hatzopoulos, P. (1997). *Statistical and Mathematical Modeling for Mortality Trends and the Comparison of Mortality Experiences through Generalised Linear Models and GLIM*, PhD Thesis, The City University, London.
- [58] Havrda, J. and Charvat, F. (1967). Quantification method of classification processes: concept of structural  $\alpha$ -entropy, *Kybernetika*, 3, 30 - 35.
- [59] Hogg, R.V. and Klugman, S.A. (1984). *Loss Distributions*, Wiley, New York.
- [60] Jeffreys H. (1946). An invariant form for the prior probability in estimation problems, *Proceedings of the Royal Society of London Ser. A*, Vol. 186, 453 - 461.
- [61] Joe, H. (1989). Estimation of entropy and other functionals of a multivariate density, *Annals of the Institute of Statistical Mathematics*, 41, 683 - 697.
- [62] Jhonson, R.C. and Jhonson, N.L. (1980). *Survival models and data analysis*, Wiley, New York.
- [63] Jones, L.K. and Byrne, C.L. (1990). General entropy criteria for inverse problems, with applications to data compression, pattern classification and cluster analysis, *IEEE Transactions on Information Theory*, 36, 23 - 30.
- [64] Kagan, M. (1963). On the theory of Fisher's amount of information, *Soviet Mathematics Doklady*, Vol. 4, 991 - 993.
- [65] Kapur, J.N. (1989). *Maximum-Entropy Models in Science and Engineering*, Wiley, New York.

- [66] Karagrigoriou, A. and Papaioannou, T. (2008). On measures of information and divergence and model selection criteria, In *Statistical Models and Methods for Biomedical and Technical Systems*, (Eds., F. Vonta, M. Nikulin, N. Limnios and C. Huber-Carol), 503 - 518, Birkhauser, Boston.
- [67] Kateri, M. and Papaioannou, T. (1994). f-divergence association models. *Int. J. Math. Stat. Sci.*, 3, 179 - 203.
- [68] Kateri, M. and Papaioannou, T. (1997). Asymmetry models for contingency tables, *J. Amer. Statist. Assoc.*, 92, 1124 - 1131.
- [69] Kullback, S. (1954). Certain Inequalities in Information Theory and the Cramer-Rao Inequality, *The Annals of Mathematical Statistics*, 25, 745 - 751.
- [70] Kullback, S. (1959). *Information Theory and Statistics*, John Wiley & Sons, New York.
- [71] Kullback, S. and Leibler, R.A. (1951). On Information and Sufficiency, *Annals of Mathematical Statistics*, 22, 79 - 86.
- [72] Landsman, L. and Makov, U. (1998). Exponential dispersion models and credibility. *Scandinavian Actuarial Journal*, 1, 89 - 96.
- [73] Landsman, L. and Makov, U. (1999). Credibility evaluation for the exponential dispersion family, *Insurance: Mathematics and Economics* 24, 23 - 29.
- [74] Lazo, A.C.G. and Rathie, P.N. (1978). On the entropy of continuous probability distributions, *IEEE Transactions on Information Theory*, IT-24, 120 - 122.
- [75] Laurent, P.J. (1972). *Approximation et Optimisation*, Hermann, Paris.
- [76] Lee, W.C. (1999). Selecting diagnostic tests for ruling out or ruling in disease: the use of the Kullback - Leibler distance, *International Journal of Epidemiology*, 28, 521 - 525.
- [77] Liese F., Vajda, I. (1987). *Convex Statistical Distances*, B. G. Teubner, Leipzig.
- [78] Lin, J. (1991). Divergence measures based on the Shannon entropy, *IEEE Trans. Information Theory*, Vol. 37, No. 1, 145 - 151.
- [79] Lindley D.V. (1956). On a measure of the information provided by an experiment, *The Annals of Mathematical Statistics*, 27, 986 - 1005.
- [80] London, D. (1985). *Graduation: The Revision of Estimates*, ACTEX Publications, Winsted, Connecticut.

- [81] Luthi, H.J. and Doege, J. (2005). Convex risk measures for portfolio optimization and concepts of flexibility, *Mathematical Programming Series, B*, 104, 541 - 559.
- [82] Mallows, C. (1959). The information in an experiment, *J. Roy. Statis. & Soc., Ser. B*, 21, 67 - 72.
- [83] Martin, S. and Durrani, T.S. (2007). A new divergence measure for medical image registration, *IEEE Transactions on Image Processing*, vol. 16, 4, 957 - 966.
- [84] Mathai, A.M., Rathie, P.N. (1975). *Basic Concepts in Information Theory and Statistics*, Wiley, New Delhi.
- [85] Mattheou, K. (2007). *On New Developments in Statistical Inference for Measures of Divergence*, Ph.D Thesis, University of Cyprus, Nicosia, Cyprus.
- [86] Mattheou, K., Lee, S. and Karagrigoriou, A. (2008). A model selection criterion based on the BHHJ measure of divergence, *Journal of Statistical Planning and Inference*, doi:10.1016/j.jspi.2008.04.022.
- [87] Matusita, K. (1951). On the theory of statistical decision functions, *Annals of the Institute of the Statistical Mathematics*, III, 17 - 35.
- [88] McCulloch, R.E. (1988). Local model influence, *Journal of the American Statistical Association*, 84, 406, 473 - 478.
- [89] Mac Lane, S. (1986). *Mathematics Form and Function*, Springer - Verlag, New York.
- [90] Menendez, M.L., Morales, D., Pardo, L. and Salicru, M. (1995). Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: a unified study, *Statistical Papers*, 36, 1 - 29.
- [91] Menendez, M.L., Pardo, J.A., Pardo, L. and Pardo, M.C. (1997). The Jensen-Shannon Divergence, *J. Franklin Inst.*, Vol. 334B, No. 2, 307 - 318.
- [92] Micheas, A. and Zografos, K. (2006). Measuring stochastic dependence using  $\phi$ -divergence, *Journal of Multivariate Analysis*, 97, 765 - 784.
- [93] Miller, M.D. (1949). *Elements of Graduation*, Actuarial Society of America, New York.
- [94] Moore, P.G. (1980). Are objectivism and subjectivism compatible concepts?, *JIA*, 107, 11 - 47.

- [95] Nadarajah, S. and Zografos, K. (2003). Formulas for Rényi information and related measures for univariate distributions, *Information Sciences: an International Journal*, 155, n.1-2, p.119-138.
- [96] Nadarajah, S. and Zografos, K. (2005). Expressions for Rényi and Shannon entropies for bivariate distributions, *Information Sciences*, 170, 173-189.
- [97] Neves, C.d.R., Migon, H.S. (2007). Bayesian Graduation of Mortality Rates: an application to mathematical reserve evaluation, *Insurance: Mathematics and Economics*, 40, 424 - 434.
- [98] Nielsen, J.P. (2003). Smoothing and Prediction with a View to Actuarial Science, Biostatistics and Finance, *Scandinavian Actuarial Journal*, 1, 51 - 74.
- [99] Nyquist, H. (1924). Certain factors affecting telegraph speed, *Bell System Technical Journal*, 3, 324 - 346.
- [100] Papaioannou (1970). On Statistical Information Theory and Related Measures of Information, *PhD Thesis*, Iowa State University AMES, IOWA.
- [101] Papaioannou, T. (1985). Measures of Information, *Encyclopedia of Statistical Sciences*, 5, (Eds., S. Kotz and N.L. Johnson), 391 - 497, John Wiley & Sons, New York.
- [102] Papaioannou, T. (2001). On Distances and Measures of Information: A Case of Diversity, In *Probability and Statistical Models with Applications*, (Eds., C.A Charalambides, M.V. Koutras and N. Balakrishnan), 503 - 515, Chapman & Hall/CRC.
- [103] Papaioannou, T. and Ferentinos, K. (2002). Is Fisher's Information Number a Measure of Statistical Information?, *Technical Report*, Department of Statistics & Insurance Science, University of Piraeus.
- [104] Papaioannou, T., Ferentinos, K. (2005). On Two Forms of Fisher's Information Number, *Communications in Statistics - Theory and Methods*, 34, 1461 - 1470.
- [105] Papaioannou, T. and Kempthorne, O. (1971). On Statistical Information Theory and Related Measures of Information, *Technical Report No ARL. 71-0059*, Aerospace Research Laboratories, Wright - Patterson A.F.B., Ohio.
- [106] Papaioannou, T., Tsairidis (2001). A Note on Defining Information in random Censoring, *Technical Report*, Department of Mathematics, University of Ioannina.
- [107] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures*, Chapman & Hall/CRC.



- [108] Pardo, L., Morales, D., Salicru, M. and Menendez, M.L. (1993). The  $\phi$ -divergence statistics in bivariate multinomial populations including stratification, *Metrika*, 40, 223 - 235.
- [109] Piantadosi, S. (2005). Translational clinical trials: an entropy-based approach to sample size, *Clinical Trials*, 2, 182 - 192.
- [110] Pereyra, M.E., Lamberti, P.W. and Rosso, O.A. (2007). Wavelet Jensen-Shannon divergence as a tool for studying the dynamics of frequency band components in EEG epileptic seizures, *Physica A*, 379, 122 - 132.
- [111] Promislow, S.D. and Young, V.R. (2000). Equity and Exact Credibility, *ASTIN Bulletin*, 30(1), 3 - 11.
- [112] Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, John Wiley, New York.
- [113] Read, T.R.C., Cressie, N.A.C. (1988). *Goodness - of - Fit Statistics for Discrete Multivariate Data*, Springer - Verlag, New York.
- [114] Reesor, R.M. and McLeish, D.L. (2002). Risk, Entropy, and the Transformation of Distributions, *Bank of Canada Working Paper*.
- [115] Renshaw, A.E., Haberman, S. and Hatzopoulos, P. (1996). On the duality of assumptions underpinning the construction of life tables, *ASTIN BULLETIN*, Vol 27, No I, 5 - 22.
- [116] Renyi, A. (1961). On measures of entropy and information, *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 547 - 561.
- [117] Rockafellar, J.H. (1970). *Convex Analysis*, Princeton University Press, Princeton, NJ.
- [118] Rohatgi, Y.K. (2001). *An Introduction to Probability and Statistics*, John Wiley & Sons, New York.
- [119] Sachlas, A. (2004). *Graduation of Mortality Tables*, MSc Thesis, Department of Statistics and Insurance Science, University of Piraeus, Piraeus.
- [120] Sachlas, A. and Papaioannou, T. (2008). On a minimization problem involving divergences and its applications, *Advances in Data Analysis: Theory and Applications to Reliability and Inference, Data Mining, Bioinformatics, Lifetime Data and Neural Networks*, 83 - 97, (C. H. Skiadas, Ed.), Birkhauser, Boston.

- [121] Sachlas, A. and Papaioannou, T. (2009a). Divergences without probability vectors and their applications, *Applied Stochastic Models in Business and Industry*, doi: 10.1002/asmb.803.
- [122] Sachlas, A. and Papaioannou, T. (2009b). On the entropy of loss distributions under inflation and truncation, *Submitted paper*.
- [123] Salicru M., Menendez, M.L., Morales, D. and Pardo, L. (1993). Asymptotic distribution of  $(h, \phi)$ -entropies, *Communications in Statistics - Theoretical and Methodological*, 22, 2015 - 2031.
- [124] Shannon, C.E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, 27, 379 - 432.
- [125] Sharma, B.D. and Mital, D.P. (1977). New non-additive measures of relative information, *Journ. Comb. Inf. Sys. Sci.*, 2, (4), 122 - 132.
- [126] Simpson, E.H. (1949). Measurement of diversity, *Nature*, 163, 688.
- [127] Soofi, E.S. (1994). Capturing the Intangible Concept of Information, *Journal of the American Statistical Association*, Vol. 89, No. 428, 1244 - 1254.
- [128] Taneja, I.J. (1987). Statistical aspects of divergence measures, *Journal of Statistical Planning & Inference*, Vol. 16, 135 - 145.
- [129] Teboulle, M. (1989). A Simple Duality Proof for Quadratically Constrained Entropy Functionals and Extension to Convex Constraints, *SIAM J. Appl. Math.*, Vol. 49, No. 6, 1845 - 1850.
- [130] Theil, H. (1970). On the Estimation of Relationships Involving Qualitative Variables, *American Journal of Sociology*, 76, 103 - 154.
- [131] The Actuarial Society of Hong Kong (2001). *Report on Hong Kong Assured Lives Mortality*. <http://www.actuaries.org.hk>.
- [132] Tsairidis, Ch., Ferentinos, K. and Papaioannou, T. (1996). Information and random censoring, *Information Sciences*, 92, 159 - 174.
- [133] Tsairidis, Ch., Zografos, K., Ferentinos, K. and Papaioannou, T. (2001). Information in Quantal Response Data and Random Censoring, *Annals of the Institute of Statistical Mathematics*, 53(3), 528 - 542.
- [134] Vajda, I. (1973).  $\chi^2$ -divergence and generalized Fisher's information, in *Transactions of the 6th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Prague, 873 - 886.

- [135] Vajda, I. (1987). *Convex Statistical Distances*, B. G. Teubner, Leipzig.
- [136] Vajda, I. (1989). *Theory of Statistical Inference and Information*, Kluwer Academic Publishers, Dordrecht.
- [137] Vos, P.D. (1992). Minimum  $f$ -divergence estimators and quasi-likelihood functions, *Annals of the Institute of Statistical Mathematics*, Vol. 44, pp. 261 - 279.
- [138] Wang, J.L. (1998). Smoothing Hazard Rates, *Encyclopedia of Biostatistics*, 5, (Eds., P. Armitage and Th. Colton), 4140 - 4150, John Wiley & Sons, New York.
- [139] Weber, E.J. (1976). *Mathematical Analysis. Business and Economic Applications*, 4th Edition, Harper and Row Publishers, New York.
- [140] Wiener, N. (1948). *Cybernetics*, Wiley, New York.
- [141] Xu, L., Bricker, D.L. and Kortanek, K.O. (1987). Bounds for stop-loss premium under restrictions on I-divergence, *Insurance: Mathematics and Economics*, 23, 119 - 139.
- [142] Yafune, A. and Ishiguro, M. (2001). Modification of pharmacokinetic sampling schedules in clinical phase I trials: a use of Kullback - Leibler information, *Statistics in Medicine*, 20, 1033 - 1049.
- [143] Zhang, J. and Brockett, P.L. (1987). Quadratically Constrained Information Theoretic Analysis, *SIAM Journal on Applied Mathematics*, 47, no 4, 871 - 885.
- [144] Zografos, K. (1999). On Maximum Entropy Characterization of Pearson's Type II and VII Multivariate Distributions, *Journal of Multivariate Analysis*, 71, 67 - 75.
- [145] Zografos, K. (2008a). On Mardia's and Song's measures of kurtosis in elliptical distributions, *Journal of Multivariate Analysis*, 99, 858 - 879.
- [146] Zografos, K. (2008b). On some entropy and divergence type measures of variability and dependence for mixed continuous and discrete variables, *Journal of Statistical Planning and Inference*, doi: 10.1016/j.jspi.2008.02.011.
- [147] Zografos, K. and Nadarajah, S. (2005a). Expressions for Renyi and Shannon entropies for multivariate distributions, *Statistics & Probability Letters*, 71, 71 - 84.
- [148] Zografos, K. and Nadarajah, S. (2005b). Survival exponential entropies, *IEEE Transactions on Information Theory*, vol. 51, no. 3, 1239 - 1246.

- [149] Zografos, K., Ferentinos, K., Papaioannou, T. (1989). Limiting Properties of Some Measures of Information, *Annals of the Institute of Statistical Mathematics*, B, Vol. 41, No. 3, 451 - 460.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΠΑ

## ΠΑΡΑΡΤΗΜΑ Α

# ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΥΙΚΗΣ ΘΕΩΡΙΑΣ ΓΙΑ ΤΗΝ ΕΛΑΧΙΣΤΟΠΟΙΗΣΗ ΤΗΣ ΑΠΟΚΛΙΣΗΣ ΤΩΝ CRESSIE AND READ

Η Λαγκρανζιανή συνάρτηση του προβλήματος ( $P$ ) είναι η

$$\begin{aligned} L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) &= \frac{1}{r(r+1)} \sum_{j=1}^n x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right] + \sum_{i=1}^m \lambda_i \left( \frac{1}{2} \mathbf{u}_i^T \mathbf{u}_i + \mathbf{b}_i^T \mathbf{x} + c_i \right) \mathbf{x} \\ &\quad + \sum_{i=1}^m \mathbf{y}_i^T (\mathbf{A}_i \mathbf{x} - \mathbf{u}_i) \\ &= \frac{1}{r(r+1)} \sum_{j=1}^n x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right] + \sum_{i=1}^m \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i + \sum_{i=1}^m \lambda_i \mathbf{b}_i^T \mathbf{x} \\ &\quad + \sum_{i=1}^m \lambda_i c_i + \sum_{i=1}^m \mathbf{y}_i^T \mathbf{A}_i \mathbf{x} - \sum_{i=1}^m \mathbf{y}_i^T \mathbf{u}_i \\ &= \frac{1}{r(r+1)} \sum_{j=1}^n x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right] + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i) \mathbf{x} \\ &\quad + \boldsymbol{\lambda}^T \mathbf{c} + \sum_{i=1}^m \left( \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right), \end{aligned}$$

όπου τα  $\mathbf{y}_i \in R^{n_i}$ ,  $i = 1, 2, \dots, m$  είναι διανυσματικοί λαγκρανζιανοί πολλαπλασιαστές.

Ας συμβολίζουμε το πρώτο μέρος του δεξιού μέλους της παραπάνω ισότητας με

$$f(\mathbf{x}; \boldsymbol{\lambda}, \mathbf{y}) = \frac{1}{r(r+1)} \sum_{j=1}^n x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right] + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i) \mathbf{x}. \quad (\text{A.1})$$

Έχουμε ότι

$$\begin{aligned}
 \frac{\partial}{\partial x_j} f(\mathbf{x}; \boldsymbol{\lambda}, \mathbf{y}) &= \frac{1}{r(r+1)} \left( (x_j)' \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right] + x_j \left[ \left( \frac{x_j}{d_j} \right)^r - 1 \right]' \right) \\
 &\quad + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \\
 &= \frac{1}{r(r+1)} \left( \left( \frac{x_j}{d_j} \right)^r - 1 + r \left( \frac{x_j}{d_j} \right)^r \right) + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \\
 &= \frac{1}{r} \left( \frac{x_j}{d_j} \right)^r - \frac{1}{r(r+1)} + \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j,
 \end{aligned}$$

όπου ο τόνος ( $\prime$ ) συμβολίζει την παράγωγο. Για να βρούμε το βέλτιστο σημείο, θέτουμε την παραπάνω ισότητα ίση με μηδέν, οπότε έχουμε

$$\begin{aligned}
 \frac{\partial}{\partial x_j} f(\mathbf{x}; \boldsymbol{\lambda}, \mathbf{y}) &= 0 \Leftrightarrow \\
 \frac{1}{r} \left( \frac{x_j}{d_j} \right)^r &= \frac{1}{r(r+1)} - \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \Leftrightarrow \\
 \frac{x_j}{d_j} &= \left( \frac{1}{r+1} - r \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right)^{\frac{1}{r}}
 \end{aligned}$$

και τελικά παίρνουμε

$$x_j = d_j \left( \frac{1}{r+1} - r \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j \right)^{\frac{1}{r}}. \quad (\text{A.2})$$

Αντικαθιστώντας την Εξίσωση A.2 στην Εξίσωση A.1, και θέτοντας  $s_j = \sum_{i=1}^m (\lambda_i \mathbf{b}_i^T + \mathbf{y}_i^T \mathbf{A}_i)_j$ ,

$j = 1, 2, \dots, n$  έχουμε ότι

$$\begin{aligned}
& \frac{1}{r(r+1)} \sum_{j=1}^n \left\{ d_j \left( \frac{1}{r+1} - rs_j \right)^{\frac{1}{r}} \left[ \left( \frac{d_j \left( \frac{1}{r(r+1)} - rs_j \right)^{\frac{1}{r}}}{d_j} \right)^r - 1 \right] \right. \\
& \quad \left. + r(r+1)s_j d_j \left( \frac{1}{r(r+1)} - rs_j \right)^{\frac{1}{r}} \right\} \\
&= \frac{1}{r(r+1)} \sum_{j=1}^n \left\{ d_j \left( \frac{1}{r+1} - rs_j \right)^{\frac{1}{r}} \left( \frac{1}{r+1} - rs_j - 1 + r(r+1)s_j \right) \right\} \\
&= \frac{1}{r(r+1)} \sum_{j=1}^n d_j \left( \frac{1}{r+1} - rs_j \right)^{\frac{1}{r}} \frac{r(r(r+1)s_j - 1)}{r+1} \\
&= \frac{1}{(r+1)^2} \sum_{j=1}^n d_j \left( \frac{1}{r+1} - rs_j \right)^{\frac{1}{r}} (r(r+1)s_j - 1),
\end{aligned}$$

το οποίο είναι η ελάχιστη τιμή του πρώτου infimum στην Εξίσωση 5.13.

Θέτοντας  $g(\mathbf{u}; \boldsymbol{\lambda}, \mathbf{y})$  για τον τελευταίο όρο της Λαγκρανζιανής συνάρτησης  $L(\mathbf{x}, \mathbf{u}; \boldsymbol{\lambda}, \mathbf{y})$ , δηλαδή

$$g(\mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) = \sum_{i=1}^m \left( \frac{1}{2} \lambda_i \mathbf{u}_i^T \mathbf{u}_i - \mathbf{y}_i^T \mathbf{u}_i \right) \quad (\text{A.3})$$

έχουμε ότι

$$\frac{\partial}{\partial \mathbf{u}_i} g(\mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) = \lambda_i \mathbf{u}_i - \mathbf{y}_i.$$

Για να βρούμε το βέλτιστο σημείο, θέτουμε την παραπάνω ισότητα ίση με μηδέν, οπότε έχουμε

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{u}_i} g(\mathbf{u}; \boldsymbol{\lambda}, \mathbf{y}) &= \mathbf{0} \Leftrightarrow \\
\lambda_i \mathbf{u}_i - \mathbf{y}_i &= \mathbf{0} \Leftrightarrow \\
\lambda_i \mathbf{u}_i &= \mathbf{y}_i,
\end{aligned}$$

το οποίο σημαίνει ότι

$$\mathbf{u}_i = \frac{1}{\lambda_i} \mathbf{y}_i. \quad (\text{A.4})$$

Αντικαθιστώντας την Εξίσωση A.4 στην Εξίσωση A.3 έχουμε ότι

$$\begin{aligned}\sum_{i=1}^m \left( \frac{1}{2} \lambda_i \left( \frac{1}{\lambda_i} \mathbf{y}_i \right)^T \left( \frac{1}{\lambda_i} \mathbf{y}_i \right) - \mathbf{y}_i^T \left( \frac{1}{\lambda_i} \mathbf{y}_i \right) \right) &= \sum_{i=1}^m \left( \frac{1}{2} \frac{1}{\lambda_i} \mathbf{y}_i^T \mathbf{y}_i - \frac{1}{\lambda_i} \mathbf{y}_i^T \mathbf{y}_i \right) \\ &= \sum_{i=1}^m \frac{1}{\lambda_i} \mathbf{y}_i^T \mathbf{y}_i \left( \frac{1}{2} - 1 \right) \\ &= -\frac{1}{2} \sum_{i=1}^m \frac{1}{\lambda_i} \mathbf{y}_i^T \mathbf{y}_i \\ &= -\frac{1}{2} \sum_{i=1}^m \frac{\|\mathbf{y}_i\|^2}{\lambda_i},\end{aligned}$$

το οποίο είναι η ελάχιστη τιμή του δεύτερου infimum στην Εξίσωση 5.13.



## ΠΑΡΑΡΤΗΜΑ Β

### ΥΠΟΛΟΓΙΣΜΟΣ ΤΗΣ ΕΝΤΡΟΠΙΑΣ

#### Β.1 Υπολογισμός της εντροπίας με αποκοπή από κάτω

##### Β.1.1 Εκθετική Κατανομή

Για την εκθετική κατανομή έχουμε

$$\int_0^d f_X(x) \ln f_X(x) dx = \ln \lambda F_X(d) - \lambda \int_0^d x f_X(x) dx,$$

όπου

$$\int_0^d x f_X(x) dx = \int_0^d x \lambda e^{-\lambda x} dx = \frac{1 - e^{-\lambda d} (1 + \lambda d)}{\lambda}.$$

Συνεπώς, επειδή  $F_X(d) = 1 - e^{-\lambda d}$ , έχουμε ότι

$$\begin{aligned} \int_0^d f_X(x) \ln f_X(x) dx &= \ln \lambda (1 - e^{-\lambda d}) - 1 + e^{-\lambda d} (1 + \lambda d) \\ &= \ln \lambda - 1 + e^{-\lambda d} (1 + \lambda d - \ln \lambda) \end{aligned}$$

και από το Λήμμα 7.2 έχουμε ότι

$$\begin{aligned}
 H_Y &= \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_0^d f_X(x) \ln f_X(x) dx + \ln \bar{F}_X(d) \left( 1 - \int_0^d f_X(x) dx \right) \right] \\
 &= \frac{1}{e^{-\lambda d}} [1 - \ln \lambda + \ln \lambda - 1 + e^{-\lambda d}(1 + \lambda d - \ln \lambda) - \lambda d(1 - 1 + e^{-\lambda d})] \\
 &= \frac{1}{e^{\lambda d}} e^{-\lambda d} (1 - \ln \lambda) = 1 - \ln \lambda.
 \end{aligned}$$

Παρατηρούμε ότι η εντροπία της εκθετικής κατανομής υπό αποκοπή από κάτω, είναι σταθερή, και δεν εξαρτάται από την τιμή του αφαιρετέου ποσού  $d$ .

### B.1.2 Κατανομή Pareto

Για την κατανομή Pareto ισχύει ότι

$$\int_0^d f_X(x) \ln f_X(x) dx = (\ln \alpha + \alpha \ln \lambda) F_X(d) - (\alpha + 1) \int_0^d \ln(\lambda + x) f_X(x) dx,$$

όπου

$$\begin{aligned}
 \int_0^d \ln(\lambda + x) f_X(x) dx &= \int_0^d \ln(\lambda + x) \alpha \lambda^\alpha (\lambda + x)^{-(\alpha+1)} dx \\
 &= \alpha \lambda^\alpha \int_\lambda^{\lambda+d} \ln y y^{-(\alpha+1)} dy \\
 &= \frac{1}{\alpha} \left[ 1 + \alpha \ln \lambda - \lambda^\alpha (\lambda + d)^{-\alpha} (1 + \alpha \ln(\lambda + d)) \right].
 \end{aligned}$$

Συνεπώς, επειδή  $F_X(d) = 1 - \left(\frac{\lambda}{\lambda+d}\right)^\alpha$  έχουμε ότι

$$\begin{aligned}
 \int_0^d f_X(x) \ln f_X(x) dx &= (\ln \alpha + \alpha \ln \lambda) \left[ 1 - \left(\frac{\lambda}{\lambda+d}\right)^\alpha \right] \\
 &\quad - \frac{(\alpha + 1)}{\alpha} \left[ 1 + \alpha \ln \lambda - \frac{\lambda^\alpha (1 + \alpha \ln(\lambda + d))}{(\lambda + d)^\alpha} \right]
 \end{aligned}$$

και από το Λήμμα 7.2 έχουμε ότι

$$\begin{aligned}
 H_Y &= \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_0^d f_X(x) \ln f_X(x) dx + \bar{F}_X(d) \ln \bar{F}_X(d) \right] \\
 &= \frac{1}{\left(\frac{\lambda}{\lambda+d}\right)^\alpha} \left\{ 1 + \frac{1}{\alpha} + \ln \lambda - \ln \alpha + (\ln \alpha + \alpha \ln \lambda) \left( 1 - \left(\frac{\lambda}{\lambda+d}\right)^\alpha \right) \right. \\
 &\quad \left. - \frac{\alpha+1}{\alpha} \left[ 1 + \alpha \ln \lambda - \frac{\lambda^\alpha (1 + \alpha \ln(\lambda+d))}{(\lambda+d)^\alpha} \right] + \left(\frac{\lambda}{\lambda+d}\right)^\alpha \ln \left(\frac{\lambda}{\lambda+d}\right)^\alpha \right\} \\
 &= 1 + \frac{1}{\alpha} - \ln \alpha + (1 + 2\alpha) \ln(\lambda+d).
 \end{aligned}$$

### B.1.3 Κατανομή Weibull

Για την κατανομή Weibull ισχύει ότι

$$\int_0^d f(x) \ln f(x) dx = (\ln c + \ln \tau) F_X(d) + (\tau - 1) \int_0^d \ln x f(x) dx - \int_0^d cx^\tau f(x) dx.$$

όπου

$$\begin{aligned}
 \int_0^d \ln x f(x) dx &= \int_0^d \ln x c \tau x^{\tau-1} e^{-cx^\tau} dx = \int_0^d \ln x \tau \frac{cx^\tau}{x} e^{-cx^\tau} dx \\
 &= \int_0^{cd^\tau} \frac{1}{\tau} \ln \frac{y}{c} \tau y \left(\frac{y}{c}\right)^{\frac{1}{\tau}-1} e^{-y} \frac{1}{c} \frac{1}{\tau} \left(\frac{y}{c}\right)^{\frac{1}{\tau}-1} dy \\
 &= \frac{1}{\tau} \int_0^{cd^\tau} (\ln y - \ln c) e^{-y} dy \\
 &= -\frac{1}{\tau} (\gamma + \Gamma(0, cd^\tau) + \ln c + e^{-cd^\tau} \tau \ln d),
 \end{aligned}$$

και

$$\begin{aligned}
 \int_0^d cx^\tau f(x) dx &= \int_0^d cx^\tau c \tau x^{\tau-1} e^{-cx^\tau} dx = \int_0^{cd^\tau} y \tau y \frac{1}{\left(\frac{y}{c}\right)^{\frac{1}{\tau}}} e^{-y} \frac{1}{c} \frac{1}{\tau} \left(\frac{y}{c}\right)^{\frac{1}{\tau}-1} dx \\
 &= \int_0^{cd^\tau} y^2 \left(\frac{y}{c}\right)^{-1} e^{-y} \frac{1}{c} dy = \int_0^{cd^\tau} ye^{-y} dy \\
 &= 1 - (1 + cd^\tau) e^{-cd^\tau}.
 \end{aligned}$$

Συνεπώς, επειδή  $F_X(d) = 1 - e^{-cd^\tau}$  έχουμε ότι

$$\int_0^d f_X(x) \ln f_X(x) dx = (\ln c + \ln \tau) (1 - e^{-cd^\tau}) - 1 + (1 + cd^\tau) e^{-cd^\tau} - \frac{\tau - 1}{\tau} (\gamma + \Gamma(0, cd^\tau) + \ln c + e^{-cd^\tau} \tau \ln d)$$

και από το Λήμμα 7.2 έχουμε ότι

$$\begin{aligned} H_Y &= \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_0^d f_X(x) \ln f_X(x) dx + \bar{F}_X(d) \ln \bar{F}_X(d) \right] \\ &= \frac{1}{e^{-cd^\tau}} \left[ 1 - \ln c - \ln \tau + (\tau - 1) \frac{\gamma + \ln c}{\tau} + (\ln c + \ln \tau) (1 - e^{-cd^\tau}) - 1 + (1 + cd^\tau) e^{-cd^\tau} - \frac{\tau - 1}{\tau} (\gamma + \Gamma(0, cd^\tau) + \ln c + e^{-cd^\tau} \tau \ln d) - cd^\tau e^{-cd^\tau} \right] \\ &= 1 - \ln c - \ln \tau - (\tau - 1) \ln d - \frac{(\tau - 1) e^{cd^\tau} \Gamma(0, cd^\tau)}{\tau}. \end{aligned}$$

### B.1.4 Κατανομή Gamma

Για την κατανομή Gamma ισχύει ότι

$$\int_0^d f(x) \ln f(x) dx = (\alpha \ln \lambda - \ln \Gamma(\alpha)) F_X(d) + (\alpha - 1) \int_0^d \ln x f(x) dx - \lambda \int_0^d x f(x) dx.$$

όπου

$$\begin{aligned} \int_0^d x f(x) dx &= \frac{1}{\Gamma(\alpha)} \int_0^d x \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx = \frac{1}{\Gamma(\alpha)} \int_0^d (\lambda x)^\alpha e^{-\lambda x} dx \\ &= \frac{1}{\lambda \Gamma(\alpha)} \int_0^{\lambda d} y^\alpha e^{-y} dy = \frac{\Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda d)}{\lambda \Gamma(\alpha)} \end{aligned}$$

και

$$\begin{aligned} \int_0^d \ln x f(x) dx &= \frac{1}{\Gamma(\alpha)} \int_0^d \ln x \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx = \frac{1}{\Gamma(\alpha)} \int_0^d \ln x \lambda (\lambda x)^{\alpha-1} e^{-\lambda x} dx \\ &= -\frac{1}{\Gamma(\alpha)} \left[ \Gamma(\alpha, \lambda d) \ln d + G_{2,0}^{3,0} \left( \lambda d \middle| \begin{matrix} \mathbf{a} \\ \mathbf{b} \end{matrix} \right) - \Gamma(\alpha) (\psi(\alpha) + \ln \lambda) \right], \end{aligned}$$

όπου  $G_{2,0}^{3,0} \left( \lambda d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right)$  είναι η συνάρτηση  $G$  του Meijer με  $\mathbf{a} = (1, 1)$  και  $\mathbf{b} = (0, 0, \alpha)$ .  
 Συνεπώς, επειδή  $F_X(d) = \Gamma(\alpha, \lambda d)$ , έχουμε ότι

$$\int_0^d f(x) \ln f(x) dx = (\alpha \ln \lambda - \ln \Gamma(\alpha)) \Gamma(\alpha, \lambda d) - \frac{\Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda d)}{\Gamma(\alpha)} \\ - \frac{\alpha - 1}{\Gamma(\alpha)} \left[ \Gamma(\alpha, \lambda d) \ln d + G_{2,0}^{3,0} \left( \lambda d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) - \Gamma(\alpha) (\psi(\alpha) + \ln \lambda) \right]$$

και από το Λήμμα 7.2 έχουμε ότι

$$H_Y = \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_0^d f_X(x) \ln f_X(x) dx + \bar{F}_X(d) \ln \bar{F}_X(d) \right] \\ = \frac{1}{1 - \Gamma(\alpha, \lambda d)} \left\{ \alpha + \ln \Gamma(\alpha) - (\alpha - 1) \psi(\alpha) - \ln \lambda + (\alpha \ln \lambda - \ln \Gamma(\alpha)) \Gamma(\alpha, \lambda d) \right. \\ \left. - \frac{\alpha - 1}{\Gamma(\alpha)} \left[ \Gamma(\alpha, \lambda d) \ln d + G_{2,0}^{3,0} \left( \lambda d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) - \Gamma(\alpha) (\psi(\alpha) + \ln \lambda) \right] \right. \\ \left. - \frac{\Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda d)}{\Gamma(\alpha)} + (1 - \Gamma(\alpha, \lambda d)) \ln (1 - \Gamma(\alpha, \lambda d)) \right\} \\ = \frac{1}{\varphi(\alpha, \lambda d)} \left\{ \Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda d) + (\alpha - 1) \Gamma(\alpha, \lambda d) \ln d \right. \\ \left. - \Gamma(\alpha) \left[ \alpha + (\alpha - 2) \ln \lambda + \ln \xi(\alpha, \lambda d) + \Gamma(\alpha, \lambda d) \ln \frac{\lambda^\alpha}{\xi(\alpha, \lambda d)} \right] \right. \\ \left. + (\alpha - 1) G_{2,0}^{3,0} \left( \lambda d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) \right\},$$

όπου  $\varphi(\alpha, \lambda d) = \Gamma(\alpha) (\Gamma(\alpha, \lambda d) - 1)$  και  $\xi(\alpha, \lambda d) = \Gamma(\alpha) (1 - \Gamma(\alpha, \lambda d))$ . Για  $\alpha$  ακέραιο παίρνουμε την εντροπία για την αποκομμένη από κάτω Erlang κατανομή.

### B.1.5 Transformed Gamma Κατανομή

Για την Transformed Gamma κατανομή ισχύει ότι

$$\int_0^d f(x) \ln f(x) dx = (\tau \alpha \ln \lambda + \ln \tau - \ln \Gamma(\alpha)) F_X(d) \\ + (\tau \alpha - 1) \int_0^d \ln x f(x) dx - \int_0^d (\lambda x)^\tau f(x) dx.$$

όπου

$$\begin{aligned}
 \int_0^d \ln x f(x) dx &= \frac{1}{\Gamma(\alpha)} \int_0^d \ln x \lambda^{\tau\alpha} \tau x^{\tau\alpha-1} e^{-(\lambda x)^\tau} dx \\
 &= \frac{1}{\tau\Gamma(\alpha)} \int_0^{(\lambda d)^\tau} (\ln z - \tau \ln \lambda) z^{\alpha-1} e^{-z} \frac{1}{\lambda^\tau} dz \\
 &= -\frac{1}{\tau\Gamma(\alpha)} \left[ \tau \ln d \Gamma(\alpha, (\lambda d)^\tau) + G_{2,0}^{3,0} \left( (\lambda d)^\tau \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) \right. \\
 &\quad \left. + \Gamma(\alpha) (\tau \ln \lambda - \psi(\alpha)) \right],
 \end{aligned}$$

όπου  $G_{2,0}^{3,0} \left( (\lambda d)^\tau \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right)$  είναι η συνάρτηση  $G$  του Meijer με  $\mathbf{a} = (1, 1)$  και  $\mathbf{b} = (0, 0, \alpha)$  και

$$\begin{aligned}
 \int_0^d (\lambda x)^\tau f(x) dx &= \frac{1}{\Gamma(\alpha)} \int_0^d \lambda^\tau x^\tau \lambda^{\tau\alpha} \tau x^{\tau\alpha-1} e^{-(\lambda x)^\tau} dx \\
 &= \frac{1}{\Gamma(\alpha)} \int_0^{(\lambda d)^\tau} \lambda^\tau \lambda^{\tau\alpha} \frac{z^\alpha}{\lambda^{\tau\alpha}} e^{-z} \frac{1}{\lambda^\tau} dz \\
 &= \frac{1}{\Gamma(\alpha)} \int_0^{(\lambda d)^\tau} z^\alpha e^{-z} dz \\
 &= \frac{\Gamma(a+1) - \Gamma(a+1, (\lambda d)^\tau)}{\Gamma(a)}.
 \end{aligned}$$

Συνεπώς, επειδή  $F_X(d) = \Gamma(\alpha, (\lambda d)^\tau)$ , έχουμε ότι

$$\begin{aligned}
 \int_0^d f(x) \ln f(x) dx &= (\tau\alpha \ln \lambda + \ln \tau - \ln \Gamma(\alpha)) \Gamma(\alpha, (\lambda d)^\tau) \\
 &\quad - \frac{\Gamma(a+1) - \Gamma(a+1, (\lambda d)^\tau)}{\Gamma(a)} \\
 &\quad - \frac{\tau\alpha - 1}{\tau\Gamma(\alpha)} \left[ \tau \ln d \Gamma(\alpha, (\lambda d)^\tau) + G_{2,0}^{3,0} \left( (\lambda d)^\tau \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) \right. \\
 &\quad \left. + \Gamma(\alpha) (\tau \ln \lambda - \psi(\alpha)) \right]
 \end{aligned}$$

και από το Λήμμα 7.2 έχουμε ότι

$$\begin{aligned}
H_Y &= \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_0^d f_X(x) \ln f_X(x) dx + \bar{F}_X(x) \ln \bar{F}_X(d) \right] \\
&= \frac{1}{1 - \Gamma(\alpha, (\lambda d)^\tau)} \left\{ \alpha - \ln \tau - \ln \lambda + \ln \Gamma(\alpha) + \left( \frac{1}{\tau} - \alpha \right) \psi(\alpha) \right. \\
&\quad \left. + (\tau \alpha \ln \lambda + \ln \tau - \ln \Gamma(\alpha)) \Gamma(\alpha, (\lambda d)^\tau) - \frac{\Gamma(\alpha + 1) - \Gamma(\alpha + 1, (\lambda d)^\tau)}{\Gamma(\alpha)} \right. \\
&\quad \left. - \frac{\tau \alpha - 1}{\tau \Gamma(\alpha)} \left[ \tau \Gamma(\alpha, (\lambda d)^\tau) \ln d + G_{2,0}^{3,0} \left( (\lambda d)^\tau \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) + \Gamma(\alpha) (\tau \ln \lambda - \psi(\alpha)) \right] \right. \\
&\quad \left. + (1 - \Gamma(\alpha, (\lambda d)^\tau)) \ln (1 - \Gamma(\alpha, (\lambda d)^\tau)) \right\} \\
&= \frac{1}{\varphi(\alpha, (\lambda d)^\tau)} \left\{ \frac{\tau \alpha - 1}{\tau} G_{2,0}^{3,0} \left( (\lambda d)^\tau \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) + \Gamma(\alpha + 1) \right. \\
&\quad \left. - \Gamma(\alpha + 1, (\lambda d)^\tau) + (\tau \alpha - 1) \Gamma(\alpha, (\lambda d)^\tau) \ln d \right. \\
&\quad \left. - \Gamma(\alpha) \left[ \alpha + \ln \frac{\lambda^{\tau \alpha} \tau}{\xi(\alpha, (\lambda d)^\tau)} + \Gamma(\alpha, (\lambda d)^\tau) \ln \frac{\lambda^{\tau \alpha} \tau}{\xi(\alpha, (\lambda d)^\tau)} \right] \right\},
\end{aligned}$$

όπου  $\varphi(\alpha, (\lambda d)^\tau) = \Gamma(\alpha) (\Gamma(\alpha, (\lambda d)^\tau) - 1)$  και  $\xi(\alpha, (\lambda d)^\tau) = \Gamma(\alpha) (1 - \Gamma(\alpha, (\lambda d)^\tau))$ .

### B.1.6 Loggamma Κατανομή

Για την Loggamma κατανομή ισχύει ότι

$$\begin{aligned}
\int_1^d f(x) \ln f(x) dx &= (\alpha \ln \lambda - \ln \Gamma(\alpha)) F_X(d) + (\alpha - 1) \int_1^d \ln(\ln x) f(x) dx \\
&\quad - (\lambda + 1) \int_1^d \ln x f(x) dx.
\end{aligned}$$

όπου

$$\begin{aligned}
\int_0^d \ln x f(x) dx &= \int_1^d \ln x \frac{\lambda^\alpha (\ln x)^{\alpha-1}}{x^{\lambda+1} \Gamma(\alpha)} dx = \frac{1}{\Gamma(\alpha)} \int_1^d \frac{(\lambda \ln x)^\alpha}{x^{\lambda+1}} dx \\
&= \frac{1}{\lambda \Gamma(\alpha)} \int_0^{\lambda \ln d} y^\alpha e^{-y} dy \\
&= \frac{\Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda \ln d)}{\lambda \Gamma(\alpha)}
\end{aligned}$$

και

$$\begin{aligned}
\int_1^d \ln(\ln x) f(x) dx &= \int_1^d \ln(\ln x) \frac{\lambda^\alpha (\ln x)^{\alpha-1}}{x^{\lambda+1} \Gamma(\alpha)} dx \\
&= \frac{1}{\Gamma(\alpha)} \int_0^{\lambda \ln d} \lambda^\alpha \ln \frac{z}{\lambda} \left(\frac{z}{\lambda}\right)^{\alpha-1} e^{-z} \frac{1}{\lambda} dz \\
&= \frac{1}{\Gamma(\alpha)} \int_0^{\lambda \ln d} (\ln z - \ln \lambda) z^{\alpha-1} e^{-z} dz \\
&= \frac{\Gamma(\alpha)\Gamma(\lambda) + \Gamma(\alpha, \lambda \ln d) \ln(\ln d) + G_{2,0}^{3,0} \left( \lambda \ln d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right)}{\Gamma(\alpha)} + \psi(\alpha),
\end{aligned}$$

όπου  $G_{2,0}^{3,0} \left( \lambda \ln d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right)$  είναι η συνάρτηση  $G$  του Meijer με  $\mathbf{a} = (1, 1)$  και  $\mathbf{b} = (0, 0, \alpha)$ .  
 Συνεπώς, επειδή  $F_X(d) = \Gamma(\alpha, \lambda \ln d)$ , έχουμε ότι

$$\begin{aligned}
\int_1^d f(x) \ln f(x) dx &= (\alpha \ln \lambda - \ln \Gamma(\alpha)) \Gamma(\alpha, \lambda \ln d) \\
&\quad - \frac{\alpha - 1}{\Gamma(\alpha)} \left( \Gamma(\alpha)\Gamma(\lambda) + \Gamma(\alpha, \lambda \ln d) \ln(\ln d) + G_{2,0}^{3,0} \left( \lambda \ln d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) \right) \\
&\quad + (\alpha - 1)\psi(\alpha) - \frac{\lambda + 1}{\lambda \Gamma(\alpha)} (\Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda \ln d))
\end{aligned}$$

και από το Λήμμα 7.2 έχουμε ότι

$$\begin{aligned}
H_Y &= \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_1^d f_X(x) \ln f_X(x) dx + \bar{F}_X(d) \ln \bar{F}_X(d) \right] \\
&= \frac{1}{1 - \Gamma(\alpha, \lambda \ln d)} \left[ \alpha \left( 1 + \frac{1}{\lambda} \right) + \ln \Gamma(\alpha) - (\alpha - 1)\psi(\alpha) - \ln \lambda + \alpha \ln \lambda \Gamma(\alpha, \lambda \ln d) \right. \\
&\quad \left. - \frac{\alpha - 1}{\Gamma(\alpha)} \left( \Gamma(\alpha)\Gamma(\lambda) + \Gamma(\alpha, \lambda \ln d) \ln(\ln d) + G_{2,0}^{3,0} \left( \lambda \ln d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) \right) \right. \\
&\quad \left. + (\alpha - 1)\psi(\alpha) - \frac{\lambda + 1}{\lambda \Gamma(\alpha)} (\Gamma(\alpha + 1) - \Gamma(\alpha + 1, \lambda \ln d)) - \ln \Gamma(\alpha)\Gamma(\alpha, \lambda \ln d) \right. \\
&\quad \left. + (1 - \Gamma(\alpha, \lambda \ln d)) \ln(1 - \Gamma(\alpha, \lambda \ln d)) \right] \\
&= \frac{1}{\lambda \varphi(\alpha, \lambda \ln d)} \left\{ \lambda(\alpha - 1) G_{2,0}^{3,0} \left( \lambda \ln d \left| \begin{array}{c} \mathbf{a} \\ \mathbf{b} \end{array} \right. \right) + (\alpha - 1) \lambda \Gamma(\alpha, \lambda \ln d) \ln(\ln d) \right. \\
&\quad \left. - (\lambda + 1) \Gamma(\alpha + 1, \lambda \ln d) \right. \\
&\quad \left. + \lambda (1 - \Gamma(\alpha, \lambda \ln d)) [\Gamma(\alpha + 1) \ln \lambda - \Gamma(\alpha) \ln \xi(\alpha, \lambda \ln d)] \right\},
\end{aligned}$$



όπου  $\varphi(\alpha, \lambda \ln d) = \Gamma(\alpha) (\Gamma(\alpha, \lambda \ln d) - 1)$  και  $\xi(\alpha, \lambda \ln d) = \Gamma(\alpha) (1 - \Gamma(\alpha, \lambda \ln d))$ .

### B.1.7 Lognormal Κατανομή

Για την Lognormal κατανομή έχουμε ότι

$$\begin{aligned} \int_0^d f(x) \ln f(x) dx &= \left( -\ln \sigma - \ln \sqrt{2\pi} - \frac{\mu^2}{\sigma^2} \right) F_X(d) - \frac{1}{2\sigma^2} \int_0^d (\ln x)^2 f(x) dx \\ &+ \left( \frac{\mu}{\sigma^2} - 1 \right) \int_0^d \ln x f(x) dx. \end{aligned}$$

όπου

$$\begin{aligned} \int_0^d \ln x f(x) dx &= \int_0^d \ln x \frac{\exp \left\{ -\frac{1}{2} \left( \frac{\ln x - \mu}{\sigma} \right)^2 \right\}}{x \sigma \sqrt{2\pi}} dx \\ &= \int_{-\infty}^{\ln d} y \frac{1}{e^y \sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (y - \mu)^2 \right\} e^y dy \\ &= \int_{-\infty}^{\ln d} y \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (y - \mu)^2 \right\} dy \\ &= -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln d)^2}{2\sigma^2}} + \frac{\mu}{2} - \frac{\mu}{2} \text{Erf} \left( \frac{\mu - \ln d}{\sqrt{2}\sigma} \right), \end{aligned}$$

όπου  $\text{Erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$  είναι η συνάρτηση σφάλματος (error function) (Abramowitz and Stegun, 1972) και

$$\begin{aligned} \int_0^d (\ln x)^2 f(x) dx &= \int_0^d (\ln x)^2 \frac{\exp \left\{ -\frac{1}{2} \left( \frac{\ln x - \mu}{\sigma} \right)^2 \right\}}{x \sigma \sqrt{2\pi}} dx \\ &= \frac{\mu^2 + \sigma^2}{2} \text{Erfc} \left( \frac{\mu - \ln d}{\sqrt{2}\sigma} \right) - \frac{\sigma(\mu + \ln d)}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln d)^2}{2\sigma^2}}, \end{aligned}$$

όπου  $\text{Erfc}(z) = 1 - \text{Erf}(z)$  είναι η συμπληρωματική συνάρτηση σφάλματος (complementary error function) (Abramowitz and Stegun, 1972). Συνεπώς, επειδή  $F_X(d) = \Phi \left( \frac{\ln d - \mu}{\sigma} \right)$  έχουμε ότι

$$\begin{aligned} \int_0^d f(x) \ln f(x) dx &= \left( -\ln \sigma - \ln \sqrt{2\pi} - \frac{\mu^2}{\sigma^2} \right) \Phi \left( \frac{\ln d - \mu}{\sigma} \right) \\ &- \frac{1}{2\sigma^2} \left[ \frac{\mu^2 + \sigma^2}{2} \text{Erfc} \left( \frac{\mu - \ln d}{\sqrt{2}\sigma} \right) - \frac{\sigma(\mu + \ln d)}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln d)^2}{2\sigma^2}} \right] \\ &+ \left( \frac{\mu}{\sigma^2} - 1 \right) \left[ -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln d)^2}{2\sigma^2}} + \frac{\mu}{2} - \frac{\mu}{2} \text{Erf} \left( \frac{\mu - \ln d}{\sqrt{2}\sigma} \right) \right] \end{aligned}$$

και από το Λήμμα 7.2 έχουμε ότι

$$\begin{aligned}
H_Y &= \frac{1}{\bar{F}_X(d)} \left[ H_X + \int_0^d f_X(x) \ln f_X(x) dx + \bar{F}_X(d) \ln \bar{F}_X(d) \right] \\
&= \frac{1}{1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right)} \left\{ \frac{1}{2} + \mu + \ln(\sigma\sqrt{2\pi}) - \left( \ln(\sigma\sqrt{2\pi}) + \frac{\mu^2}{\sigma^2} \right) \Phi\left(\frac{\ln d - \mu}{\sigma}\right) \right. \\
&\quad - \frac{1}{2\sigma^2} \left[ \frac{\mu^2 + \sigma^2}{2} \operatorname{Erfc}\left(\frac{\mu - \ln d}{\sqrt{2}\sigma}\right) - \frac{\sigma(\mu + \ln d)}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln d)^2}{2\sigma^2}} \right] \\
&\quad + \left( \frac{\mu}{\sigma^2} - 1 \right) \left[ -\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{(\mu - \ln d)^2}{2\sigma^2}} + \frac{\mu}{2} - \frac{\mu}{2} \operatorname{Erf}\left(\frac{\mu - \ln d}{\sqrt{2}\sigma}\right) \right] \\
&\quad \left. + \left( 1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right) \right) \ln \left( 1 - \Phi\left(\frac{\ln d - \mu}{\sigma}\right) \right) \right\}.
\end{aligned}$$

### B.1.8 Κατανομή Burr

Για την κατανομή Burr έχουμε ότι

$$\begin{aligned}
\int_0^d f(x) \ln f(x) dx &= (\ln \alpha + \ln \tau + \alpha \ln \lambda) F_X(d) + (\tau - 1) \int_0^d \ln x f(x) dx \\
&\quad - (\alpha + 1) \int_0^d (\lambda + x^\tau) f(x) dx,
\end{aligned}$$

όπου  $F_X(d) = 1 - \left(\frac{\lambda}{\lambda + d^\tau}\right)^\alpha$ . Όμως το ολοκλήρωμα

$$\int_0^d \ln x f(x) dx$$

δεν μπορεί να υπολογιστεί σε κλειστή μορφή. Συνεπώς η εντροπία της αποκομμένης από κάτω κατανομής Burr δεν μπορεί να υπολογιστεί σε κλειστή μορφή.

### B.1.9 Γενικευμένη Pareto Κατανομή

Για τη γενικευμένη Pareto κατανομή ισχύει ότι

$$\begin{aligned}
\int_0^d f(x) \ln f(x) dx &= \left( \ln \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\Gamma(k)} + \alpha \ln \lambda \right) F_X(d) + (k - 1) \int_0^d \ln x f(x) dx \\
&\quad - (k + \alpha) \int_0^d \ln(\lambda + x) f(x) dx,
\end{aligned}$$

όπου  $F_X(d) = B(k, \alpha, \frac{d}{\lambda+d})$ . Όμως τα ολοκληρώματα

$$\int_0^d \ln x f(x) dx \quad \text{και} \quad \int_0^d \ln(\lambda + x) f(x) dx$$

δεν μπορούν να υπολογιστούν σε κλειστή μορφή. Συνεπώς η εντροπία δεν μπορεί να υπολογιστεί σε κλειστή μορφή.

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΡΑΙΑ

# РАНЕЕЗНАМО ПЕРПАА

# ΠΑΡΑΡΤΗΜΑ Γ

## ΚΩΔΙΚΕΣ ΓΙΑ ΤΟΝ ΥΠΟΛΟΓΙΣΜΟ ΤΩΝ ΕΞΟΜΑΛΥΜΕΝΩΝ ΤΙΜΩΝ

### Γ.1 Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με το μέτρο των Cressie and Read και 5 περιορισμούς

Για να υπολογίσουμε τις εξομαλυμένες τιμές με το μέτρο των Cressie and Read και 5 περιορισμούς, για τα τρία σύνολα δεδομένων, γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, τους παρακάτω κώδικες:

#### Γ.1.1 Σύνολο Δεδομένων L85

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 5 constraints for the L85 data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.03870 0.03909 0.04609 0.04839 0.06644 0.06091 0.07376 0.08019 0.08874 0.07455  
0.11195 0.12320 0.14031 0.14879 0.19020 0.21628 0.20718 0.28866 0.13362 0.38265; ! The ungraduated values;

$L = 15012\ 11871\ 10002\ 8949\ 7751\ 6140\ 4718\ 3791\ 2806\ 2240\ 1715\ 1388\ 898\ 578\ 510\ 430$   
 $362\ 291\ 232\ 196$ ; ! The number of people at risk;  
 $AGE = 75\ 76\ 77\ 78\ 79\ 80\ 81\ 82\ 83\ 84\ 85\ 86\ 87\ 88\ 89\ 90\ 91\ 92\ 93\ 94$ ; ! The ages;  
 $S = 2$ ; ! The value of parameter;  
 ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;  
 $MIN = (1/(S * (S + 1))) * @SUM(SET : Y * ((Y/X)^S - 1));$   
 ! The graduated values should lie between 0 and 1;  
 $@FOR(SET : @BND(0, Y, 1));$   
 ! Subject to the smoothness constraint;  
 $@FOR(SET(K)|K#LE#1 :$   
      $@SUM(SET(I)|I#LE#17 :$   
          $(Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))^2 <= .00004);$   
 !... the monotonicity constraint;  
 $@FOR(SET(I)|I#LE#19 :$   
      $Y(I + 1) - Y(I) >= 0);$   
 !... the convexity constraint;  
 $@FOR(SET(I)|I#LE#18 :$   
      $Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);$   
 !... the first actuarial constraint;  
 $@FOR(SET : @SUM(SET(I)|I#LE#20 : L(I) * X(I)) =$   
      $@SUM(SET(I)|I#LE#20 : L(I) * Y(I));$   
 !... the second actuarial constraint;  
 $@FOR(SET : @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * X(I)) =$   
      $@SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * Y(I));$   
 END

### Γ.1.2 Σύνολο Δεδομένων HK01M

MODEL:  
 ! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 5 constraints for the HK01M data set;  
 SETS:  
 ! There are 16;

```

SET / 1..16 /: X, ! ungraduated values;
           Y, ! graduated values;
           L, ! people at risk;
           AGE; ! ages;

ENDSETS

DATA:
X = 0.0192308 0.0256304 0.0299252 0.035854 0.0390041 0.0353063 0.055409 0.0493601
0.059126 0.0750853 0.0484581 0.116667 0.0681818 0.234568 0.117647 0.294118; ! The
ungraduated values;
L = 12860 2419 2005 1534 1205 963 758 547 389 293 227 180 132 81 51 34; ! The number
of people at risk;
AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85; ! The ages;
S = 2; ! The value of parameter;
ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;
MIN = (1/(S * (S + 1))) * @SUM(SET : Y * ((Y/X)S - 1));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
      @SUM(SET(I)|I#LE#13 :
            (Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))2) <= .00003);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#15 :
      Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#14 :
      Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the first actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#16 : L(I) * X(I)) =
          @SUM(SET(I)|I#LE#16 : L(I) * Y(I)));

!... the second actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#16 : AGE(I) * L(I) * X(I)) =
          @SUM(SET(I)|I#LE#16 : AGE(I) * L(I) * Y(I)));

END

```

### Γ.1.3 Σύνολο Δεδομένων ΗΚ01F

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 5 constraints for the ΗΚ01F data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.011344 0.016704 0.013775 0.018219 0.022583 0.021531 0.035387 0.031236 0.036882  
0.039539 0.058961 0.028872 0.033420 0.073951 0.101383 0.077762 0.057996 0.114286 0.18465

0.2; ! The ungraduated values;

L = 3747 3263 2759 2141 1638 1300 1046 800 569 379 237 173 120 68 39 26 17 9 5 5; !

The number of people at risk;

AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89; ! The ages;

S = 2; ! The value of parameter;

ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;

MIN = (1/(S \* (S + 1))) \* @SUM(SET : Y \* ((Y/X)<sup>S</sup> - 1));

! The graduated values should lie between 0 and 1;

@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;

@FOR(SET(K)|K#LE#1 :

@SUM(SET(I)|I#LE#17 :

(Y(I+3) - 3\*Y(I+2) + 3\*Y(I+1) - Y(I))<sup>2</sup> <= .000015);

!... the monotonicity constraint;

@FOR(SET(I)|I#LE#19 :

Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;

@FOR(SET(I)|I#LE#18 :



$$Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);$$

!... the first actuarial constraint;

```
@FOR(SET : @SUM(SET(I)|I#LE#20 : L(I) * X(I)) =
    @SUM(SET(I)|I#LE#20 : L(I) * Y(I)));
```

!... the second actuarial constraint;

```
@FOR(SET : @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * X(I)) =
    @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * Y(I)));
```

END

## Γ.2 Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με το μέτρο των Cressie and Read και 6 περιορισμούς

Για να υπολογίσουμε τις εξομαλυμένες τιμές με το μέτρο των Cressie and Read και 6 περιορισμούς, για τα τρία σύνολα δεδομένων, γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, τους παρακάτω κώδικες:

### Γ.2.1 Σύνολο Δεδομένων L85

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 6 constraints for the L85 data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.03870 0.03909 0.04609 0.04839 0.06644 0.06091 0.07376 0.08019 0.08874 0.07455  
0.11195 0.12320 0.14031 0.14879 0.19020 0.21628 0.20718 0.28866 0.13362 0.38265; ! The  
ungraduated values;

L = 15012 11871 10002 8949 7751 6140 4718 3791 2806 2240 1715 1388 898 578 510 430  
362 291 232 196; ! The number of people at risk;

AGE = 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94; ! The ages;

S = 2; ! The value of parameter;

ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;

$MIN = (1/(S * (S + 1))) * @SUM(SET : Y * ((Y/X)^S - 1));$

! The graduated values should lie between 0 and 1;

@FOR(SET : @BND(0,Y,1));

! Subject to the smoothness constraint;

@FOR(SET(K)|K#LE#1 :

@SUM(SET(I)|I#LE#17 :

$(Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))^2 <= .00004);$

!... the monotonicity constraint;

@FOR(SET(I)|I#LE#19 :

$Y(I + 1) - Y(I) >= 0);$

!... the convexity constraint;

@FOR(SET(I)|I#LE#18 :

$Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);$

!... the first actuarial constraint;

@FOR(SET : @SUM(SET(I)|I#LE#20 : L(I) \* X(I)) =

@SUM(SET(I)|I#LE#20 : L(I) \* Y(I));

!... the second actuarial constraint;

@FOR(SET : @SUM(SET(I)|I#LE#20 : AGE(I) \* L(I) \* X(I)) =

@SUM(SET(I)|I#LE#20 : AGE(I) \* L(I) \* Y(I));

!... the additional constraint;

@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =

@SUM(SET(I)|I#LE#20 : Y(I));

END

## Γ.2.2 Σύνολο Δεδομένων HK01M

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 6 constraints for the HK01M data set;

SETS:

! There are 16;

SET / 1..16 /: X, ! ungraduated values;

```

        Y, ! graduated values;
        L, ! people at risk;
        AGE; ! ages;
ENDSETS

DATA:
X = 0.0192308 0.0256304 0.0299252 0.035854 0.0390041 0.0353063 0.055409 0.0493601
0.059126 0.0750853 0.0484581 0.116667 0.0681818 0.234568 0.117647 0.294118; ! The
ungraduated values;
L = 12860 2419 2005 1534 1205 963 758 547 389 293 227 180 132 81 51 34; ! The number
of people at risk;
AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85; ! The ages;
S = 2; ! The value of parameter;
ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;
MIN = (1/(S * (S + 1))) * @SUM(SET : Y * ((Y/X)S - 1));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
    @SUM(SET(I)|I#LE#13 :
        (Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))2) <= .00003);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#15 :
    Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#14 :
    Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the first actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#16 : L(I) * X(I)) =
    @SUM(SET(I)|I#LE#16 : L(I) * Y(I)));

!... the second actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#16 : AGE(I) * L(I) * X(I)) =
    @SUM(SET(I)|I#LE#16 : AGE(I) * L(I) * Y(I)));

!... the additional constraint;

```

```

@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =
        @SUM(SET(I)|I#LE#20 : Y(I)));
END

```

### Γ.2.3 Σύνολο Δεδομένων HK01F

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 6 constraints for the HK01F data set;

SETS:

```

! There are 20;
SET / 1..20 /: X, ! ungraduated values;
           Y, ! graduated values;
           L, ! people at risk;
           AGE; ! ages;

```

ENDSETS

DATA:

```

X = 0.011344 0.016704 0.013775 0.018219 0.022583 0.021531 0.035387 0.031236 0.036882
0.039539 0.058961 0.028872 0.033420 0.073951 0.101383 0.077762 0.057996 0.114286 0.18465
0.2; ! The ungraduated values;
L = 3747 3263 2759 2141 1638 1300 1046 800 569 379 237 173 120 68 39 26 17 9 5 5; !
The number of people at risk;
AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89; ! The ages;
S = 2; ! The value of parameter;
ENDDATA

```

! Minimize the Cressie and Read power divergence with parameter S;

```

MIN = (1/(S*(S+1))) * @SUM(SET : Y * ((Y/X)S - 1));

```

! The graduated values should lie between 0 and 1;

```

@FOR(SET : @BND(0, Y, 1));

```

! Subject to the smoothness constraint;

```

@FOR(SET(K)|K#LE#1 :
        @SUM(SET(I)|I#LE#17 :
                (Y(I+3)-3*Y(I+2)+3*Y(I+1)-Y(I))2) <= .000015);

```

!... the monotonicity constraint;

```

@FOR(SET(I)|I#LE#19 :

```

```

        Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#18 :
        Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the first actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : L(I) * X(I)) =
        @SUM(SET(I)|I#LE#20 : L(I) * Y(I)));

!... the second actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * X(I)) =
        @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * Y(I)));

!... the additional constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =
        @SUM(SET(I)|I#LE#20 : Y(I)));

END

```

### Γ.3 Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με το μέτρο των Cressie and Read και 4 περιορισμούς

Για να υπολογίσουμε τις εξομαλυμένες τιμές με το μέτρο των Cressie and Read και 4 περιορισμούς, για τα τρία σύνολα δεδομένων, γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, τους παρακάτω κώδικες:

#### Γ.3.1 Σύνολο Δεδομένων L85

```

MODEL:
! This model determines the graduated values by the minimization of the Cressie and
Read power divergence with parameter S subject to 4 constraints for the L85 data set;

SETS:
! There are 20;
SET / 1..20 /: X, ! ungraduated values;
        Y, ! graduated values;
        L, ! people at risk;
        AGE; ! ages;

ENDSETS

DATA:

```

```

X = 0.03870 0.03909 0.04609 0.04839 0.06644 0.06091 0.07376 0.08019 0.08874 0.07455
0.11195 0.12320 0.14031 0.14879 0.19020 0.21628 0.20718 0.28866 0.13362 0.38265; ! The
ungraduated values;
L = 15012 11871 10002 8949 7751 6140 4718 3791 2806 2240 1715 1388 898 578 510 430
362 291 232 196; ! The number of people at risk;
AGE = 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94; ! The ages;
S = 2; ! The value of parameter;
ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;
MIN = (1/(S * (S + 1))) * @SUM(SET : Y * ((Y/X)S - 1));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
    @SUM(SET(I)|I#LE#17 :
        (Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))2) <= .00004);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#19 :
    Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#18 :
    Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the additional constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =
    @SUM(SET(I)|I#LE#20 : Y(I)));

END

```

### Γ.3.2 Σύνολο Δεδομένων HK01M

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 4 constraints for the HK01M data set;

SETS:

! There are 16;

SET / 1..16 /: X, ! ungraduated values;

```

        Y, ! graduated values;
        L, ! people at risk;
        AGE; ! ages;
ENDSETS

DATA:
X = 0.0192308 0.0256304 0.0299252 0.035854 0.0390041 0.0353063 0.055409 0.0493601
0.059126 0.0750853 0.0484581 0.116667 0.0681818 0.234568 0.117647 0.294118; ! The
ungraduated values;
L = 12860 2419 2005 1534 1205 963 758 547 389 293 227 180 132 81 51 34; ! The number
of people at risk;
AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85; ! The ages;
S = 2; ! The value of parameter;
ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;
MIN = (1/(S * (S + 1))) * @SUM(SET : Y * ((Y/X)S - 1));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
    @SUM(SET(I)|I#LE#13 :
        (Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))2) <= .00003);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#15 :
    Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#14 :
    Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the additional constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =
    @SUM(SET(I)|I#LE#20 : Y(I)));

END

```

### Γ.3.3 Σύνολο Δεδομένων HK01F

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 4 constraints for the HK01F data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.011344 0.016704 0.013775 0.018219 0.022583 0.021531 0.035387 0.031236 0.036882  
0.039539 0.058961 0.028872 0.033420 0.073951 0.101383 0.077762 0.057996 0.114286 0.18465  
0.2; ! The ungraduated values;

L = 3747 3263 2759 2141 1638 1300 1046 800 569 379 237 173 120 68 39 26 17 9 5 5; !

The number of people at risk;

AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89; ! The ages;

S = 2; ! The value of parameter;

ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;

MIN = (1/(S \* (S + 1))) \* @SUM(SET : Y \* ((Y/X)<sup>S</sup> - 1));

! The graduated values should lie between 0 and 1;

@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;

@FOR(SET(K)|K#LE#1 :

@SUM(SET(I)|I#LE#17 :

(Y(I+3) - 3\*Y(I+2) + 3\*Y(I+1) - Y(I))<sup>2</sup> <= .000015);

!... the monotonicity constraint;

@FOR(SET(I)|I#LE#19 :

Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;

@FOR(SET(I)|I#LE#18 :

Y(I + 2) - 2 \* Y(I + 1) + Y(I) >= 0);

!... the additional constraint;

@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =

@SUM(SET(I)|I#LE#20 : Y(I));



END

#### Γ.4 Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με το εκτεταμένο μέτρο των Cressie and Read

Για να υπολογίσουμε τις εξομαλυμένες τιμές με το εκτεταμένο μέτρο των Cressie and Read, για τα τρία σύνολα δεδομένων, γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, τους παρακάτω κώδικες:

##### Γ.4.1 Σύνολο Δεδομένων L85

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 5 constraints for the L85 data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.03870 0.03909 0.04609 0.04839 0.06644 0.06091 0.07376 0.08019 0.08874 0.07455  
0.11195 0.12320 0.14031 0.14879 0.19020 0.21628 0.20718 0.28866 0.13362 0.38265; ! The  
ungraduated values;

L = 15012 11871 10002 8949 7751 6140 4718 3791 2806 2240 1715 1388 898 578 510 430  
362 291 232 196; ! The number of people at risk;

AGE = 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94; ! The ages;

S = 2; ! The value of parameter;

ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;

MIN = (1/(S \* (S + 1))) \* @SUM(SET : Y \* ((Y/X)<sup>S</sup> - 1));

! The graduated values should lie between 0 and 1;

@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;

@FOR(SET(K)|K#LE#1 :

```

@SUM(SET(I)|I#LE#17 :
      (Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))^2) <= .00004);
!... the monotonicity constraint;
@FOR(SET(I)|I#LE#19 :
      Y(I + 1) - Y(I) >= 0);
!... the convexity constraint;
@FOR(SET(I)|I#LE#18 :
      Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);
!... the first actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : L(I) * X(I)) =
      @SUM(SET(I)|I#LE#20 : L(I) * Y(I)));
!... the second actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * X(I)) =
      @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * Y(I)));
END

```

#### Γ.4.2 Σύνολο Δεδομένων HK01M

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 5 constraints for the HK01M data set;

SETS:

! There are 16;

SET / 1..16 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.0192308 0.0256304 0.0299252 0.035854 0.0390041 0.0353063 0.055409 0.0493601  
0.059126 0.0750853 0.0484581 0.116667 0.0681818 0.234568 0.117647 0.294118; ! The  
ungraduated values;

L = 12860 2419 2005 1534 1205 963 758 547 389 293 227 180 132 81 51 34; ! The number  
of people at risk;

AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85; ! The ages;

```

S = 2; ! The value of parameter;
ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;
MIN = (1/(S * (S + 1))) * @SUM(SET : Y * ((Y/X)S - 1));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0,Y,1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
    @SUM(SET(I)|I#LE#13 :
        (Y(I+3)-3*Y(I+2)+3*Y(I+1)-Y(I))2) <= .00003);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#15 :
    Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#14 :
    Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the first actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#16 : L(I) * X(I)) =
    @SUM(SET(I)|I#LE#16 : L(I) * Y(I));

!... the second actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#16 : AGE(I) * L(I) * X(I)) =
    @SUM(SET(I)|I#LE#16 : AGE(I) * L(I) * Y(I));

END

```

### Γ.4.3 Σύνολο Δεδομένων HK01F

MODEL:

! This model determines the graduated values by the minimization of the Cressie and Read power divergence with parameter S subject to 5 constraints for the HK01F data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

$X = 0.011344 \ 0.016704 \ 0.013775 \ 0.018219 \ 0.022583 \ 0.021531 \ 0.035387 \ 0.031236 \ 0.036882$   
 $0.039539 \ 0.058961 \ 0.028872 \ 0.033420 \ 0.073951 \ 0.101383 \ 0.077762 \ 0.057996 \ 0.114286 \ 0.18465$

0.2; ! The ungraduated values;

$L = 3747 \ 3263 \ 2759 \ 2141 \ 1638 \ 1300 \ 1046 \ 800 \ 569 \ 379 \ 237 \ 173 \ 120 \ 68 \ 39 \ 26 \ 17 \ 9 \ 5 \ 5$ ; !

The number of people at risk;

$AGE = 70 \ 71 \ 72 \ 73 \ 74 \ 75 \ 76 \ 77 \ 78 \ 79 \ 80 \ 81 \ 82 \ 83 \ 84 \ 85 \ 86 \ 87 \ 88 \ 89$ ; ! The ages;

$S = 2$ ; ! The value of parameter;

ENDDATA

! Minimize the Cressie and Read power divergence with parameter S;

$MIN = (1/(S * (S + 1))) * @SUM(SET : Y * ((Y/X)^S - 1));$

! The graduated values should lie between 0 and 1;

@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;

@FOR(SET(K)|K#LE#1 :

    @SUM(SET(I)|I#LE#17 :

$(Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))^2 <= .000015$ );

!... the monotonicity constraint;

@FOR(SET(I)|I#LE#19 :

$Y(I + 1) - Y(I) >= 0$ );

!... the convexity constraint;

@FOR(SET(I)|I#LE#18 :

$Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0$ );

!... the first actuarial constraint;

@FOR(SET : @SUM(SET(I)|I#LE#20 :  $L(I) * X(I)$ ) =

    @SUM(SET(I)|I#LE#20 :  $L(I) * Y(I)$ ));

!... the second actuarial constraint;

@FOR(SET : @SUM(SET(I)|I#LE#20 :  $AGE(I) * L(I) * X(I)$ ) =

    @SUM(SET(I)|I#LE#20 :  $AGE(I) * L(I) * Y(I)$ ));

END

## Γ.5 Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με τη διαφορά του Jensen και 5 περιορισμούς

Για να υπολογίσουμε τις εξομαλυμένες τιμές με τη διαφορά του Jensen και 5 περιορισμούς, για τα τρία σύνολα δεδομένων, γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, τους παρακάτω κώδικες:

### Γ.5.1 Σύνολο Δεδομένων L85

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 5 constraints for the L85 data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.03870 0.03909 0.04609 0.04839 0.06644 0.06091 0.07376 0.08019 0.08874 0.07455  
0.11195 0.12320 0.14031 0.14879 0.19020 0.21628 0.20718 0.28866 0.13362 0.38265; ! The  
ungraduated values;

L = 15012 11871 10002 8949 7751 6140 4718 3791 2806 2240 1715 1388 898 578 510 430  
362 291 232 196; ! The number of people at risk;

AGE = 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94; ! The ages;

S = 2; ! The value of parameter;

ENDDATA

! Minimize the Jensen's difference;

MIN = -@SUM(SET : 0.5 \* (X + Y) \* @LOG(0.5 \* (X + Y)))  
+ 0.5 \* (@SUM(SET : Y \* @LOG(Y)) + @SUM(SET : X \* @LOG(X)));

! The graduated values should lie between 0 and 1;

@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;

@FOR(SET(K)|K#LE#1 :

@SUM(SET(I)|I#LE#17 :

$$(Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))^2 \leq .00004;$$

!... the monotonicity constraint;

```
@FOR(SET(I)|I#LE#19 :
      Y(I + 1) - Y(I) >= 0);
```

!... the convexity constraint;

```
@FOR(SET(I)|I#LE#18 :
      Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);
```

!... the first actuarial constraint;

```
@FOR(SET : @SUM(SET(I)|I#LE#20 : L(I) * X(I)) =
      @SUM(SET(I)|I#LE#20 : L(I) * Y(I)));
```

!... the second actuarial constraint;

```
@FOR(SET : @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * X(I)) =
      @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * Y(I)));
```

END

## Γ.5.2 Σύνολο Δεδομένων HK01M

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 5 constraints for the HK01M data set;

SETS:

! There are 16;

SET / 1..16 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.0192308 0.0256304 0.0299252 0.035854 0.0390041 0.0353063 0.055409 0.0493601  
0.059126 0.0750853 0.0484581 0.116667 0.0681818 0.234568 0.117647 0.294118; ! The  
ungraduated values;

L = 12860 2419 2005 1534 1205 963 758 547 389 293 227 180 132 81 51 34; ! The number  
of people at risk;

AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85; ! The ages;

S = 2; ! The value of parameter;

ENDDATA

! Minimize the Jensen's difference;

$MIN = -@SUM(SET : 0.5 * (X + Y) * @LOG(0.5 * (X + Y)))$   
 $+ 0.5 * (@SUM(SET : Y * @LOG(Y)) + @SUM(SET : X * @LOG(X)));$

! The graduated values should lie between 0 and 1;

$@FOR(SET : @BND(0, Y, 1));$

! Subject to the smoothness constraint;

$@FOR(SET(K) | K \# LE \# 1 :$

$@SUM(SET(I) | I \# LE \# 13 :$

$(Y(I+3) - 3 * Y(I+2) + 3 * Y(I+1) - Y(I))^2 <= .00003);$

!... the monotonicity constraint;

$@FOR(SET(I) | I \# LE \# 15 :$

$Y(I + 1) - Y(I) >= 0);$

!... the convexity constraint;

$@FOR(SET(I) | I \# LE \# 14 :$

$Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);$

!... the first actuarial constraint;

$@FOR(SET : @SUM(SET(I) | I \# LE \# 16 : L(I) * X(I)) =$

$@SUM(SET(I) | I \# LE \# 16 : L(I) * Y(I));$

!... the second actuarial constraint;

$@FOR(SET : @SUM(SET(I) | I \# LE \# 16 : AGE(I) * L(I) * X(I)) =$

$@SUM(SET(I) | I \# LE \# 16 : AGE(I) * L(I) * Y(I));$

END

### Γ.5.3 Σύνολο Δεδομένων HK01F

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 5 constraints for the HK01F data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

$X = 0.011344 \ 0.016704 \ 0.013775 \ 0.018219 \ 0.022583 \ 0.021531 \ 0.035387 \ 0.031236 \ 0.036882$   
 $0.039539 \ 0.058961 \ 0.028872 \ 0.033420 \ 0.073951 \ 0.101383 \ 0.077762 \ 0.057996 \ 0.114286 \ 0.18465$

0.2; ! The ungraduated values;

$L = 3747 \ 3263 \ 2759 \ 2141 \ 1638 \ 1300 \ 1046 \ 800 \ 569 \ 379 \ 237 \ 173 \ 120 \ 68 \ 39 \ 26 \ 17 \ 9 \ 5 \ 5$ ; !

The number of people at risk;

$AGE = 70 \ 71 \ 72 \ 73 \ 74 \ 75 \ 76 \ 77 \ 78 \ 79 \ 80 \ 81 \ 82 \ 83 \ 84 \ 85 \ 86 \ 87 \ 88 \ 89$ ; ! The ages;

$S = 2$ ; ! The value of parameter;

ENDDATA

! Minimize the Jensen's difference;

$MIN = -@SUM(SET : 0.5 * (X + Y) * @LOG(0.5 * (X + Y)))$   
 $+ 0.5 * (@SUM(SET : Y * @LOG(Y)) + @SUM(SET : X * @LOG(X)));$

! The graduated values should lie between 0 and 1;

$@FOR(SET : @BND(0, Y, 1));$

! Subject to the smoothness constraint;

$@FOR(SET(K) | K \# LE \# 1 :$

$@SUM(SET(I) | I \# LE \# 17 :$

$(Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))^2 <= .000015);$

!... the monotonicity constraint;

$@FOR(SET(I) | I \# LE \# 19 :$

$Y(I+1) - Y(I) >= 0);$

!... the convexity constraint;

$@FOR(SET(I) | I \# LE \# 18 :$

$Y(I+2) - 2 * Y(I+1) + Y(I) >= 0);$

!... the first actuarial constraint;

$@FOR(SET : @SUM(SET(I) | I \# LE \# 20 : L(I) * X(I)) =$

$@SUM(SET(I) | I \# LE \# 20 : L(I) * Y(I));$

!... the second actuarial constraint;

$@FOR(SET : @SUM(SET(I) | I \# LE \# 20 : AGE(I) * L(I) * X(I)) =$

$@SUM(SET(I) | I \# LE \# 20 : AGE(I) * L(I) * Y(I));$

END



## Γ.6 Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με τη διαφορά του Jensen και 6 περιορισμούς

Για να υπολογίσουμε τις εξομαλυμένες τιμές με τη διαφορά του Jensen και 6 περιορισμούς, για τα τρία σύνολα δεδομένων, γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, τους παρακάτω κώδικες:

### Γ.6.1 Σύνολο Δεδομένων L85

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 6 constraints for the L85 data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.03870 0.03909 0.04609 0.04839 0.06644 0.06091 0.07376 0.08019 0.08874 0.07455  
0.11195 0.12320 0.14031 0.14879 0.19020 0.21628 0.20718 0.28866 0.13362 0.38265; ! The  
ungraduated values;

L = 15012 11871 10002 8949 7751 6140 4718 3791 2806 2240 1715 1388 898 578 510 430  
362 291 232 196; ! The number of people at risk;

AGE = 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94; ! The ages;

S = 2; ! The value of parameter;

ENDDATA

! Minimize the Jensen's difference;

MIN = -@SUM(SET : 0.5 \* (X + Y) \* @LOG(0.5 \* (X + Y)))  
+ 0.5 \* (@SUM(SET : Y \* @LOG(Y)) + @SUM(SET : X \* @LOG(X)));

! The graduated values should lie between 0 and 1;

@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;

@FOR(SET(K)|K#LE#1 :

@SUM(SET(I)|I#LE#17 :

$$(Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))^2 \leq .00004;$$

!... the monotonicity constraint;

```
@FOR(SET(I)|I#LE#19 :
      Y(I + 1) - Y(I) >= 0);
```

!... the convexity constraint;

```
@FOR(SET(I)|I#LE#18 :
      Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);
```

!... the first actuarial constraint;

```
@FOR(SET : @SUM(SET(I)|I#LE#20 : L(I) * X(I)) =
      @SUM(SET(I)|I#LE#20 : L(I) * Y(I));
```

!... the second actuarial constraint;

```
@FOR(SET : @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * X(I)) =
      @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * Y(I));
```

!... the additional constraint;

```
@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =
      @SUM(SET(I)|I#LE#20 : Y(I));
```

END

## Γ.6.2 Σύνολο Δεδομένων HK01M

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 6 constraints for the HK01M data set;

SETS:

! There are 16;

SET / 1..16 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.0192308 0.0256304 0.0299252 0.035854 0.0390041 0.0353063 0.055409 0.0493601  
0.059126 0.0750853 0.0484581 0.116667 0.0681818 0.234568 0.117647 0.294118; ! The  
ungraduated values;

L = 12860 2419 2005 1534 1205 963 758 547 389 293 227 180 132 81 51 34; ! The number  
of people at risk;

```

AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85; ! The ages;
S = 2; ! The value of parameter;
ENDDATA

! Minimize the Jensen's difference;
MIN = -@SUM(SET : 0.5 * (X + Y) * @LOG(0.5 * (X + Y)))
      +0.5 * (@SUM(SET : Y * @LOG(Y)) + @SUM(SET : X * @LOG(X)));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
      @SUM(SET(I)|I#LE#13 :
            (Y(I+3)-3*Y(I+2)+3*Y(I+1)-Y(I))^2) <= .00003);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#15 :
      Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#14 :
      Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the first actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#16 : L(I) * X(I)) =
      @SUM(SET(I)|I#LE#16 : L(I) * Y(I)));

!... the second actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#16 : AGE(I) * L(I) * X(I)) =
      @SUM(SET(I)|I#LE#16 : AGE(I) * L(I) * Y(I)));

!... the additional constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =
      @SUM(SET(I)|I#LE#20 : Y(I)));

END

```

### Γ.6.3 Σύνολο Δεδομένων HK01F

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 6 constraints for the HK01F data set;

SETS:

```

! There are 20;
SET / 1..20 /: X, ! ungraduated values;
           Y, ! graduated values;
           L, ! people at risk;
           AGE; ! ages;
ENDSETS

DATA:
X = 0.011344 0.016704 0.013775 0.018219 0.022583 0.021531 0.035387 0.031236 0.036882
0.039539 0.058961 0.028872 0.033420 0.073951 0.101383 0.077762 0.057996 0.114286 0.18465
0.2; ! The ungraduated values;
L = 3747 3263 2759 2141 1638 1300 1046 800 569 379 237 173 120 68 39 26 17 9 5 5; !
The number of people at risk;
AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89; ! The ages;
S = 2; ! The value of parameter;
ENDDATA

! Minimize the Jensen's difference;
MIN = -@SUM(SET : 0.5 * (X + Y) * @LOG(0.5 * (X + Y)))
      +0.5 * (@SUM(SET : Y * @LOG(Y)) + @SUM(SET : X * @LOG(X)));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
      @SUM(SET(I)|I#LE#17 :
            (Y(I+3)-3*Y(I+2)+3*Y(I+1)-Y(I))^2) <= .000015);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#19 :
      Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#18 :
      Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the first actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : L(I) * X(I)) =
          @SUM(SET(I)|I#LE#20 : L(I) * Y(I)));

!... the second actuarial constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * X(I)) =

```

```
@SUM(SET(I)|I#LE#20 : AGE(I) * L(I) * Y(I));
```

!... the additional constraint;

```
@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =  
@SUM(SET(I)|I#LE#20 : Y(I)));
```

END

## Γ.7 Κώδικες για τον υπολογισμό των εξομαλυμένων τιμών με τη διαφορά του Jensen και 4 περιορισμούς

Για να υπολογίσουμε τις εξομαλυμένες τιμές με τη διαφορά του Jensen και 4 περιορισμούς, για τα τρία σύνολα δεδομένων, γράψαμε στο πρόγραμμα μη γραμμικού προγραμματισμού LINGO ver. 10, τους παρακάτω κώδικες:

### Γ.7.1 Σύνολο Δεδομένων L85

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 6 constraints for the L85 data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.03870 0.03909 0.04609 0.04839 0.06644 0.06091 0.07376 0.08019 0.08874 0.07455  
0.11195 0.12320 0.14031 0.14879 0.19020 0.21628 0.20718 0.28866 0.13362 0.38265; ! The  
ungraduated values;

L = 15012 11871 10002 8949 7751 6140 4718 3791 2806 2240 1715 1388 898 578 510 430  
362 291 232 196; ! The number of people at risk;

AGE = 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94; ! The ages;

S = 2; ! The value of parameter;

ENDDATA

! Minimize the Jensen's difference;

```
MIN = -@SUM(SET : 0.5 * (X + Y) * @LOG(0.5 * (X + Y)))
```

```

+0.5 * (@SUM(SET : Y * @LOG(Y)) + @SUM(SET : X * @LOG(X)));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
    @SUM(SET(I)|I#LE#17 :
        (Y(I+3)-3*Y(I+2)+3*Y(I+1)-Y(I))^2) <= .00004);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#19 :
    Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#18 :
    Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the additional constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =
    @SUM(SET(I)|I#LE#20 : Y(I)));

END

```

## Γ.7.2 Σύνολο Δεδομένων HK01M

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 4 constraints for the HK01M data set;

SETS:

! There are 16;

SET / 1..16 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

X = 0.0192308 0.0256304 0.0299252 0.035854 0.0390041 0.0353063 0.055409 0.0493601  
0.059126 0.0750853 0.0484581 0.116667 0.0681818 0.234568 0.117647 0.294118; ! The  
ungraduated values;

L = 12860 2419 2005 1534 1205 963 758 547 389 293 227 180 132 81 51 34; ! The number  
of people at risk;

```

AGE = 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85; ! The ages;
S = 2; ! The value of parameter;
ENDDATA

! Minimize the Jensen's difference;
MIN = -@SUM(SET : 0.5 * (X + Y) * @LOG(0.5 * (X + Y)))
      +0.5 * (@SUM(SET : Y * @LOG(Y)) + @SUM(SET : X * @LOG(X)));

! The graduated values should lie between 0 and 1;
@FOR(SET : @BND(0, Y, 1));

! Subject to the smoothness constraint;
@FOR(SET(K)|K#LE#1 :
      @SUM(SET(I)|I#LE#13 :
            (Y(I+3) - 3*Y(I+2) + 3*Y(I+1) - Y(I))^2) <= .00003);

!... the monotonicity constraint;
@FOR(SET(I)|I#LE#15 :
      Y(I + 1) - Y(I) >= 0);

!... the convexity constraint;
@FOR(SET(I)|I#LE#14 :
      Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);

!... the additional constraint;
@FOR(SET : @SUM(SET(I)|I#LE#20 : X(I)) =
      @SUM(SET(I)|I#LE#20 : Y(I)));

END

```

### Γ.7.3 Σύνολο Δεδομένων HK01F

MODEL:

! This model determines the graduated values by the minimization of the Jensen's difference subject to 4 constraints for the HK01F data set;

SETS:

! There are 20;

SET / 1..20 /: X, ! ungraduated values;

Y, ! graduated values;

L, ! people at risk;

AGE; ! ages;

ENDSETS

DATA:

$X = 0.011344 \ 0.016704 \ 0.013775 \ 0.018219 \ 0.022583 \ 0.021531 \ 0.035387 \ 0.031236 \ 0.036882$   
 $0.039539 \ 0.058961 \ 0.028872 \ 0.033420 \ 0.073951 \ 0.101383 \ 0.077762 \ 0.057996 \ 0.114286 \ 0.18465$

0.2; ! The ungraduated values;

$L = 3747 \ 3263 \ 2759 \ 2141 \ 1638 \ 1300 \ 1046 \ 800 \ 569 \ 379 \ 237 \ 173 \ 120 \ 68 \ 39 \ 26 \ 17 \ 9 \ 5 \ 5$ ; !

The number of people at risk;

$AGE = 70 \ 71 \ 72 \ 73 \ 74 \ 75 \ 76 \ 77 \ 78 \ 79 \ 80 \ 81 \ 82 \ 83 \ 84 \ 85 \ 86 \ 87 \ 88 \ 89$ ; ! The ages;

$S = 2$ ; ! The value of parameter;

ENDDATA

! Minimize the Jensen's difference;

$MIN = -@SUM(SET : 0.5 * (X + Y) * @LOG(0.5 * (X + Y)))$   
 $+ 0.5 * (@SUM(SET : Y * @LOG(Y)) + @SUM(SET : X * @LOG(X)));$

! The graduated values should lie between 0 and 1;

$@FOR(SET : @BND(0, Y, 1));$

! Subject to the smoothness constraint;

$@FOR(SET(K) | K \# LE \# 1 :$

$@SUM(SET(I) | I \# LE \# 17 :$

$(Y(I+3) - 3 * Y(I+2) + 3 * Y(I+1) - Y(I))^2 <= .000015);$

!... the monotonicity constraint;

$@FOR(SET(I) | I \# LE \# 19 :$

$Y(I + 1) - Y(I) >= 0);$

!... the convexity constraint;

$@FOR(SET(I) | I \# LE \# 18 :$

$Y(I + 2) - 2 * Y(I + 1) + Y(I) >= 0);$

!... the additional constraint;

$@FOR(SET : @SUM(SET(I) | I \# LE \# 20 : X(I)) =$

$@SUM(SET(I) | I \# LE \# 20 : Y(I)));$

END