Π Α Ν Ε Π Ι Σ Τ Η Μ Ι Ο Π Ε Ι Ρ Α Ι Ω Σ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ **ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

ΗΜΙΠΑΡΑΜΕΤΡΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΓΙΑ ΜΟΝΤΕΛΑ ΕΠΙΤΑΧΥΝΟΜΕΝΩΝ ΧΡΟΝΩΝ ΖΩΗΣ

Ειρήνη Σταματοπούλου

Διπλωματική Εργασία
που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς
Σεπτέμβριος 2016

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

Σχολή Χρηματοοικονομικής και Στατιστικής

Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ
ΣΤΗΝ **ΕΦΑΡΜΟΣΜΕΝΗ ΣΤΑΤΙΣΤΙΚΗ**

ΗΜΙΠΑΡΑΜΕΤΡΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ ΓΙΑ ΜΟΝΤΕΛΑ ΕΠΙΤΑΧΥΝΟΜΕΝΩΝ ΧΡΟΝΩΝ ΖΩΗΣ

Ειρήνη Σταματοπούλου

Διπλωματική Εργασία

που υποβλήθηκε στο Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς ως μέρος των απαιτήσεων για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στην *Εφαρμοσμένη Στατιστική*

Πειραιάς
Σεπτέμβριος 2016

Η παρούσα Διπλωματική Εργασία εγκρίθηκε ομόφωνα από την Τριμελή Εξεταστική Επιτροπή που ορίσθηκε από τη ΓΣΕΣ του Τμήματος Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς στην υπ’ αριθμ. …….. συνεδρίασή του σύμφωνα με τον Εσωτερικό Κανονισμό Λειτουργίας του Προγράμματος Μεταπτυχιακών Σπουδών στην Εφαρμοσμένη Στατιστική

Τα μέλη της Επιτροπής ήταν:
- Καθηγητής, Ηλιόπουλος Γεώργιος (Επιβλέπων)
- Αναπληρωτής Καθηγητής, Αντζουλάκος Δημήτριος
- Επίκουρος Καθηγητής, Μπούτσικας Μιχαήλ

Η έγκριση της Διπλωματική Εργασίας από το Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης του Πανεπιστημίου Πειραιώς δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα.

UNIVERSITY OF PIRAEUS

School of Finance and Statistics

Department of Statistics and Insurance Science

POSTGRADUATE PROGRAM IN
APPLIED STATISTICS

SEMIPARAMETRIC INFERENCE FOR ACCELERATED LIFE TEST MODELS

By

Eirini Stamatopoulou

MSc Dissertation

submitted to the Department of Statistics and Insurance Science of the University of Piraeus in partial fulfilment of the requirements for the degree of Master of Science in Applied Statistics

Piraeus, Greece
September 2016

Τίποτα δεν είναι εύκολο ή δύσκολο.
Όλα είναι γνωστά ή άγνωστα.

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω από καρδιάς τον επιβλέποντα καθηγητή μου, κο Γ. Ηλιόπουλο για την καθοδήγησή του, την πολύτιμη βοήθειά του και την υπομονή του κατά τη διάρκεια της εκπόνησης της μεταπτυχιακής μου διατριβής. Μέσα από αυτήν την συνεργασία, με ενέπνευσε και μου προσέφερε πολλές ευκαιρίες και συμβουλές, καθοριστικές για την συγγραφή αυτής της εργασίας αλλά και γενικά της πορείας μου. Τον ευχαριστώ για τις γνώσεις που μου μετέφερε, για τον τρόπο διδασκαλίας του, για την θέληση και όρεξη που έδειξε και που πίστεψε σε μένα.

Επίσης, θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής, τον Αναπληρωτή Καθηγητή κ. Δ. Αντζουλάκο και τον Επίκουρο Καθηγητή κ. Μ. Μπούτσικα για τη συμμετοχή τους στην επιτροπή καθώς και για όλα αυτά που έμαθα από εκείνους στα πλαίσια του μεταπτυχιακού αυτού προγράμματος.

Δεν θα μπορούσα να παραβλέψω τον Επίκουρο Καθηγητή κο Σ. Μπερσίμη, για την πολύτιμη βοήθειά του κατά την διάρκεια των σπουδών μου και τις ευκαιρίες που μου έδωσε.

Παράλληλα, θα ήθελα να ευχαριστήσω τον φίλο και συμφοιτητή μου Νάσο Παλαιοδημόπουλο για την πολύτιμη βοήθειά του, τα εποικοδομητικά του σχόλια αλλά και την στήριξή του, όλο αυτό το διάστημα.

Ολοκληρώνοντας, θα ήθελα να αναφερθώ και να ευχαριστήσω θερμά την οικογένειά μου για δύναμη που μου έδωσε προκειμένου να ολοκληρώσω αυτήν την προσπάθεια και που στάθηκε δίπλα μου από την αρχή.

Περίληψη

Τα μοντέλα επιταχυνόμενων χρόνων ζωής χρησιμοποιούνται προκειμένου να εκτιμήσουμε κατανομές ζωής συστημάτων συντομότερα, χωρίς να περιμένουμε να παρατηρήσουμε τους χρόνους ζωής τους υπό κανονικές συνθήκες λειτουργίας. Τα μοντέλα αυτά εμφανίσθηκαν εδώ και πολλές δεκαετίες και εφαρμόζονται συχνότερα στους κλάδους της βιοστατιστικής και της μηχανολογίας. Ο σκοπός αυτής της διπλωματικής εργασίας είναι να εισαγάγουμε τον αναγνώστη σε τέτοια μοντέλα και να εφαρμόσουμε έναν ημιπαραμετρικό τρόπο εκτίμησης της κατανομής του χρόνου ζωής.

Στο πρώτο κεφάλαιο ξεκινάμε με την περιγραφή των μοντέλων επιταχυνόμενων χρόνων ζωής σχολιάζοντας την χρησιμότητά τους και συνεχίζουμε με τον ορισμό βασικών εννοιών που χρησιμοποιούνται στην εργασία.

Στο δεύτερο κεφάλαιο παρουσιάζουμε το μοντέλο λόγου πυκνοτήτων που εισήγαγαν οι Fokianos et al. (2001) βασιζόμενοι στις ιδέες των Qin and Lawless (1994) και εξηγούμε λεπτομερώς πώς γίνεται η εκτίμηση των παραμέτρων του μέσω κατάτμησης της (μη παραμετρικής) πιθανοφάνειας. Βασιζόμενοι σε αυτές εκτιμούμε την άγνωστη συνάρτηση κατανομής. Η μέθοδος εφαρμόζεται σε ένα παράδειγμα με πραγματικά δεδομένα.

Στο τρίτο κεφάλαιο χρησιμοποιούμε το μοντέλο λόγου πυκνοτήτων μοντελοποιώντας περαιτέρω τις παραμέτρους του. Το απλούστερο μοντέλο συγκρίνεται με το βασικό μέσω ασυμπτωτικού ελέγχου και ελέγχου που βασίζεται στη μέθοδο bootstrap.

Στο τέταρτο κεφάλαιο θεωρούμε το μοντέλο λόγου πυκνοτήτων στην περίπτωση που τα δεδομένα υπόκεινται σε (πιθανή) τυχαία λογοκρισία. Για την εκτίμηση των παραμέτρων σε αυτήν την περίπτωση βασιζόμαστε σε έναν αλγόριθμο EM που πρότειναν πρόσφατα οι Wei and Zhou (2016).

Για την εκτίμηση των μοντέλων και γενικότερα για όλους τους υπολογισμούς χρησιμοποιήθηκε το πρόγραμμα R και πιο συγκεκριμένα η συνάρτηση optim. Οι συναρτήσεις που δίνουν τον λογάριθμο της πιθανοφάνειας και των παραγώγων της παρουσιάζονται στο τέταρτο κεφάλαιο.

Η διπλωματική εργασία ολοκληρώνεται με καταγραφή των ερωτημάτων και σκέψεων που γεννήθηκαν κατά την εκπόνησή της καθώς και ιδεών για μελλοντική ερευνητική εργασία.

Abstract

Accelerated life time models are used in order to faster estimate life distributions of systems without waiting to observe their lifetimes under normal operating conditions. These models appeared many decades ago and are applied mainly in the areas of biostatistics and mechanical engineering. The purpose of this dissertation is to introduce the reader to this kind of models and to apply a semiparametric method of estimating the life time distribution.

In the first chapter we start with a description of accelerated life test models by commenting on their usefulness and we continue with basic definitions of notions which are used in the dissertation.

In the second chapter we present the density ratio model introduced by Fokianos et al. (2001) based on ideas of Qin and Lawless (1994) and we explain in detail the estimation process of its parameters via profiling the (nonparametric) likelihood. Based on the estimates of the parameters we estimate the unknown distribution function. The method is applied to a real data example.

In the third chapter we discuss the density ratio model by further modeling its parameters. The simpler model is compared to the basic one asymptotically and using the bootstrap method.

In the fourth chapter we consider the density ratio model in the case where the data are subject to (possible) random censoring. In this case our estimation procedures are based on a particular EM algorithm recently proposed by Wei and Zhou (2016).

For the estimation of the models and all other calculations we used R and more specifically the function optim. The procedures which return the loglikelihood and its derivatives are presented in the fourth chapter.

We conclude by discussing some questions and thoughts as well some ideas for future work.

Contents

Chapter 1

Introduction

1.1 Accelerated life test models

The revolution of technology and mechanism demands the most accurate and precise estimation of reliability of a product or a system. The classical methods of estimation emphasize in the analysis of lifetime data under normal operation conditions, something that in most of the times is very difficult or even unreachable. This problem mostly relies on the long life of the products so, in order to obtain an estimation of the lifetime distribution of a product in a timely manner, we use Accelerated Life Tests (ALT). The ALT models have been studied for many decades and one of the first researchers in this area was Nelson (1980, 2004).

1.1.1 How an ALT model works

The experimenter applies severe stresses on the parameters of the lifetime distributions to obtain information more quickly than would be possible if the products were under the normal operating conditions. Our purpose is to induce earlier failure because if we depended on normal operating conditions a very large sample and of course a lot of time would be necessary and most of the times this is not feasible. Engineers have used accelerated tests for diverse products as well as this type of stress has been used for material testing (metals, plastic, ceramics, food and drugs, cement, nuclear reactor materials). We will give only one example just to clarify how such a model works. Imagine that we have metal in our disposal. Accelerating stresses include mechanical stress, temperature, specimen geometry and surface finish. Chemical acceleration factors include salt, humidity, corrosives and acids.

As mentioned above, an ALT is a method for estimating the reliability of products at normal operating conditions from the failure data obtained at the severe conditions. Nowadays, such approach is used in many science fields such as biology, demography, gerontology, dynamic of population, genetics, reliability, survival analysis, radio-biology, biophysics and so many other fields where scientists study the longevity, aging and degradation using stochastic models. In order to obtain information about all these, reliability theory seems to be a great method combined with parametric, non parametric and semi parametric accelerated life models used in accelerating life testing.

1.1.2 Types of ALT models

There are different types of ALT models as there are different types of stresses that are applied in many ways. ALT methods can be either qualitative or quantitative. Qualitative methods are used when we have a small sample and we are interested in finding possible failure mode in the product. If there is one, managers and engineers try to improve the product design. Nevertheless, a failure mode does not imply anything about the reliability status of the product. Some well known and relatively new qualitative methods are:

- Highly Accelerated Life Testing (HALT)
- Highly Accelerated Stress Screen (HASS)
- Elephant and Torture Tests
- Shake and Bake Tests
- Reliability Demonstration Test (RDT)

- Reliability Acceptance Test
- Burn-In Test
- Reliability Growth Test (RGT)

From the above we will say a few words for the most common of these methods starting with the "Elephant tests". In the bibliography one could see that this specific method has a lot of names such as torture test, killer test, design limit/margin test. Such a test steps on the product with an elephant, figuratively speaking. If the item/product survives, it passes the test otherwise the engineer must redesign or improve the product. The test generally involves one specimen which will be subject to a single (or more) severe level of stress. Most of the time, different elephant tests may be used to reveal different failure modes. Apart from "Elephant tests", we have the "Burn-In" test. As Nelson mentions, such a test consists of running units under design or accelerated conditions for a suitable length of time. Burn-in is a manufacturing operation that is intended to fail short-lived units (freaks) so if the burn-in test works, the surviving units that go into service have a few early failures. These are the units that have manufacturing defects.

On the other hand, quantitative methods focus in the analysis of what we cited above, namely, test units are exposed to a greater pressure than the predefined in order to increase the likelihood of observing failures. A Quantitative Accelerated Life Test model (QALT) actually tries to quantify the life of the product and provide numerical estimations of reliability characteristics and the statistical inference for the functioning of the product under normal use conditions. Generally an ALT model is used during the design phase and before the release of a product. If we consider the industrial sector and if we think that ALT models were introduced in order to ensure the reliability of electronic products that have been used in the military, aerospace, automotive and mobile applications, we can see that the model tries to improve product reliability efficiently and to determine whether safety and reliability goals are being met in order to assess warranty risk.

At QALT we can distinguish different stress strategies like:

- Parallel Constant-Stress ALT (PC-SALT)
- Step-Stress ALT (SSALT)
- Accelerated Failure Time Model (AFT)
- Cumulative Exposure Model (CEM)

In PC-SALT test units are divided into groups and each group is tested under different stress level. At SSALT all test units are exposed in a pre-specified stress level. For those test units that survive up to a pre-specified time, the stress level is changed and held constant until all units fail (uncensored test) or until a pre-specified test termination time is attained (censored test). The most widely used class of ALT model is AFT models. The stress levels applied at the accelerated conditions are within a range of true acceleration and the AFT models assume that the applied stresses act multiplicatively on the failure time or linearly on the log failure time, rather than multiplicatively on the hazard rate. Finally, the CEM relates the lifetime distribution of the test units at one stress level to the distributions at previous levels assuming that the remaining lifetime depends only on the current stress level and on the current cumulative fraction of the units that failured (Markov property).

1.1.3 Types of Accelerated Test Data

Accelerated test data can be divided into two big categories. The first one is when the product characteristic of interest is ”life” and the second one when the measure of interest is

”performance”. A measure of performance could be tensile, strength or ductility. In such a case we are interested in how product performance degrades with age. Namely, specimens are aged under high stress and their performances, i.e. strength, measured at different ages. A very well known example of this method occurs temperature aging of electrical insulation and pharmaceuticals. For life data the analysis relies on the data that we have. First of all, we might have complete data (failure age) of each sample unit. However, it is very common to have censored data, namely some units are unfailed and their failure times are known only to be beyond their present running times (censored on the right or truncated). The analysis could be made with data when a failure time known only to be before a certain time (censored on the left). It is known from literature that speaking about censored data we have Type I and Type II censored data, multiply censored data, etc. Regarding accelerated data we have also competing modes when sample units fail from different causes. Also, one may know whether the failure time of a unit is before or after a certain time so we have to deal with quantal-response data. Finally, when a failure is inspected more than once and a unit failed in an interval between inspections we have interval data. Of course, in an analysis one could notice a mixture of all the above cases and so we could speak about mixture data.

1.2 Empirical likelihood

First of all we have to define the empirical cumulative distribution function and show that it is a nonparametric maximum likelihood estimator (NPMLE) of the underlying CDF. Let X be a real- (or integer-) valued random variable and let $F(x) = P(X \leq x)$, $x \in \mathbb{R}$, denote its CDF. Then, $F(x-) = \lim_{y \uparrow x} F(y) = P(X < x)$ and $F(x) - F(x-) = P(X = x)$.

Definition 1. Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} F$. The empirical cumulative distribution function

(ECDF) of $X_1, X_2, \dots, X_n$ is the function

$$\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x), \quad x \in \mathbb{R}.$$

It is well-known that, given the observed values $x_1, \dots, x_n$, the ECDF is the CDF of the (discrete) distribution which assigns mass $1/n$ to each of the x_i's, $i = 1, \dots, n$.

Definition 2. Let $X_1, X_2, \dots, X_n \overset{\text{iid}}{\sim} F$. Then, the empirical likelihood of F is

$$L(F) = \prod_{i=1}^{n}\{F(x_i) - F(x_i-)\}, \quad F \in \mathscr{F}, \tag{1.1}$$

where $\mathscr{F}$ is the set of all CDFs.

The empirical likelihood (see Owen, 2001) can be considered as the nonparametric likelihood of a CDF. It is immediately seen that $L(F) = 0$ when F is continuous. This implies that in order to maximize $L(F)$ with respect to F, i.e., to find the nonparametric maximum likelihood estimator (NPMLE) of F, we must restrict the parameter space to the set of discrete distributions and, furthermore, to the set of discrete distributions that assign positive mass to the observed values. It turns out that

Theorem 1. *The NPMLE of F is the ECDF $\widehat{F}$.*

Proof. Clearly, if F is continuous at some observed value x_i, then $F(x_i) - F(x_i-) = 0$ and so, $L(F) = 0$. Hence, the distribution maximizing (1.1) must assign positive probabilities $p_1, \dots, p_n$ to the observed values $x_1, \dots, x_n$. These probabilities must satisfy $\sum_{k=1}^{n} p_k = 1$ because if their sum is smaller than 1 we have

$$\prod_{i=1}^{n} p_i < \prod_{i=1}^{n} \frac{p_i}{\sum p_k}$$

and thus, distribution which assigns mass $p_i' = p_i / \sum p_k$ to x_i, $i = 1, \dots, n$, has larger likelihood. By symmetry we must have $p_1 = \cdots = p_n$, therefore $p_i = 1/n$, $i = 1, \dots, n$. $\square$

Kiefer and Wolfowitz (1956) were the first who noticed that the ECDF is an NPMLE. In the same year, Grenarder (1956) constructed a NPMLE for a distribution known to have a monotone decreasing density over $[0, \infty)$. Kaplan and Meier (1958) used the idea of the NPMLE and introduced the product-limit estimator of the CDF based on censored data. Johansen (1978) showed that product-limit estimators are NPMLE's for transition probabilities of continuous time Markov chains. Finally, Hartley and Rao (1968) introduced the construction of the NPMLE in the context of random sampling from a finite population. It is also worth mentioning that Brunk, Franck, Hanson, and Hogg (1966) considered nonparametric maximum likelihood estimation of the CDFs of two stochastically ordered distributions.

1.3 Biased (or weighted) distributions

There are situations where the researcher wants to estimate a CDF but the data are sampled from a biased, or weighted, version of the distribution of interest. As Patil (2002) notes,

> "weighted distributions adjust the probabilities of actual occurrence of events at a specification of the probabilities of those events have already been observed".

The basic idea of weighted distributions is based on Fisher's theory (1934) and an extension of this idea was given by Rao (1965).

Suppose that a distribution has CDF F_0 and let w be a nonnegative function. Then the distribution with CDF

$$F(x) = \frac{\int_{-\infty}^{x} w(y) \mathrm{d}F_0(y)}{\int_{-\infty}^{\infty} w(y) \mathrm{d}F_0(y)}, \tag{1.2}$$

where $E_{F_0}\{w(X)\} = \int w(y)\mathrm{d}F_0(y) \in (0, \infty)$, is called a biased (or weighted) version of F_0 with biasing (or weighting) function w. If F_0 admits a probability density (or mass) function

(PDF) f_0, the corresponding PDF of its biased version is

$$f(x) = \frac{w(x) f_0(x)}{E_0\{w(X)\}}.$$

When the support of F_0 is a subset of the nonnegative reals, the case $w(x) = x$ corresponds to its so-called length biased version. In this case,

$$F(x) = \frac{1}{E_{F_0}(X)} \int_{-\infty}^{x} y \, \mathrm{d}F_0(y),$$

provided $E_{F_0}(X) < \infty$. The function $w(x) = x^2$ gives rise to the size biased version of F_0 while $w(x) = I(x \in A)$ truncates F_0 at the set A.

In general, weighted distributions may serve as a useful tool for the selection of appropriate models for observed data drawn without a proper frame. The statistical problem which rise underneath is to determine a suitable weight function $w(x)$. Below are some of the most common weighted functions used in the literature (Liang, 2005):

- $w(x) = x^{\alpha}$
- $w(x) = x \cdot (x-1) \cdot \ldots \cdot (x-r)$, where r is an integer
- $w(x) = e^{\alpha + \beta x}$
- $w(x) = \alpha + \beta \cdot x$
- $w(x) = 1 - (1-\beta)^x$
- $w(x) = \dfrac{\alpha + \beta x}{\gamma + \delta x}$
- $w(x) = P(Z \leq x)$, for some RV Z
- $w(x) = 1 - F_0(x)$

Nonparametric maximum likelihood estimation of a CDF based on a length biased sample from it has been first considered by Cox (1969). If $x_1, \ldots, x_n$ denote the observed values of $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} F$, then the NPMLE of the baseline CDF F is given in closed form, namely,

$$\widehat{F}(x) = \frac{\sum_{i=1}^n x_i^{-1} I(x_i \le x)}{\sum_{i=1}^n x_i^{-1}} = \frac{\sum_{x_i \le x} x_i^{-1}}{\sum_{i=1}^n x_i^{-1}}.$$

In general, when the biasing function is w, it can be shown that the NPMLE of F is

$$\widehat{F}(x) = \frac{\sum_{i=1}^n w(x_i)^{-1} I(x_i \le t)}{\sum_{i=1}^n w(x_i)^{-1}} = \frac{\sum_{x_i \le x} w(x_i)^{-1}}{\sum_{i=1}^n w(x_i)^{-1}}.$$

The problem of estimating a CDF based on samples from two or more of its weighted versions has been considered by Vardi (1982), Vardi (1985), Vardi et al. (1988). In these papers, conditions for the existence and uniqueness of the NPMLEs are discussed as well as the asymptotic distributions of the NPMLEs of the expectations of the weighting functions and of the baseline CDF. More specifically, under reasonable conditions, the NPMLEs of the expectations are asymptotically normal while the NPMLE of the baseline CDF converges weakly to a pinned Gaussian process.

1.4 Semiparametric extensions

The procedures mentioned at the end of the previous section assume the function(s) w to be completely known. A first semiparametric approach introduced by Sun and Woodroofe (1997) who estimated the weight function w nonparametrically and the CDF F parametrically. Gilbert et al. (1999) introduced another semiparametric model, leaving F nonparametric but assuming that the weight function of the ith sample is $w(\cdot, \beta)$, where β is some finite dimensional parameter and the sampling density of the ith ample is proportional to $w(\cdot, \beta) f(\cdot)$. In more detail,

$$F(x; \beta) = \frac{\int_{-\infty}^{x} w(y; \beta) \mathrm{d}F_0(y)}{\int_{-\infty}^{\infty} w(y; \beta) \mathrm{d}F_0(y)} \tag{1.3}$$

or, in terms of densities,

$$f(x;\beta) = \frac{w(x;\beta)f_0(x)}{E_0\{w(X;\beta)\}}. \tag{1.4}$$

Note that for fixed β, (1.3) is a particular case of (1.2). Equation (1.4) defines a semiparametric model: the finite dimensional parameter β is its parametric while the infinite dimensional parameter F_0 is its nonparametric part.

Notice that (1.4) defines a *density ratio model* since it actually models the ratio of two densities:

$$\frac{f(x;\beta)}{f_0(x)} \propto w(x;\beta).$$

In this dissertation we will follow Fokianos et al. (2001) and set

$$w(x;\beta) = e^{\beta h(x)},$$

where $h(x)$ is an arbitrary yet known function.

Chapter 2

The density ratio model: A tool for semiparametric comparison of distributions

2.1 Intoduction to the density ratio model

In order to avoid the normality assumption in the one-way layout, Fokianos et al. (2001) considered and studied in detail a semiparametric model which now is known as *the* density ratio model. Suppose that we observe $K+1$ independent random samples from distributions with densities $f_0, f_1, \ldots, f_K$, where f_0 is completely unspecified and

$$f_i(x) = \exp\{\alpha_i + \beta_i h(x)\} f_0(x), \quad i = 1, \ldots, K, \tag{2.1}$$

for some unknown parameters α_i and β_i and given function h. The term "density ratio model" is justified by the fact that (2.1) models the ratio of the densities f_i/f_0 as

$$\frac{f_i(x)}{f_0(x)} = \exp\{\alpha_i + \beta_i h(x)\}, \quad x \in \mathrm{supp}(f_0),$$

or, equivalently,

$$\log \frac{f_i(x)}{f_0(x)} = \alpha_i + \beta_i h(x), \quad x \in \mathrm{supp}(f_0).$$

Clearly, under this model the distributions f_i, $i = 1, \ldots, K$, are biased (or weighted) versions of f_0. It is a semiparametric model since it consists of a nonparametric part, the baseline distribution with PDF f_0, and the real-valued parameters $\alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_K$.

The fact that f_i is a PDF implies that model (2.1) is overparametrized since α_i is completely determined by β_i and f_0 through the equation

$$\alpha_i = -\log\left\{\int e^{\beta_i h(x)} \mathrm{d}F_0(x)\right\}.$$

This implies that when $\beta_i = 0$ we have $\alpha_i = 0$ as well and thus $f_i = f_0$. Based on this, Fokianos et al. (2001) were able to express the hypothesis of equality of distributions

$$f_0 = f_1 = \cdots = f_K$$

as

$$0 = \beta_1 = \cdots = \beta_K.$$

Therefore, under the density ratio model (2.1), the problem of testing equality of $K+1$ distributions simplifies to testing that K real-valued parameters equal zero.

The DR model fits nicely in the context of one-parameter exponential families (see Table 2.1). In fact, by definition, the DR model assumes that the $K+1$ distributions belong to the same exponential family with canonical parameter β and sufficient statistic $h(X)$; see (2.1).

2.2 Empirical likelihood and its maximization

Assume that we observe $K+1$ independent groups

$$\begin{aligned}
&X_{01}, \ldots, X_{0n_0} \overset{\text{iid}}{\sim} f_0(x), \\
&X_{11}, \ldots, X_{1n_1} \overset{\text{iid}}{\sim} f_0(x)\exp\{\alpha_1 + \beta_1 h(x)\}, \\
&\qquad\vdots \\
&X_{K1}, \ldots, X_{Kn_K} \overset{\text{iid}}{\sim} f_0(x)\exp\{\alpha_K + \beta_K h(x)\}.
\end{aligned} \tag{2.2}$$

Distribution	$h(x)$	β
Binomial $\mathcal{B}(n,p)$	x	$\log\{p/(1-p)\}$
Poisson $\mathcal{P}(\lambda)$	x	$\log\lambda$
Negative Binomial $\mathcal{NB}(a,p)$ (a known)	x	$\log(1-p)$
Exponential $\mathcal{E}(\lambda)$	x	$-\lambda$
Gamma $\mathcal{G}(a,\lambda)$ (a known)	x	$-\lambda$
Gamma $\mathcal{G}(a,\lambda)$ (λ known)	$\log x$	a
Normal $\mathcal{N}(\mu,\sigma^2)$ (σ^2 known)	x	μ/σ^2
Normal $\mathcal{N}(\mu,\sigma^2)$ (μ known)	$(x-\mu)^2$	$-1/(2\sigma^2)$
Lognormal $\mathcal{LN}(\mu,\sigma^2)$	$\log x$	μ/σ^2
Lognormal $\mathcal{LN}(\mu,\sigma^2)$	$(\log x-\mu)^2$	$-1/(2\sigma^2)$
Inverse Gaussian $\mathcal{IG}(\mu,\lambda)$ (λ known)	x	$-\lambda/(2\mu^2)$
Inverse Gaussian $\mathcal{IG}(\mu,\lambda)$ (μ known)	$1/x$	$-\lambda/2$
Beta $\mathcal{B}eta(a,b)$ (b known)	$\log x$	a
Beta $\mathcal{B}eta(a,b)$ (a known)	$\log(1-x)$	b

Table 2.1: Function h and parameter β for several one-parameter exponential families.

For convenience, hereafter we set $\alpha_0 = \beta_0 = 0$ so that we will able to write $f_i(x) = \exp\{\alpha_i + \beta_i h(x)\} f_0(x)$ for all $i = 0, 1, \ldots, K$. Moreover, we define $\widehat{\alpha}_0 = \widehat{\beta}_0 = 0$ in order to be able to write more elegant formulas containing estimators of α's and β's.

The empirical likelihood of $F_0, \alpha = (\alpha_1, \ldots, \alpha_K), \beta = (\beta_1, \ldots, \beta_K)$ is

$$\begin{aligned} L(\alpha,\beta,F_0) &= \prod_{i=0}^{K}\prod_{j=1}^{n_i} \mathrm{d}F_i(x_{ij}) \\ &= \prod_{i=0}^{K}\prod_{j=1}^{n_i} \mathrm{d}F_0(x_{ij}) \exp\{\alpha_i + \beta_i h(x_{ij})\} \\ &= \exp\left\{ \sum_{i=1}^{K} n_i\alpha_i + \sum_{i=1}^{K} \beta_i \sum_{j=1}^{n_i} h(x_{ij}) \right\} \prod_{i=0}^{K}\prod_{j=1}^{n_i} \mathrm{d}F_0(x_{ij}), \end{aligned}$$

and the parameters satisfy the constraints $\int e^{\alpha_i+\beta_i h(x)} \mathrm{d}F_0(x) = 1$, $i = 0, 1, \ldots, K$. As usual, in the context of empirical likelihood, the maximization of L with respect to F_0 is achieved within the set of discrete CDFs with jumps at the observed values. Let $N = \sum_{i=0}^{K} n_i$ and $z_1, \ldots, z_N$ be the combined (or pooled) sample. Then, the maximization of L with respect to α, β and F_0 under the aforementioned constraints is equivalent to the maximization of

$$L^*(\alpha, \beta, p) = \exp\left\{ \sum_{i=1}^{K} n_i \alpha_i + \sum_{i=1}^{K} \beta_i \sum_{j=1}^{n_i} h(x_{ij}) \right\} \prod_{k=1}^{N} p_k,$$

where

$$p_k = \mathrm{d}F_0(z_k) = F_0(z_k) - F_0(z_k-), \quad k = 1, \ldots, N,$$

and $p = (p_1, \ldots, p_N)$, under the constraints

$$\sum_{k=1}^{N} p_k e^{\alpha_i + \beta_i h(z_k)} = 1, \quad i = 0, 1, \ldots, K, \tag{2.3}$$

or, equivalently, of its logarithm

$$\ell(\alpha, \beta, p) \equiv \log L^*(\alpha, \beta, p) = \sum_{i=1}^{K} n_i \alpha_i + \sum_{i=1}^{K} \beta_i \sum_{j=1}^{n_i} h(x_{ij}) + \sum_{k=1}^{N} \log p_k. \tag{2.4}$$

Such maximization tasks are typically done with the use of Lagrange multipliers. Here we will denote them by $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_K)$ (since there are $K+1$ constraints). The corresponding Lagrangian is

$$\begin{aligned} \mathcal{L}(\alpha, \beta, p, \lambda) = &\sum_{k=1}^{N} \log p_k + \sum_{i=1}^{K} n_i \alpha_i + \sum_{i=1}^{K} \beta_i \sum_{j=1}^{n_i} h(x_{ij}) + \\ &\lambda_0 \left(\sum_{k=1}^{N} p_k - 1 \right) + \sum_{k=1}^{K} \lambda_i \left(\sum_{k=1}^{N} p_k e^{\alpha_i + \beta_i h(z_k)} - 1 \right). \end{aligned}$$

The first partial derivatives of $\mathcal{L}$ with respect to α_i's are

$$\frac{\partial}{\partial \alpha_i} \mathcal{L}(\alpha, \beta, p, \lambda) = n_i + \lambda_i \sum_{k=1}^{N} p_k e^{\alpha_i + \beta_i h(z_k)}, \quad i = 1, \ldots, K.$$

By equating them to zero and using the constraints $\sum_{k=1}^{N} p_k e^{\alpha_i + \beta_i h(z_k)} = 1$, $i = 1, \ldots, K$, we get explicitly the Lagrange multipliers $\lambda_1, \ldots, \lambda_K$ as

$$\lambda_i = -n_i, \quad i = 1, \ldots, K. \tag{2.5}$$

Furthermore, the first partial derivative of $\mathcal{L}$ with respect to p_k is

$$\frac{\partial}{\partial p_k}\mathcal{L}(\alpha,\beta,p,\lambda) = \frac{1}{p_k} + \lambda_0 + \sum_{i=1}^{K}\lambda_i e^{\alpha_i+\beta_i h(z_k)}. \tag{2.6}$$

By multiplying by p_k and equating it to zero we get the equations

$$1 + \lambda_0 p_k + \sum_{i=1}^{K}\lambda_i p_k e^{\alpha_i+\beta_i h(z_k)} = 0, \quad k = 1,\ldots,N.$$

Summation with respect to k and use of the constraints (2.3) gives

$$N + \lambda_0 + \sum_{i=1}^{K}\lambda_i = 0$$

which together with (2.5) implies

$$\lambda_0 = -n_0. \tag{2.7}$$

If we replace the values of all Lagrange multipliers given in (2.5) and (2.7) into the derivative in (2.6) and set it equal to zero we get

$$p_k = \frac{1}{n_0 + \sum_{i=1}^{K} n_i e^{\alpha_i+\beta_i h(z_k)}} = \left\{\sum_{i=0}^{K} n_i e^{\alpha_i+\beta_i h(z_k)}\right\}^{-1} \tag{2.8}$$

(recall that we have set $\alpha_0 = \beta_0 = 0$). By replacing further these values into (2.4) we obtain the profile loglikelihood of α, β,

$$\tilde{\ell}(\alpha,\beta) = \sum_{i=1}^{K} n_i\alpha_i + \sum_{i=1}^{K}\beta_i\sum_{j=1}^{n_i} h(x_{ij}) - \sum_{k=1}^{N}\log\sum_{i=0}^{K} n_i e^{\alpha_i+\beta_i h(z_k)}. \tag{2.9}$$

The last function can be easily maximized with respect to α, β since it is essentially the log-likelihood of the parameters in a multinomial logistic regression model. More specifically, suppose that with have $K+1$ categories labeled $0, 1, \ldots, K$, and the probability $\pi_i(z)$ of observing the ith category given the covariate value z satisfies

$$\log\{\pi_i(z)/\pi_0(z)\} = \alpha'_i + \beta_i h(z),$$

where $\alpha_i' = \alpha_i + \log(n_i/n_0)$. Then, the loglikelihood of this model's parameters is exactly (2.9).

The well-known condition of "nonseparation" for the existence of the MLEs in the logistic regression models (Albert and Anderson, 1984) applies in the case of DR model as well. Hence, in order the MLEs to exist and be unique, the observations x_{ij}, $i = 0, 1, \ldots, K$, $j = 1, \ldots, n_{ij}$, must satisfy the condition

$$\text{for all } i \in \{0, 1, \ldots, K\} \text{ there exists } i' \in \{0, 1, \ldots, K\},\ i' \neq i,$$
$$\text{such that } \big(\min_j h(x_{ij}), \max_j h(x_{ij})\big) \bigcap \big(\min_j h(x_{i'j}), \max_j h(x_{i'j})\big) \neq \varnothing.$$

In simple words, this condition means that there must be no sample completely separated of the others.

The (profile) likelihood equations for α and β are

$$\frac{\partial}{\partial \alpha_i} \tilde{\ell}(\alpha, \beta) = \; n_i - \sum_{k=1}^{N} \frac{n_i e^{\alpha_i + \beta_i h(z_k)}}{\sum_{i=0}^{K} n_i e^{\alpha_i + \beta_i h(z_k)}} = 0, \quad i = 1, \ldots, K, \tag{2.10}$$

$$\frac{\partial}{\partial \beta_i} \tilde{\ell}(\alpha, \beta) = \; \sum_{j=1}^{n_i} h(x_{ij}) - \sum_{k=1}^{N} \frac{n_i h(z_k) e^{\alpha_i + \beta_i h(z_k)}}{\sum_{i=0}^{K} n_i e^{\alpha_i + \beta_i h(z_k)}} = 0, \quad i = 1, \ldots, K. \tag{2.11}$$

Once we obtain their solutions $\widehat{\alpha}_1, \ldots, \widehat{\alpha}_K, \widehat{\beta}_1, \ldots, \widehat{\beta}_K$ we plug them into (2.8) to get the MLE of p_k,

$$\widehat{p}_k = \frac{1}{\sum_{i=0}^{K} n_i e^{\widehat{\alpha}_i + \widehat{\beta}_i h(z_k)}}.$$

Thus, the MLE of F_0 is

$$\widehat{F}_0(x) = \sum_{z_k \leq x} \widehat{p}_k = \sum_{z_k \leq x} \frac{1}{\sum_{i=0}^{K} n_i e^{\widehat{\alpha}_i + \widehat{\beta}_i h(z_k)}}, \quad x \in \mathbb{R}.$$

Furthermore, the MLEs of $F_1, \ldots, F_K$ are given by

$$\widehat{F}_i(x) = \sum_{z_k \leq x} e^{\widehat{\alpha}_i + \widehat{\beta}_i h(z_k)} \widehat{p}_k = \sum_{z_k \leq x} \frac{e^{\widehat{\alpha}_i + \widehat{\beta}_i h(z_k)}}{\sum_{i=0}^{K} n_i e^{\widehat{\alpha}_i + \widehat{\beta}_i h(z_k)}}, \quad x \in \mathbb{R}, \quad i = 1, \ldots, K.$$

Fokianos et al. (2001) showed that $\widehat{\alpha}$, $\widehat{\beta}$ are asymptotically normal estimators of α, β, i.e.,

$$\sqrt{N}\begin{pmatrix}\widehat{\alpha}-\alpha\\ \widehat{\beta}-\beta\end{pmatrix}\xrightarrow{d}\mathcal{N}_{2K}(0,\Sigma) \tag{2.12}$$

as $N \to \infty$ in such a way that $n_i/N \to \rho_i \in (0,1)$, $i = 0,1,\ldots,K$. (The last condition means that all sample sizes must tend to infinity at the same rate.) The covariance matrix of the asymptotic normal distribution is given by $\Sigma = U^{-1}VU^{-1}$ where

$$U = N^{-1}E\{\nabla^2\tilde{\ell}(\alpha,\beta)\} \quad \text{and} \quad V = N^{-1}E\{\nabla\tilde{\ell}(\alpha,\beta)\nabla\tilde{\ell}(\alpha,\beta)^T\}.$$

Here, $\nabla\tilde{\ell}(\alpha,\beta)$ denotes the $(2K)$-dimensional column vector of the derivatives in (2.10), (2.11) (i.e., the gradient of $\tilde{\ell}$) while $\nabla^2\tilde{\ell}(\alpha,\beta)$ the $(2K)\times(2K)$ matrix with entries the second partial derivatives of $\tilde{\ell}$ (i.e., its Hessian). This follows from the general theory of M-estimators. It is important to note that if some β_i's coincide then the corresponding α_i's coincide as well and then the asymptotic normal distribution in (2.12) is singular. On the other hand, the general theory of semiparametric models and profile likelihood implies that $\widehat{F}_0$ converges to a pinned Gaussian process with mean function F_0.

2.3 The DR model as an ALT model

Assume that we want to put on a lifetest N items in order to estimate a baseline distribution function F_0, where the term “baseline” means that it is the distribution function of the items' lifes under standard conditions. In order to deal with time limitations, we may split the N items into $K+1$ groups of $n_0, n_1, \ldots, n_K$ items, respectively, and run the experiment under different stress levels for each group. Group 0 is put on the standard stress level, hence the distribution function generating the corresponding sample is F_0. The remaining groups are put on highter stress levels and it is assumed that the stress for group i is higher than that of group j for all $i < j$. Denote by F_i, $i = 1, \ldots, K$, the distribution functions of the remaining groups and let $\bar{F}_i = 1 - F_i$, $i = 0, 1, \ldots, K$. Since the stress level increases with i, one expects that the distributions are ordered, i.e., $F_i \succeq_{\text{st}} F_j$ for $i < j$. Recall that this means that $\bar{F}_i(x) \geq \bar{F}_j(x)$, for all x, which means that lifetimes under a higher stress level are expected to take lower values than lifetimes under a lower stress level.

The DR model in (2.2) it is a plausible model to describe such data. Assume that the lifetime distribution of interest has density f_0 while the distributions under higher stress levels belong to the same family with densities satifying (2.1). The parameters β_i are assumed to depend on the ith stress level and they are easily estimated using the procedure described in the previous section.

Although ordering of the distributions is expected to hold, sometimes the estimated parameters don't completely agree with this assumption. In cases where we strongly believe to the ordering, we could apply order restricted inference (Silvapulle and Sen, 2005) to force the estimates to satisfy this assumption. Davidov, Fokianos and Iliopoulos (2010) considered order restricted inference for the power biased model, i.e., the DR model with $h(x) = \log x$. However their procedure can be easily applied with any strictly increasing function instead of the logarithm.

2.4 Illustrative example

In order to illustrate the DR model we consider a dataset originally presented in Kalkanis and Rosso (1989) and discussed as an example in Meeker and Escobar (1998, Example 18.5). The data represent times of dielectric breakdown in minutes of units tested

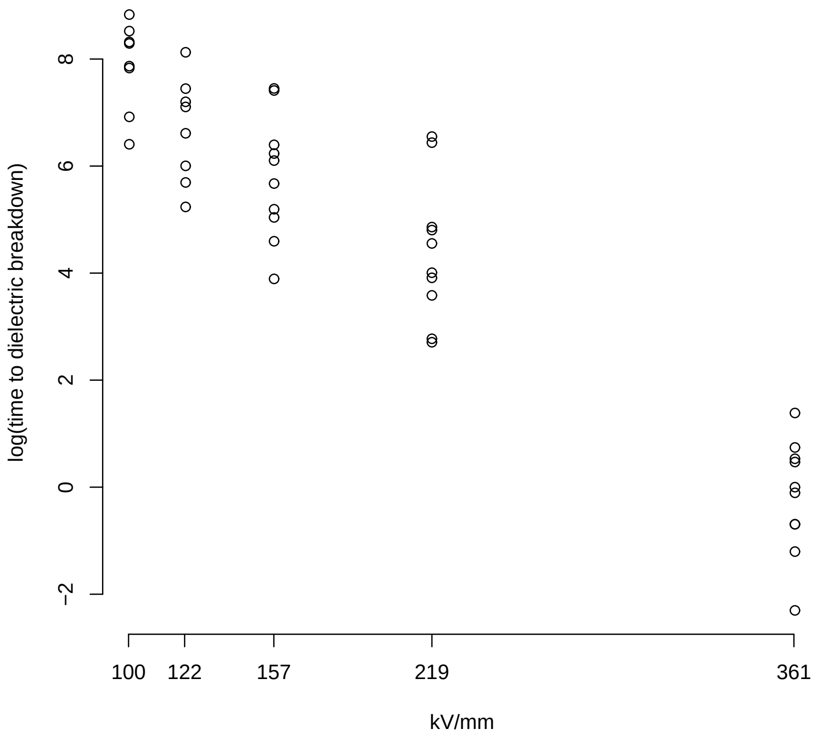

Figure 2.1: Log(times to dielectric breakdown) of units tested at $100.3, 122.4, 157.1, 219.0$, and 361.4 kV/mm.

at $100.3, 122.4, 157.1, 219.0$, and 361.4 kV/mm, i.e., under five different voltage stresses. Meeker and Escobar (1998) explain that voltage stress across a dielectric is measured in units of volts/thickness. The data are depicted in Figure 2.1.

Let us use the DR model in order to estimate the CDF of the time of breakdown under 100.3 kV/mm. Figure 2.1 shows some linearity of the data logarithms so it seems natural to take $h(x) = \log x$ and assume that

$$f_i(x) = f_0(x)e^{\alpha_i + \beta_i \log x} \propto x^{\beta_i} f_0(x), \quad i = 0, 1, \ldots, 4.$$

However, as we can clearly see from the graph, the last group corresponding to the highest stress is separated from the other four groups. This means that the MLEs of the parameters do not exist. Therefore, we will use only the first four groups to estimate the model.

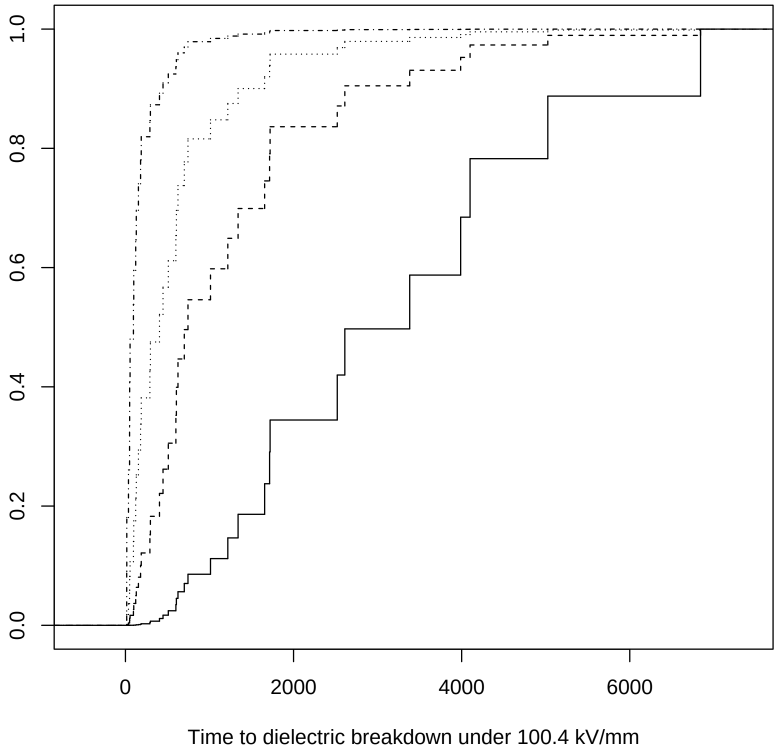

Figure 2.2: Estimated CDFs of time to dielectric breakdown at 100.3, 122.4, 157.1 and 219.0 kV/mm.

The maximization of $\tilde{\ell}$ has been done in `R` using the function `optim`. (Details on the implementation are given in the next chapter.) After running the procedure we get the estimates

$$\widehat{\alpha} = (11.72, 16.31, 21.28), \quad \widehat{\beta} = (-1.59, -2.33, -3.30)$$

and thus that the estimates of the underlying CDFs are

$$\widehat{F}_0(x) = \sum_{z_k \leqslant x} \frac{1}{8 + 8e^{11.72-1.59\log z_k} + 10e^{16.31-2.33\log z_k} + 10e^{21.28-3.30\log z_k}},$$
$$\widehat{F}_1(x) = \sum_{z_k \leqslant x} \frac{8e^{11.72-1.59\log z_k}}{8 + 8e^{11.72-1.59\log z_k} + 10e^{16.31-2.33\log z_k} + 10e^{21.28-3.30\log z_k}},$$
$$\widehat{F}_2(x) = \sum_{z_k \leqslant x} \frac{e^{16.31-2.33\log z_k}}{8 + 8e^{11.72-1.59\log z_k} + 10e^{16.31-2.33\log z_k} + 10e^{21.28-3.30\log z_k}},$$
$$\widehat{F}_3(x) = \sum_{z_k \leqslant x} \frac{e^{21.28-3.30\log z_k}}{8 + 8e^{11.72-1.59\log z_k} + 10e^{16.31-2.33\log z_k} + 10e^{21.28-3.30\log z_k}}.$$

Figure 2.2 shows the four estimated CDFs. The lowest line corresponds to $\widehat{F}_0$ while the highest line to $\widehat{F}_3$. This follows from the fact that $\widehat{\beta}_i$ decreases with i and thus, the CDFs are stochastically decreasing in i.

Chapter 3

A density ratio model with proportional β's

As mentioned in the previous section, the parameters $\beta_1, \ldots, \beta_K$ depend on the particular stress level under which we observe the "biased" lifetimes X_{ij}, $i = 1, \ldots, K$, $j = 1, \ldots, n_i$. In this section we consider a linear model for these parameters. In particular, we assume that

$$\beta_i = \phi w_i, \quad i = 1, \ldots, K,$$

where $w_1, \ldots, w_K$ are known values related to the K stress levels. For instance, if the different stress levels are formed by changing the operating temperature, w_i could be either the temperature or some simple transformation of it. Even simpler, if the different stresses could be assumed "equidistance", we could set $w_i = i$. For convenience, in the sequel we set $w_0 = 0$.

3.1 Empirical likelihood and its maximization

Under this model and with data as in Section 2.2, the empirical likelihood of $\alpha = (\alpha_1, \ldots, \alpha_K)$, ϕ and F_0 is

$$\begin{aligned} L(\alpha, \phi, F_0) &= \prod_{i=0}^{K} \prod_{j=1}^{n_i} \mathrm{d}F_i(x_{ij}) \\ &= \prod_{i=0}^{K} \prod_{j=1}^{n_i} \mathrm{d}F_0(x_{ij}) \exp\{\alpha_i + \phi w_i h(x_{ij})\} \end{aligned}$$

$$= \exp\left\{\sum_{i=1}^{K} n_i\alpha_i + \phi\sum_{i=1}^{K} w_i \sum_{j=1}^{n_i} h(x_{ij})\right\}\prod_{i=0}^{K}\prod_{j=1}^{n_i} \mathrm{d}F_0(x_{ij}),$$

with the parameters satisfying the constraints $\int e^{\alpha_i+\phi w_i h(x)}\mathrm{d}F_0(x) = 1$, $i = 0, 1, \ldots, K$. Similarly to the Section 2.2, the maximization of L is equivalent to the maximization of

$$L^*(\alpha,\phi,p) = \exp\left\{\sum_{i=1}^{K} n_i\alpha_i + \phi\sum_{i=1}^{K} w_i \sum_{j=1}^{n_i} h(x_{ij})\right\}\prod_{k=1}^{N} p_k,$$

under the constraints $\sum_{k=1}^{N} p_k e^{\alpha_i+\phi w_i h(z_k)} = 1$, $i = 0, 1, \ldots, K$. By taking the logarithm and employing Lagrange multipliers we can again profile with respect to p since the maximizer satisfies

$$p_k = \left\{\sum_{i=0}^{K} n_i e^{\alpha_i+\phi w_i h(z_k)}\right\}^{-1}, \quad k = 1, \ldots, N,$$

i.e., formula (2.8) with the ϕw_i's in the place of β_i's. By replacing these p_k's, the profile loglikelihood of α's and ϕ becomes

$$\tilde{\ell}(\alpha,\phi) = \sum_{i=1}^{K} n_i\alpha_i + \phi\sum_{i=1}^{K} w_i \sum_{j=1}^{n_i} h(x_{ij}) - \sum_{k=1}^{N}\log\sum_{i=0}^{K} n_i e^{\alpha_i+\phi w_i h(z_k)}.$$

The likelihood equations are

$$\frac{\partial}{\partial\alpha_i}\tilde{\ell}(\alpha,\phi) = n_i - \sum_{k=1}^{N}\frac{n_i e^{\alpha_i+\phi w_i h(z_k)}}{\sum_{i=0}^{K} n_i e^{\alpha_i+\phi w_i h(z_k)}} = 0, \quad i = 1, \ldots, K,$$

$$\frac{\partial}{\partial\phi}\tilde{\ell}(\alpha,\phi) = \sum_{i=1}^{K} w_i \sum_{j=1}^{n_i} h(x_{ij}) - \sum_{k=1}^{N}\frac{n_i w_i h(z_k) e^{\alpha_i+\phi w_i h(z_k)}}{\sum_{i=0}^{K} n_i e^{\alpha_i+\phi w_i h(z_k)}} = 0.$$

Their solution corresponds to the MLEs under the DR model with proportional β's.

3.2 Testing the proportionality assumption

The adequacy of the proportional β's model must be tested, given that the original DR model holds. There are two basic procedures which can be applied in order to test the hypothesis

$$H_0 : \beta_i = \phi w_i, \quad i = 1, \ldots, K \qquad \text{vs} \qquad H_1 : H_0 \text{ does not hold.} \tag{3.1}$$

First, there is the asymptotic approach. Clearly, the proportional β's model is nested to the free β's DR model. Since the profile likelihood of the DR model behaves essentially as a standard parametric likelihood, we have as $n_i \to \infty$, $i = 0, 1, \ldots, K$, such that $n_i/N \to \rho_i \in (0, 1)$ for all i,

$$T = 2\{\tilde{\ell}(\widehat{\alpha}, \widehat{\beta}) - \tilde{\ell}(\widehat{\alpha}, \widehat{\phi})\} \xrightarrow{d} \chi^2_{K-1}.$$

This means that the asymptotic approach tests the null hypothesis based on a standard chi-squared test: The asymptotic p-value is

$$\mathrm{p_A} = P(\chi^2_{K-1} \geq T_{\mathrm{obs}}), \tag{3.2}$$

where T_{obs} is the observed value of T.

The second method one can use is bootstrap. In general, in order to perform a bootstrap test for a null hypothesis based on some statistic T, one has to draw B bootstrap samples by resampling with replacement from the observed sample under the assumption that the null hypothesis holds. If $T_1^*, \ldots, T_B^*$ are the values of the statistic T based on these bootstrap samples, the bootstrap p-value is

$$\mathrm{p_B} = B^{-1} \sum_{b=1}^{B} I(T_b^* \geq T_{\mathrm{obs}}). \tag{3.3}$$

In general, in the nonparametric or semiparametric context, the exact distribution which generates the data is unknown even under the null hypothesis. Therefore, the bootstrap samples are formed from the corresponding estimated null distribution. In our case, this means that we will sample from the (discrete) distributions with CDFs

$$\widehat{F}_i(x) = \sum_{z_k \leq x} \frac{e^{\widehat{\alpha}_i + \phi w_i h(z_k)}}{\sum_{i=0}^{K} e^{\widehat{\alpha}_i + \phi w_i h(z_k)}}, \quad i = 0, 1, \ldots, K.$$

At the bth bootstrap iteration we sample n_0 z's from $\widehat{F}_0$, n_1 z's from $\widehat{F}_1$ and so on in order to form the bootstrap sample x_{ij}^*, $i = 0, 1, \ldots, K$, $j = 1, \ldots, n_i$, from which the bootstrapped

		Asymptotic			Bootstrap		
$K+1=3$	True w_i	10%	5%	1%	10%	5%	1%
$n_i=20$	i	.108	.062	.014	.103	.055	.012
	$i^{1/2}$	.212	.142	.036	.200	.131	.032
	$i^{3/2}$	.204	.120	.040	.193	.110	.039
	i^2	.447	.346	.160	.437	.325	.437
$n_i=50$	i	.093	.049	.013	.087	.046	.012
	$i^{1/2}$	.307	.213	.078	.305	.209	.081
	$i^{3/2}$	.319	.216	.091	.321	.209	.098
	i^2	.793	.715	.473	.793	.695	.474
$n_i=100$	i	.112	.059	.016	.116	.058	.017
	$i^{1/2}$	.517	.403	.194	.517	.400	.199
	$i^{3/2}$	.542	.412	.203	.537	.414	.212
	i^2	.983	.949	.842	.981	.950	.840

Table 3.1: Estimated size and power of asymptotic and bootstrap tests of nominal levels 1%, 5% and 10% for testing proportionality of β's when $K+1=3$ groups are observed based on 1000 Monte Carlo iterations. The first row for each sample size corresponds to the actual size of the tests.

value T_b^* of the (log)likelihood ratio statistic $T = 2\{\tilde{\ell}(\widehat{\alpha},\widehat{\beta}) - \tilde{\ell}(\widehat{\alpha},\widehat{\phi})\}$ is calculated. Finally, the corresponding p-value is calculated as in (3.3).

In order to evaluate the performance of the two tests, a small simulation study is conducted. As baseline distribution is taken the standard exponential, i.e.,

$$F_0(x) = 1 - e^{-x}, \quad x > 0.$$

The distributions under the other stresses were taken also exponentials which means that $h(x) = x$. Their rates are set equal to $1+\phi w_i$, $i = 1,\ldots,K$, so that

$$f_i(x) = f_0(x)e^{\log(1+\phi w_i)-\phi w_i x}, \quad i = 1,2,\ldots,$$

i.e., $\beta_i = -\phi w_i$. In our simulations we took $\phi = 2$ and $w_i = i, i^{1/2}, i^{3/2}, i^2$. However, in

		Asymptotic			Bootstrap		
$K+1=4$	True w_i	10%	5%	1%	10%	5%	1%
$n_i = 20$	i	.098	.044	.010	.083	.038	.009
	$i^{1/2}$	.252	.151	.045	.225	.142	.045
	$i^{3/2}$	.265	.181	.072	.246	.160	.063
	i^2	.722	.613	.356	.705	.586	.342
$n_i = 50$	i	.105	.052	.009	.098	.046	.007
	$i^{1/2}$	.488	.348	.155	.473	.342	.163
	$i^{3/2}$	.517	.393	.187	.512	.387	.196
	i^2	.976	.958	.866	.977	.963	.867
$n_i = 100$	i	.098	.048	.012	.099	.044	.011
	$i^{1/2}$	.724	.588	.331	.716	.584	.347
	$i^{3/2}$	.788	.694	.467	.789	.689	.460
	i^2	1	1	.999	1	1	.999

Table 3.2: Estimated size and power of asymptotic and bootstrap tests of nominal levels 1%, 5% and 10% for testing proportionality of β's when $K+1=4$ groups are observed based on 1000 Monte Carlo iterations. The first row for each sample size corresponds to the actual size of the tests.

---

fitting the model, we consider $w_i = i$ in all cases. Hence, when the true w_i is i the null hypothesis holds while for the other w_i's it does not. We considered $K=2$ and $K=3$, i.e., the cases where we have in total three and four groups and balanced samples of sizes $20, 50$ and 100. For each configuration of w_i's, K and n_i's we ran 1000 Monte Carlo iterations and in every iteration we took $B = 500$ bootstrap samples. The power of the two tests for nominal level α $(= 10\%$, 5% or $1\%)$ was estimated as follows: In the mth Monte Carlo iteration we calculated the two p-values $\mathrm{p}_{\mathrm{A}}^{(m)}$ and $\mathrm{p}_{\mathrm{B}}^{(m)}$ using (3.2) and (3.3), respectively. We then set

$$\mathrm{power}_{\mathrm{A}} = \sum_{m=1}^{1000} I(\mathrm{p}_{\mathrm{A}}^{(m)} \leq \alpha)/1000 \quad \text{and} \quad \mathrm{power}_{\mathrm{B}} = \sum_{m=1}^{1000} I(\mathrm{p}_{\mathrm{B}}^{(m)} \leq \alpha)/1000,$$

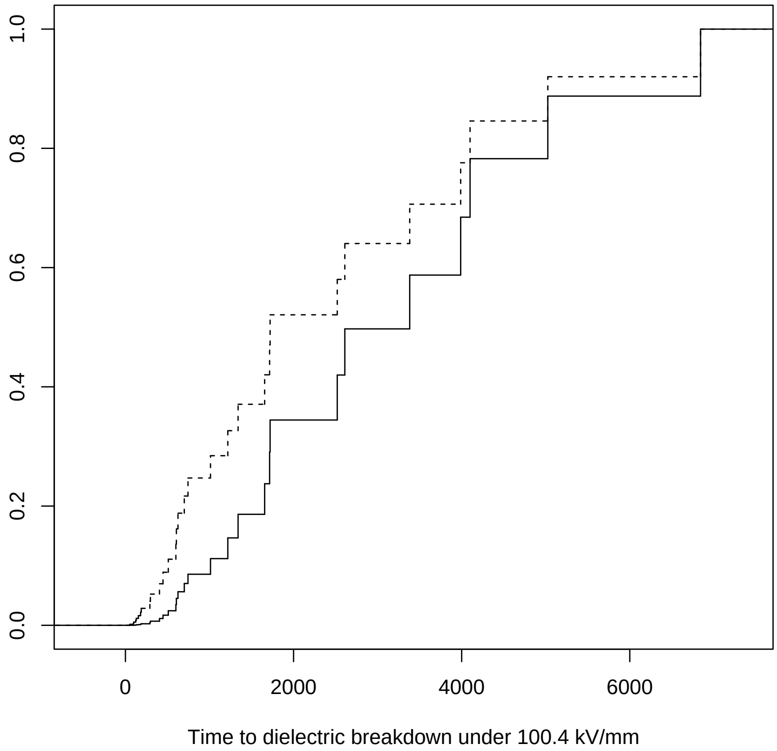

Figure 3.1: Estimates of F_0 under the original DR model (solid line) and under the proportional β's DR model (dashed line).

that is, the percentages of rejections of H_0 by the two tests at level α. The results are displayed in Tables 3.1 and 3.2. We can see that the size of both tests is close to the nominal level of significance while they have the power to reject the null when it is not true. Moreover, the asymptotic test seems to be slightly more powerful than the bootstrap one. However, it is unknown how B affects the power of the latter. This subject needs further investigation.

3.3 Illustrative example

In order to illustrate the proportinal β's DR model we use again the dataset analyzed previously. We consider again $h(x) = \log x$ and set

$$(w_1, w_2, w_3) = (122.4 - 100.3, 157.1 - 100.3, 219.0 - 100.3) = (22.1, 56.8, 118.7),$$

i.e., the differences between the kV/mm of the three higher stess levels to that of the “baseline” one.

After running the procedure with the proportional β’s DR model, we got

$$\widehat{\alpha} = (3.51, 8.47, 15.42), \quad \widehat{\phi} = -0.02194$$

so that

$$\widehat{\phi} \times (w_1, w_2, w_3) = (-0.48, -1.24, -2.60)$$

The estimated F_0 under this model is shown in Figure 3.1 together with the estimated F_0 under the first DR model.

The adequacy of proportional β’s model under the assumption that the original DR model holds has been tested using both the asymptotic approach (although the sample size is rather small) and the bootstrap procedure. The corresponding p-values were 0.169 and 0.264, respectively. This means that the proportional β’s model can describe the underlying distributions almost as well as the original DR model.

Chapter 4

A density ratio model with randomly censored data

4.1 Random censoring

Censoring is very common in life testing experiments and can occur either by design or accidentally. Here we will consider the second case and, more specifically, independent random censoring. Let X° denote the original lifetime of an item and U be an independent random variable with some unknown distribution H_U. Then, X° is subject to random censoring if we observe $X = \min\{X^\circ, U\}$ rather than X° itself. This kind of situation may occur if items can fail due to an independent cause which is irrelevant to the experiment.

4.2 Empirical likelihood and its maximization

Let X_{ij}°, $i = 0, 1, \ldots, K$, $j = 1, \ldots, n_i$, be lifetimes following model (2.2) and U_{ij} are corresponding independent random variables. Assume that we observe (X_{ij}, δ_{ij}), $i = 0, 1, \ldots, K$, $j = 1, \ldots, n_i$, where $X_{ij} = \min\{X_{ij}^\circ, U_{ij}\}$ and $\delta_{ij} = I(X_{ij}^\circ \leq U_{ij}) \equiv I(X_{ij} = X_{ij}^\circ)$. The empirical likelihood based on the data described above is

$$\begin{aligned} L(\alpha, \beta, F_0) = & \prod_{i=0}^{K} \prod_{j=1}^{n_i} \mathrm{d}F_i(x_{ij})^{\delta_{ij}} \bar{F}_i(x_{ij})^{1-\delta_{ij}} \\ = & \prod_{i=0}^{K} \prod_{j=1}^{n_i} \mathrm{d}F_0(x_{ij})^{\delta_{ij}} e^{\delta_{ij}\{\alpha_i + \beta_i h(x_{ij})\}} \bar{F}_i(x_{ij})^{1-\delta_{ij}}. \end{aligned}$$

As in the previous chapter, let $z_1, \ldots, z_N$ be the combined sample and $p_k = \mathrm{d}F_0(z_k)$. Then, maximization of L is equivalent to the maximization of

$$L^*(\alpha, \beta, p) = \Big(\prod_{k=1}^{N} p_k\Big) \exp\Big\{\sum_{i=1}^{K} \alpha_i \sum_{j=1}^{n_i} \delta_{ij} + \sum_{i=1}^{K} \beta_i \sum_{j=1}^{n_i} \delta_{ij} h(x_{ij})\Big\} \times \prod_{i=0}^{K}\prod_{j=1}^{n_i}\Big(\sum_{z_k > x_{ij}} p_k e^{\alpha_i + \beta_i h(z_k)}\Big)^{1-\delta_{ij}},$$

under the constraints $\sum_{k=1}^{N} p_k e^{\alpha_i + \beta_i h(z_k)} = 1$, $i = 0, 1, \ldots, K$.

Unfortunately, the above maximization problem is much more involved than the one we described in the previous chapter due to the presence of the last term. However, recently Wei and Zhou (2016) presented an EM algorithm for maximizing such likelihoods. More specifically, they considered a two-sample density ratio model where the data are subject to independent random censoring. Here we will slightly extend their approach to the case of more than two samples.

Recall that the EM algorithm (Dempster et al., 1977) is an iterative procedure which can be applied in models with missing data. The algorithm alternates repeatedly between the E-step (E for "expectation") and the M-step (M for "maximization") until the sequence of maximizers to converge. In the special case where the underlying distribution belongs to an exponential family, the algorithm simplifies a lot: The E-step consists of imputing as missing values their expectations by considering as true values of the parameters the maximizer of the previous step. After the imputation, the data are "complete" and the maximization is carried out as usually.

Since the underlying distributions are assumed to be continuous, each of the values $z_1, \ldots, z_N$ has been observed only once. By the discrete nature of the maximizer with respect to F_0, the only values supported by p are the observed values $z_1, \ldots, z_N$. Since the data are censored, some of these values could have been "observed" more than one time. In the sequel assume that the z-values have been sorted, i.e., that $z_1 < \cdots < z_N$. Let b_{ik} denote the multiplicity of z_k in the ith sample. Note that when the underlying distributions are continuous we have $b_{ik} = 1$ or 0 according to whether z_k comes from the ith sample or not and $\sum_{k=1}^{N} b_{ik} = n_i$. Wei and Zhou (2016) gave the expected number of observations which are equal to z_k in the group i explicitly. In our case where the underlying distributions are assumed continuous, it reads

$$E(b_{ik}) = I(z_k \text{ is an uncensored value from group } i)+ \\ p_k e^{\alpha_i+\beta_i h(z_k)} \sum_{j=1}^{k} \frac{I(z_j \text{ is a censored value from group } i)}{\sum_{\ell=j}^{N} p_\ell e^{\alpha_i+\beta_i h(z_\ell)}}. \quad (4.1)$$

In order to understand this, notice that, given the data, the expectation of the original (i.e., before possible censoring) lifetime X°_{ij} under the discrete DR model which assigns positive probabilities $p_1, \dots, p_N$ only to $z_1, \dots, z_N$ is

$$\begin{aligned}
E\{I(X^\circ_{ij} = z_k)|\text{data}\} &= E\{I(X^\circ_{ij} = z_k, \delta_{ij} = 1)|X_{ij}, \delta_{ij}\} + E\{I(X^\circ_{ij} = z_k, \delta_{ij} = 0)|X_{ij}, \delta_{ij}\} \\
&= I(X_{ij} = z_k, \delta_{ij} = 1) + I(X_{ij} \le z_k, \delta_{ij} = 0) \frac{P(X^\circ_{ij} = z_k, U_{ij} = X_{ij})}{\sum_{X_{ij} \le z_\ell} P(X^\circ_{ij} = z_\ell, U_{ij} = X_{ij})} \\
&= I(X_{ij} = z_k, \delta_{ij} = 1) + I(X_{ij} \le z_k, \delta_{ij} = 0) \frac{P(X^\circ_{ij} = z_k)\mathrm{d}H_U(X_{ij})}{\sum_{X_{ij} \le z_\ell} P(X^\circ_{ij} = z_\ell)\mathrm{d}H_U(X_{ij})} \\
&\qquad \text{(by the independence assumption of } X^\circ_{ij} \text{ and } U_{ij}) \\
&= I(X_{ij} = z_k, \delta_{ij} = 1) + I(X_{ij} \le z_k, \delta_{ij} = 0) \frac{P(X^\circ_{ij} = z_k)\mathrm{d}H_U(X_{ij})}{\sum_{X_{ij} \le z_\ell} P(X^\circ_{ij} = z_\ell)\mathrm{d}H_U(X_{ij})} \\
&= I(X_{ij} = z_k, \delta_{ij} = 1) + \sum_{m=1}^{k} I(X_{ij} = z_m, \delta_{ij} = 0) \frac{p_k e^{\alpha_i+\beta_i h(z_k)}}{\sum_{\ell=m}^{K} p_\ell e^{\alpha_i+\beta_i h(z_\ell)}}.
\end{aligned}$$

Therefore, the expected value of the number of observations in group i which are equal to z_k is

$$\sum_{j=1}^{n_i} \left\{ I(X_{ij} = z_k, \delta_{ij} = 1) + \sum_{m=1}^{k} I(X_{ij} = z_m, \delta_{ij} = 0) \frac{p_k e^{\alpha_i+\beta_i h(z_k)}}{\sum_{\ell=m}^{K} p_\ell e^{\alpha_i+\beta_i h(z_\ell)}} \right\}$$

$$= \sum_{j=1}^{n_i} I(X_{ij} = z_k, \delta_{ij} = 1) + p_k e^{\alpha_i + \beta_i h(z_k)} \sum_{m=1}^{k} \sum_{j=1}^{n_i} I(X_{ij} = z_m, \delta_{ij} = 0) \frac{1}{\sum_{\ell=m}^{K} p_\ell e^{\alpha_i + \beta_i h(z_\ell)}}$$

which is equal to (4.1) since

$$\sum_{j=1}^{n_i} I(X_{ij} = z_k, \delta_{ij} = 1) = I(z_k \text{ is an uncensored value from group } i)$$

and

$$\sum_{j=1}^{n_i} I(X_{ij} = z_m, \delta_{ij} = 0) = I(z_m \text{ is a censored value from group } i).$$

When possible multiplicities of observations are considered the likelihood based on uncensored data is

$$\begin{aligned} L(\alpha, \beta, F_0) = & \prod_{k=1}^{N} \prod_{i=0}^{K} \mathrm{d}F_i(z_k)^{b_{ik}} \\ = & \prod_{k=1}^{N} \prod_{i=0}^{K} \mathrm{d}F_0(z_k)^{b_{ik}} e^{b_{ik}\{\alpha_i + \beta_i h(z_k)\}} \\ = & \Big(\prod_{k=1}^{N} \mathrm{d}F_0(z_k)^{b_k} \Big) \exp\Big\{ \sum_{i=1}^{K} \alpha_i \sum_{k=1}^{N} b_{ik} + \sum_{i=1}^{K} \beta_i \sum_{k=1}^{N} b_{ik} h(z_k) \Big\} \\ = & \Big(\prod_{k=1}^{N} \mathrm{d}F_0(z_k)^{b_k} \Big) \exp\Big\{ \sum_{i=1}^{K} n_i \alpha_i + \sum_{i=1}^{K} \beta_i \sum_{k=1}^{N} b_{ik} h(z_k) \Big\} \\ & \text{(recall that } \textstyle\sum_{k=1}^{N} b_{ik} = n_i\text{).} \end{aligned}$$

As before, its maximization with respect to α, β and F_0 is equivalent to the maximization of

$$L^*(\alpha, \beta, p) = \Big(\prod_{k=1}^{N} p_k^{b_k} \Big) \exp\Big\{ \sum_{i=1}^{K} n_i \alpha_i + \sum_{i=1}^{K} \beta_i \sum_{j=1}^{n_i} b_{ik} h(z_k) \Big\}$$

with respect to α, β and p, under the constraints $\sum_{k=1}^{N} p_k e^{\alpha_i + \beta_i h(z_k)} = 1$, $i = 0, 1, \ldots, K$. By using again Lagrange multipliers and repeating the procedure of the previous chapter we get that

$$p_k = b_k \Big\{ \sum_{i=0}^{K} n_i e^{\alpha_i + \beta_i h(z_k)} \Big\}^{-1}, \quad k = 1, \ldots, N.$$

By replacing p_k into L^* and taking the logarithm we get the profile loglikelihood of α, β,

$$\tilde{\ell}(\alpha, \beta) = \sum_{i=1}^{K} n_i \alpha_i + \sum_{i=1}^{K} \beta_i \sum_{k=1}^{N} b_{ik} h(z_k) - \sum_{k=1}^{N} b_k \log \sum_{i=0}^{K} n_i e^{\alpha_i + \beta_i h(z_k)}. \tag{4.2}$$

In this case the likelihood equations become

$$\frac{\partial}{\partial \alpha_i} \tilde{\ell}(\alpha, \beta) = n_i - \sum_{k=1}^{N} b_k \frac{n_i e^{\alpha_i + \beta_i h(z_k)}}{\sum_{i=0}^{K} n_i e^{\alpha_i + \beta_i h(z_k)}} = 0, \quad i = 1, \dots, K,$$

$$\frac{\partial}{\partial \beta_i} \tilde{\ell}(\alpha, \beta) = \sum_{k=1}^{N} b_{ik} h(z_k) - \sum_{k=1}^{N} b_k \frac{n_i h(z_k) e^{\alpha_i + \beta_i h(z_k)}}{\sum_{i=0}^{K} n_i e^{\alpha_i + \beta_i h(z_k)}} = 0, \quad i = 1, \dots, K.$$

Following Wei and Zhou (2016), we run the following EM algorithm for maximizing L^*:

0. Initialization

Set $b_{ik}^{(0)} = I(z_k$ is an uncensored value from group $i)$, $i = 0, 1, \dots, K+1$, $k = 1, \dots, N$.

1. At iteration $m = 1, 2, \dots$:

M-step

Maximize $\tilde{\ell}(\alpha, \beta)$ in (4.2) with $b_{ik}^{(m-1)}$ in the place of b_{ik}.

Denote by $\alpha^{(m)}, \beta^{(m)}$ the maximizers and set

$$p_k^{(m)} = b_k^{(m-1)} \Big\{ \sum_{i=0}^{K} n_i e^{\alpha_i^{(m)} + \beta_i^{(m)} h(z_k)} \Big\}^{-1}, \quad k = 1, \dots, N.$$

E-step

Set

$$b_{ik}^{(m)} = I(z_k \text{ is an uncensored value from group } i) + p_k^{(m)} e^{\alpha_i^{(m)} + \beta_i^{(m)} h(z_k)} \sum_{j=1}^{k} \frac{I(z_j \text{ is a censored value from group } i)}{\sum_{\ell=j}^{N} p_\ell^{(m)} e^{\alpha_i^{(m)} + \beta_i^{(m)} h(z_\ell)}}$$

2. Repeat until some convergence criterion is met.

Natural convergence criteria could be $\sum_{i=1}^{K} \{|\alpha_i^{(m)} - \alpha_i^{(m-1)}| + |\beta_i^{(m)} - \beta_i^{(m-1)}|\} < \varepsilon$,

$\sum_{k=1}^{N} |p_k^{(m)} - p_k^{(m-1)}| < \varepsilon$, $\tilde{\ell}(\alpha^{(m)}, \beta^{(m)}) - \tilde{\ell}(\alpha^{(m-1)}, \beta^{(m-1)}) < \varepsilon$ where $\epsilon > 0$ is some tolerance as well as combinations of them.

Once the EM algorithm converges, we take as MLEs of α, β and p their current values.

4.3 Implementation in R

The maximization of the likelihood in R is done with the help of function `optim`. Given the data (x_{ij}, δ_{ij}), $i = 0, 1, \ldots, K$, $j = 1, \ldots, n_i$, we first define the vectors

$$z = (z_1, \ldots, z_N) \equiv (x_{01}, \ldots, x_{0n_0}, x_{11}, \ldots, x_{1n_1}, \ldots, x_{K1}, \ldots, x_{Kn_K}),$$

$$\delta = (\delta_{01}, \ldots, \delta_{0n_0}, \delta_{11}, \ldots, \delta_{1n_1}, \ldots, \delta_{K1}, \ldots, \delta_{Kn_K})$$

as well as the $N \times (K+1)$ "membership" matrix B_0 with (k,i)-entry

$$b_{0,ki} = I(z_k \text{ comes from sample } i), \quad k = 1, \ldots, N, \ i = 0, 1, \ldots, K.$$

Moreover, let z^{ord} be the sorted version of z and $\delta^{\text{ord}}, B_0^{\text{ord}}$ be δ, B_0 but permuted to much the ordering of z. In particular, in R we set

$$\texttt{z.ord} = \texttt{sort}(\texttt{z}), \quad \texttt{del.ord} = \texttt{del}[\texttt{order}(\texttt{z})] \quad \text{and} \quad \texttt{B0.ord} = \texttt{B0}[\texttt{order}(\texttt{z}),],$$

respectively (`del` and `B0` are δ and B_0). We also define the vectors $\texttt{hz} = h(\texttt{z})$, $\texttt{hz.ord} = h(\texttt{z.ord})$. We finally set $\texttt{B} = \texttt{B0}$: this will be the matrix of $b_{ik}^{(m)}$'s in the implementation of the EM algorithm.

The R function giving the (profile) loglikelihood of $\theta = (\alpha, \beta)$ is

```
loglikelihood = function(theta){
alf = c(0,theta[1:K]); bet = c(0,theta[(K+1):(2*K)])
sum(n*alf)+sum(bet*colSums(B*hz))-
sum(rowSums(B)*log(colSums(exp(alf+log(n)+outer(bet,hz)))))
}
```

while its gradient (the score function),

```
score = function(theta){
alf = c(0,theta[1:K]); bet = c(0,theta[(K+1):(2*K)])
a.plus.beta.h = t(alf+log(n)+outer(bet,hz))
c( (n-colSums(rowSums(B)*exp(a.plus.beta.h)/
              rowSums(exp(a.plus.beta.h))))[-1],
   (colSums(B*hz)-colSums(rowSums(B)*hz*exp(a.plus.beta.h)/
                          rowSums(exp(a.plus.beta.h))))[-1] )
}
```

The EM algorithm is implemented by repeated calls of the function

```
estimate = function(){
ahat.old <<- ahat; bhat.old <<- bhat
res = optim(p=c(ahat,bhat),
        fn=loglikelihood, g=score, method="BFGS",
        control=list(abstol=1e-16,reltol=1e-16,fnscale=-1))
ahat <<- res$par[1:K]; bhat <<- res$par[(K+1):(2*K)]
ahat0 = c(0,ahat); bhat0 = c(0,bhat)
phat <<- rowSums(B)*exp(t(ahat0+outer(bhat0,hz)))/
              rowSums(exp(t(ahat0+log(n)+outer(bhat0,hz))));
phat = t(t(phat)/colSums(phat))
phat.ord = phat[order(z),]
Phatbar.ord = apply(phat.ord[N:1,],2,cumsum)[N:1,]
B.ord = B0.ord*del.ord+
    phat.ord*apply(B0.ord*(1-del.ord)/Phatbar.ord,2,cumsum)
B <<- B.ord[rank(z),]
}
```

as follows:

```
while(sum(abs(ahat-ahat.old)+abs(bhat-bhat.old))>1e-8)
      estimate()
```

Before running the above `while` command, we assign initial values

$$\texttt{ahat.old} = \texttt{bhat.old} = \texttt{rep}(1, \texttt{K}) \quad \text{and} \quad \texttt{ahat} = \texttt{bhat} = \texttt{rep}(0, \texttt{K})$$

This ensures that `R` will proceed to the call of `estimate` at least once.

As we can see, the function `estimate` returns the maximizers $\alpha^{(m)}, \beta^{(m)}$ (variables `ahat`, `bhat`) together with $p^{(m)}$ (variable `phat`) as well as the updated matrix `B` (which contains $b_{ik}^{(m)}$'s as mentioned above), making them global variables via the assignment "`<<-`". Note finally that if δ contains only ones (i.e., there is no censoring in our data) then the function `estimate` returns the MLEs of α and β in just one step.

4.4 Illustrative example

In order to get an idea how censoring can affect the estimate of α, β and F_0 we consider again the dielectric breakdowns dataset analyzed in the previous chapter. There is no censoring in these data therefore we introduced artificial censoring random times as follows. We simulated exponential random variates U_{ij}, $i = 0, 1, 2, 3$, $j = 1, \ldots, n_i$, such that the mean of U_{ij} to be equal to a particular quantile of the corresponding group of observations. We then set x_{ij} the minimum of the (ij)th observation and U_{ij}.

We censored artificially the dataset using the third quartiles and the medians of the groups of the observations. This implies that we introduced about 25% and 50% censoring to the data, respectively. Here we report the results of a single run. When we used the third quartiles we got the MLEs

$$\widehat{\alpha} = (7.46, 14.47, 25.59), \quad \widehat{\beta} = (-1.00, -2.09, -4.46)$$

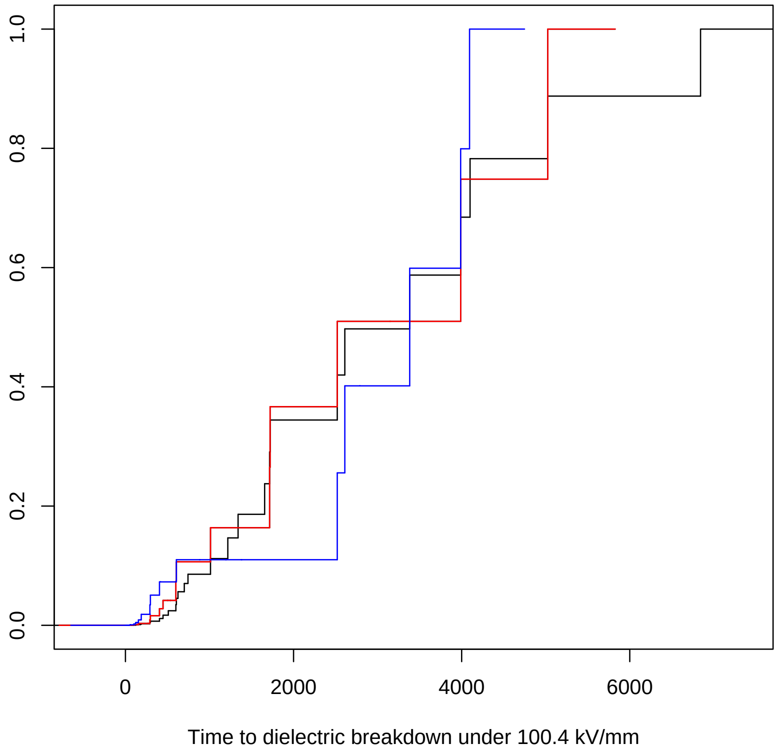

Figure 4.1: Estimates of F_0 based on the complete data (black line) and under 25% and 50% censoring rate (red and blue lines, respectively).

---

while when we used the median we got

$$\widehat{\alpha} = (6.43, 10.91, 19.24), \quad \widehat{\beta} = (-0.86, -1.57, -3.29).$$

(Recall that the MLEs based on the complete data were $\widehat{\alpha} = (11.72, 16.31, 21.28)$, $\widehat{\beta} = (-1.59, -2.33, -3.30)$.)

The estimates of F_0 are shown in Figure 4.1 using red and blue color, respectively. We can see that the red line is much closer to the black than the blue. This is expected since the less censored values the closer the estimate should be to the one we get from the complete data.

Final comments and thoughts for future work

As already mentioned, accelerated life test models cover a considerable amount of the literature on reliability and they are used in many scientific fields. In this dissertation we fitted the density ratio model to estimate a baseline distribution function using either complete or censored data. Howerver, many questions arise from this procedure. For instance, there is the problem of choice of the function h. This is a well-known problem in the DR model literature with no "correct" answer. Several suggestions have been made, including preliminary estimation of the densities and plotting their ratios. Another issue is the choice of w_i's in the proportional β's DR model.

There are other interesting types of reliability data where the DR model can be applied. In particular, we are interested in using it on accelerated life data in the presence of competing risks. To be more specific, assume that within each of the $K+1$ stresses there exist up to $m \geq 2$ different risk factors that may cause failure. Then, every datapoint contributes to the likelihood the density of the distribution related to a particular factor (the one that causes its failure) and the survival functions of the distributions related to the rest of the factors. We believe that if we model the data via a DR model then the EM algorithm described in the last chapter can be adapted in order to estimate its parameters.

Bibliography

Albert A., Anderson J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, **71**, 1–10.

Brunk H. D., Franck W. E, Hanson D. L, Hogg R. V. (1966). Maximum likelihood estimation of the distributions of two stochastically ordered random variables. *Journal of the American Statistical Association*, **61**, 1067–1080.

Cox, D. R. (1969). Some sampling problems in technology. In: Johnson, N.L., Smith, J.H. (Eds.), *New Developments in Survey Sampling*. Wiley–Interscience, New York, pp. 506–527.

Davidov O., Fokianos K., Iliopoulos G. (2010). Order-restricted semiparametric inference for the power bias model. *Biometrics*, **66**, 549–557.

Dempster A. P., Laird N. M., Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, **39**, 1–38.

Efron B., Tibshirani R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, **1**, 54–77.

Fisher R. A. (1934). The effect of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, **6**, 13–25

Fokianos K., Qin J., Kedem B., Short A. D. (2001). A semiparametric approach to the one-way Layout. *Technometrics*,**43**, 56–65.

Gilbert B. P. (2000). Large sample theory of maximum likelihood estimates in semiparametric biased sampling. *The Annals of Statistics*, **28**, 151–194.

Gilbert B. P., Lele R. S., Vardi Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with applications to AIDS vaccine trials. *Biometrika*, **86**, 27–43.

Gill D. R., Vardi Y., Wellner A. J., (1988). Large sample theory of empirical distributions in biased sampling models. *The Annals of Statistics*, **16**, 1069–1112.

Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skandinavian Aktuarial Journal*, **39**, 125–153.

Hartley, H. O. & Rao, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika*, **55**, 547–557.

Johansen S., (1978). The product limit estimator as maximum likelihood estimator. *Scandinavian Journal of Statistics*, **5**, 195–199.

Kalkanis, G.,and Rosso, E. (1989). The inverse power law model for the lifetime of a mylar-polyurethane laminated DC HV insulating structure. *Nuclear Instruments and Methods in Physics Research*, **A281**, 489–496.

Kaplan E.L. and Meier P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

Kiefer J., Wolfowitz J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics*, **18**, 600–685.

Liang S. (2005). Generalized linear and additive models with weighted distribution. M.Sc. Thesis, National University of Singapore.

Meeker W.Q., Escobar L.A. (1998). *Statistical Methods for Reliability Data*, John Wiley & Sons.

Nelson W. (1980). Accelerated life testing — Step-stress models and data analysis. *IEEE Transactions on Reliability*, **29**, 103–108

Nelson W. (2004). *Accelerated testing: Statistical models, test plans, and data analysis.* John Wiley & Sons, New Jersey.

Owen B. A. (2001). *Empirical likelihood*, Champman & Hall, CRC.

Patil G. P. (2002). Weighted distributions. *Encyclopedia of Environmetrics*, John Wiley & Sons, Chichester, **4**, 2369–2377.

Qin, J. and Lawless, J. F. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, **22**, 300–325.

Qin, J. and Zhang, B.,(1997). A goodness of fit test for logistic regression models based on case-control data. *Biometrica*, **84**, 609–618.

Rao C. (1965). *Linear Statistical Inference and its Applications*. Wiley & Sons, Canada.

Silvapulle M. J. and Sen P. K. (2005). *Constrained Statistical Inference: inequality, order, and shape restrictions.* Wiley, Hoboken, NJ.

Sun J., Woodroofe M. (1997). Semi-parametric estimates under biased sampling. *Statistica Sinica*, **7**, 545–575.

Vardi Y., (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics*, **10**, 616–620.

Vardi Y. (1985). Empirical distributions in selection bias models. *The Annals of Statistics*, **13**, 178–203.

Wei W., Zhou Y. (2016). Semiparametric maximum likelihood estimation for a two-sample density ratio model with right-censored data. *The Canadian Journal of Statistics*, **44**, 58–81.