האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

CROSS-SECTIONAL SAMPLING, BIAS, DEPENDENCE, AND COMPOSITE LIKELIHOOD

By

MICHA MANDEL and YOSEF RINOTT

Discussion Paper # 614 July 2012

מרכז לחקר הרציונליות

CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Cross-Sectional Sampling, Bias, Dependence, and Composite Likelihood

Micha Mandel and Yosef Rinott

The Hebrew University of Jerusalem

July 9, 2012

Abstract

A population that can be joined at a known sequence of discrete times is sampled cross-sectionally, and the sojourn times of individuals in the sample are observed. It is well known that cross-sectioning leads to length-bias, but less well known that it may result also in dependence among the observations, which is often ignored. It is therefore important to understand and to account for this dependence when estimating the distribution of sojourn times in the population.

In this paper, we study conditions under which observed sojourn times are independent and conditions under which treating observations as independent, using the product of marginals in spite of dependence, results in proper inference. The latter is known as the Composite Likelihood approach. We study parametric and nonparametric inference based on Composite Likelihood, and provide conditions for consistency, and further asymptotic properties, including normal and non-normal distributional limits of estimators. We show that Composite Likelihood leads to good estimators under certain conditions, and illustrate that it may fail without them. The theoretical study is supported by simulations. We apply the proposed methods to two data sets collected by cross-sectional designs: data on hospitalization time after bowel and hernia surgeries, and data on service times at our university.

KEY WORDS: Discrete entrance process, Length bias, Poisson cohort distribution, Truncation.

1 Introduction

Consider a population $\mathcal{S}$, e.g., patients in a hospital, that can be joined at fixed and known sequence of time-points. The population is cross-sectioned at a random time and individuals in $\mathcal{S}$ present at that time comprise the sample. The main aim is to estimate the sojourn time distribution G of individuals in $\mathcal{S}$. Cross-sectioning biases G, and also results in thinned information on the entrance process. Moreover, as discussed below, sojourn times in the cross-sectional sample may be dependent even when sojourn times in $\mathcal{S}$ are independent. This dependence, which seems to be overlooked in part of the literature, plays an important role in this paper.

Let $N_1, \ldots, N_K$ denote the random numbers of individuals who enter $\mathcal{S}$ at known fixed time points, $-a_K < \cdots < -a_1 \le 0$, where 0 is set to be the cross-sectioning time, with corresponding sojourn times $X_{ki} \sim G$, $i = 1, \ldots, N_k$, $k = 1, \ldots, K$. Figure 1 illustrates the setup with $K = 4$, and $(N_1, N_2, N_3, N_4) = (3, 5, 2, 3)$. It is sometimes simpler to replace the index ki of individuals with a single noninformative index j in a random permutation of the population, and denote the sojourn times of individuals in $\mathcal{S}$ by X_j. Their corresponding ages at time 0, that is, the time from entering $\mathcal{S}$ to sampling time, is then denoted by $A_j \in \{a_1, \ldots, a_K\}$, where A_j may be smaller or larger than X_j. The cross-sectional sample consists of those individuals for whom $A_j \le X_j$. With some abuse of notation, these observations will be denoted by (A_j^*, X_j^*). We shall use the generic notation $X \sim G$, and $X^* \sim G^*$. The observed number of subjects who joined $\mathcal{S}$ at $-a_k$ and are still in $\mathcal{S}$ at time 0 is "thinned" relative to the unobserved N_k, and is denoted by N_k^*. Observed sojourn times are depicted in Figure 1 by solid horizontal lines, unobserved by dashed lines, and $(N_1^*, N_2^*, N_3^*, N_4^*) = (3, 4, 0, 1)$. The total population and sample sizes are $M = \sum_k N_k$ and $M^* = \sum_k N_k^*$. The laws of X_j^* and N_k^* differ from those of X_j and N_k, and will be derived in the sequel.

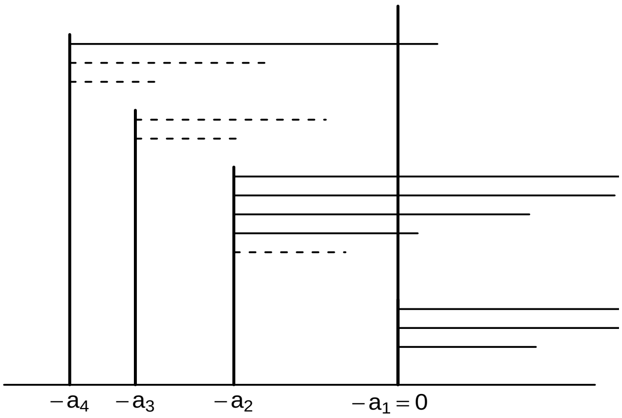

Figure 1: An example with $K = 4$ entrance time-points.

This kind of sampling is frequently studied in the framework of left truncation models; the standard approach for estimating G assumes that conditionally on M^*, the observations (A_j^*, X_j^*) are independent, having the law of $(A, X)|A \leq X$. Inference on G is carried out conditionally on M^* and the ages at sampling, $\{A_j^*\}$, by considering the likelihood $\prod_j \frac{dG(X_j^*)}{\bar{G}(A_j^*-)}$, where $\bar{G} = 1 - G$. This likelihood can be maximized over G in a parametric family $\{G(\cdot\,;\theta)\}$; the nonparametric estimator is based on:

$$\hat{h}_G(t) = \frac{\#\{X_j^* = t\}}{\#\{A_j^* \leq t \leq X_j^*\}}, \tag{1}$$

where $\hat{h}_G$ is an estimator of the hazard function of G. This estimator was briefly mentioned by Kaplan and Meier (1956), and was further studied by Lynden-Bell (1971), Woodroofe (1985), Wang, Jewell and Tsai (1986) and others. We refer to (1) as the conditional estimator since it is based on conditioning on ages at sampling and assumes nothing on the ages' law.

In general $\{A_j^*\}$ are not ancillary, that is, their distribution depends on G. Therefore conditioning on them may lead to loss of information that can be useful under suitable assumptions on the entrance process.

The most common unconditional approach assumes that given $M^* = m^*$, $X_1^*, \ldots, X_m^*$, are independent observations from the length biased distribution $dG^*(t) = tdG(t)/EX$, and $A_j^* \mid X_j^* = x \sim \mathrm{U}(0, x)$. The likelihood $\prod_j dG(X_j^*)$ can be maximized over a parametric family, while the nonparametric estimator is $\hat{G}(t) = \sum_{j=1}^{m^*} \left(X_j^*\right)^{-1} \mathrm{I}\{X_j^* \le t\} / \sum_{j=1}^{m^*} \left(X_j^*\right)^{-1}$, where $\mathrm{I}\{\cdot\}$ denotes an indicator function (see Cox, 1969, or Vardi, 1985). The assumptions on X_j^* and A_j^* are justified when entrances to $\mathcal{S}$ follow a homogeneous Poisson process or when the distribution of A is uniform, see Laslett (1982), Vardi (1989), and Asgharian, M'Lan and Wolfson (2002). The Poisson process assumption is often unsuitable as an entrance model. For example, entrances to hospitals for operations occur on specific days of the week, with limited variability; the same is true for promotion dates of faculty members in a university, which in ours occur once a year. These two examples motivated this paper, see Section 5. Hence we study a discrete entrance process, where individuals enter $\mathcal{S}$ at fixed time-points, and show that it possesses several interesting theoretical properties that should affect inference.

We show that in our models observed sojourn times are generally not independent. An important aspect of this paper is the study of Composite Likelihood that in spite of the dependence, considers estimators based on products of marginals. We study these estimators in parametric and nonparametric models, and conditions for consistency and for asymptotic normality. We indicate situations where Composite Likelihood methods are more efficient than the standard conditional approach.

The paper is organized as follows: In Section 2 we present the model and calculate the likelihood. We show that only under a Poisson assumption sojourn times in the cross-sectional population are independent. We then derive the relation between the sojourn time marginal distributions in the population and the cross-sectional sample. In Section 3, we provide estimators for the sojourn time distribution under the Poisson assumption on the entrance process, and in Section 4 we study inference for more general models, based on the concept of Composite Likelihood. The method is applied to two data sets in Section 5. A simulation study that compares the conditional and unconditional approaches under various models for the entrance process is reported in Section 6. Section 7 concludes the paper with several remarks. Proofs are given in the Appendix.

2 Model and Likelihood

2.1 Discrete Entrance Process

The assumption below on the sojourn times is made throughout the paper. Assumptions on the entrance process $(N_1, \ldots, N_K)$ are specified when needed.

Assumption 2.1. *The sojourn times* $\{X_j\}$ *are iid with* $X_j \sim G$*, independently of the entrance process* $N_1, \ldots, N_K$*.*

The cross-sectional sample does not include duration times smaller than a_1 and therefore $G(x)$ is not estimable for $x \leq a_1$ and the estimable function is $P(X \leq x)/P(X \geq a_1)$. For notational convenience, we henceforth assume that $a_1 = 0$. The data $\{(a_k^*, x_{ki}^*) : k = 1, \ldots, K, i = 1, \ldots, n_k^*\}$ comprise the ages and sojourn times of the $m^* = \sum_k n_k^*$ individuals in $\mathcal{S}$ at time 0, where n_k^* is the number of individuals of age a_k^* in the cross-sectional sample, having observed sojourn times x_{ki}^*. The likelihood is given by

$$\mathcal{L}_k = \sum_{\mathbf{n} \geq \mathbf{n}^*} P(\mathbf{N} = \mathbf{n}) \prod_{k=1}^{K} \Big\{ \binom{n_k}{n_k^*} G(a_k-)^{n_k - n_k^*} \prod_{i=1}^{n_k^*} dG(x_{ki}^*) \Big\}, \tag{2}$$

where boldface $\mathbf{N}, \mathbf{n}$ and $\mathbf{n}^*$ denote K-vectors, e.g., $\mathbf{N} = (N_1, \ldots, N_K)$. Unless a simple parametric model for $\mathbf{N}$ is assumed, the likelihood (2) is complex and $P(\mathbf{N} = \mathbf{n})$ contributes a large number of nuisance parameters.

If the actual numbers of individuals entering at the K time-points $n_1, \ldots, n_K$ were known, the likelihood would be proportional to

$$\prod_{k=1}^{K} G(a_k-)^{n_k-n_k^*} \prod_{k=1}^{K}\prod_{i=1}^{n_k^*} G(dx_{ki}^*) = \prod_{k=1}^{K} G(a_k-)^{n_k} \prod_{k=1}^{K}\prod_{i=1}^{n_k^*} \frac{G(dx_{ki}^*)}{G(a_k-)}, \tag{3}$$

so the left-hand side of (3) reduces to likelihood of left censored data. However, such knowledge is not assumed here. Ignoring the unknown quantity $\prod_{k=1}^{K} G(a_k-)^{n_k}$ on the left-hand side of (3), and maximizing only the rightmost product, yields the conditional maximum likelihood estimator, whose hazard function is given by (1).

2.2 Independence

Equation (2) shows that, in general, individual sojourn times in the sample are not independent. Since conditions for independence are often not stated clearly in the literature, it is important to characterize cases where it holds. Proofs are given in the Appendix.

Theorem 2.1. *If $N_1, \ldots, N_K$ are independent, then the pairs $(A_1, X_1), \ldots, (A_M, X_M)$ are independent conditionally on M, if and only if $N_k \sim \text{Poisson}(\lambda_k),\ k = 1, \ldots, K$.*

This independence carries over to the cross-sectional sample.

Corollary 2.1. *If $N_1, \ldots, N_K$ are independent, then the pairs $(A_1^*, X_1^*), \ldots, (A_{M^*}^*, X_{M^*}^*)$ are independent conditionally on M^*, if and only if $N_k \sim \text{Poisson}(\lambda_k),\ k = 1, \ldots, K$.*

The sojourn times $\{X_j\}$ are always independent by Assumption 2.1. However, only under the independent Poisson entrance process, the sojourn times in the cross-sectional sample, $\{X_j^*\}$, are independent. On the other hand, conditionally on $\{A_j^*\}$, the residual sojourn times $\{X_j^* - A_j^*\}$ are independent under any distribution of $\{N_k\}$. Hence, the conditional approach that is based on (1) is robust with respect to the entrance process model. This will be demonstrated further in the simulation study of Section 6.

The importance of the theorem and corollary above is that when entrances are according to a Poisson distribution, the complicated likelihood (2) can be replaced by a product of marginals, and inference becomes much simpler, see Section 3, but one should be aware of the assumptions behind such simplification. The consequences of using a product of marginals when the Poisson assumption fails are studied in Section 4.

2.3 The Marginal Distribution

We study the distribution of age of a subject j chosen at random among those who entered $\mathcal{S}$ at one of the points $-a_K, \ldots, -a_1$; if no such subject exists because $N_1 = \cdots = N_K = 0$, we formally define $A_j = \infty$.

Let $\eta_k = E(N_k / \sum_s N_s)$, where we define $0/0 = 0$. Then

$$P(A_j = a_k) \quad = \quad \sum_{\mathcal{N}} \frac{n_k}{\sum_{s=1}^K n_s} P(N_1 = n_1, \ldots, N_K = n_K) = \eta_k,$$

where the sum is over $\mathcal{N} = \Big\{(n_1, \ldots, n_K) : \sum_{k=1}^K n_k \geq 1 \ , \ n_k \in \{0, 1, 2, \ldots\} \quad k = 1, \ldots, K\Big\}$.

Recall that $X_j \sim G$ and that A_j and X_j are independent by assumption 2.1. The joint density of (A_j^*, X_j^*), the age at sampling and sojourn time for the jth subject in the cross-sectional sample (i.e., under the condition $A_j \leq X_j$) is given by

$$P(A_j^* = a_k, X_j^* \in [x, x+dx]) = P(A_j = a_k, X_j \in [x, x+dx] \mid A_j \leq X_j) \underset{dx \downarrow 0}{\longrightarrow} \frac{dG(x)\eta_k}{\beta} \mathrm{I}\{a_k \leq x\}, \tag{4}$$

where $\beta = P(A_j \leq X_j) = \sum_{k=1}^K \eta_k \bar{G}(a_k-)$.

The marginal density of an observed sojourn time is obtained by summing over a_k:

$$dG^*(x) = \frac{\sum_{k|a_k \leq x} \eta_k dG(x)}{\beta} = \frac{w(x) dG(x)}{\beta}, \tag{5}$$

a weighted version of G, where the weight is the step-function $w(x) = \sum_{k|a_k \leq x} \eta_k$.

The derivation of (5) holds for a general joint distribution of $N_1, \ldots, N_K$. Of special interest in this paper is the case $\eta_k = 1/K$, which holds, for example, when $N_1, \ldots, N_K$ are exchangeable. In this case, the right-hand sides of (4) and (5) reduce to

$$(a) \quad \frac{\frac{1}{K} dG(x)}{\beta} \mathrm{I}\{a_k \leq x\} = \frac{dG(x)}{\mu} \mathrm{I}\{a_k \leq x\}, \quad (b) \quad dG^*(x) = \frac{\lfloor x \rceil dG(x)}{\mu}, \tag{6}$$

where $\lfloor x \rceil = \sum_{k=1}^{K} \mathrm{I}\{a_k \le x\} = \max\{k | a_k \le x\}$, and $\mu = K\beta = E\lfloor X \rceil = \sum_{k=1}^{K} \bar{G}(a_k-)$. Thus, the weight function at x is the potential number of time-points at which an individual having sojourn time of length x can enter the population and still be included in the sample. When $a_k = k$ and $\mathrm{supp}(G) = \{1, 2, \ldots\}$, G^* is the standard length-biased version of G.

The following corollary summarizes the results of Sections 2.2 and 2.3 and shows that for a discrete entrance process, the standard assumptions of independence and length-biased sojourn times hold only under rather restrictive conditions.

Corollary 2.2. *Assume $N_1, \ldots, N_k$ are independent. Sojourn times in the cross-sectional sample are independent conditionally on $\{M^* = m^*\}$, and have the length-biased type distribution G^* of (6) if and only if $N_k \sim \mathrm{Poisson}(\lambda)$ for $k = 1, \ldots, K$ and some $\lambda > 0$.*

3 Inference Under the Poisson Assumption

3.1 The Homogeneous Case

Due to Corollary 2.2, the model of independent $N_k \sim \mathrm{Poisson}(\lambda)$ for $k = 1, \ldots, K$ is attractive. Next, we discuss maximum likelihood estimation of (G, λ) for this model.

The likelihood of the data is a product of terms as (6a) multiplied by the likelihood of having m^* observations which is $e^{-\lambda\mu}[\lambda\mu]^{m^*}/m^*!$, yielding

$$\mathcal{L}(G, \lambda) = e^{-\lambda\mu}[\lambda\mu]^{m^*}/m^*! \prod_{j=1}^{m^*} \frac{\lfloor x_j^* \rceil dG(x_j^*)}{\mu} \prod_{i=j}^{m^*} \left(\lfloor x_j^* \rceil\right)^{-1}.$$

Laslett (1982) and many others considered a continuous Poisson entrance process, and arrived at a likelihood similar to the above, with a length bias weight x rather than $\lfloor x \rceil$.

For any μ, $\mathcal{L}(G, \lambda)$ is maximized by $\hat{\lambda} = m^*/\mu$, reducing $e^{-\lambda\mu}[\lambda\mu]^{m^*}/m^*!$ to $e^{-m^*}m^{*m^*}/m^*!$. Thus, estimation is done by first maximizing $\prod_{j=1}^{m^*} \lfloor x_j^* \rceil dG(x_j^*)/\mu$ for G and then using the resulting $\hat{\mu}$ to estimate λ by $\hat{\lambda} = m^*/\hat{\mu}$. Parametric inference for G is straightforward, and general maximum likelihood principles apply. For nonparametric inference, note that the likelihood at $(G, \hat{\lambda})$ is proportional to a product of weighted densities so the resulting estimator is (e.g., Vardi, 1985):

$$\hat{G}(t) = \frac{\sum_{j=1}^{M^*} \left(\lfloor X_j^* \rceil\right)^{-1} \mathrm{I}\{X_j^* \le t\}}{\sum_{j=1}^{M^*} \left(\lfloor X_j^* \rceil\right)^{-1}}. \tag{7}$$

Writing

$$\sqrt{M^*}(\hat{G}(t) - G(t)) = \frac{\frac{1}{\sqrt{M^*}} \sum_{j=1}^{M^*} \left(\lfloor X_j^* \rceil\right)^{-1} \left(\mathrm{I}\{X_j^* \le t\} - G(t)\right)}{\frac{1}{M^*} \sum_{j=1}^{M^*} \left(\lfloor X_j^* \rceil\right)^{-1}},$$

we obtain consistency of $\hat{G}$, and asymptotic normality of $\sqrt{M^*}(\hat{G}(t) - G(t))$, with zero mean and a variance that can be estimated by $\hat{\mu}^2 \frac{1}{M^*} \sum_{j=1}^{M^*} \left\{ \left(\lfloor X_j^* \rceil\right)^{-1} \left(\mathrm{I}\{X_j^* \le t\} - \hat{G}(t)\right) \right\}^2$.

3.2 The Inhomogeneous Case

When $N_1, \ldots, N_k$ are independent and $N_k \sim \text{Poisson}(\lambda_k)$ then $\eta_k = E\left(N_k / \sum_s N_s\right) = \lambda_k / \sum_s \lambda_s$, and (4) becomes

$$P(A_j^* = a_k, x \le X_j^* < x + dx) \underset{dx \downarrow 0}{\longrightarrow} \frac{\lambda_k dG(x)}{\sum_{\ell=1}^{K} \lambda_\ell \bar{G}(a_\ell -)} I\{a_k \le x\}. \tag{8}$$

Also, $M^* \sim \text{Poisson}(\sum_{k=1}^{K} \lambda_k \bar{G}(a_k -))$, and the observations are independent by Corollary 2.1. The likelihood of the data is readily computed to be

$$\mathcal{L}(G, \lambda_1, \ldots, \lambda_k) = \prod_{k=1}^{K} \frac{e^{-\lambda_k \bar{G}(a_k -)} \{\lambda_k \bar{G}(a_k -)\}^{n_k^*}}{m^*!} \times \prod_{k=1}^{K} \prod_{i=1}^{n_k^*} \frac{dG(x_{ki}^*)}{\bar{G}(a_k -)}.$$

Obviously, $\hat{\lambda}_k = n_k^* / \hat{\bar{G}}(a_k -)$, and plugging it in shows that the MLE of G is obtained by maximizing the right term of the likelihood above. This term is just the conditional likelihood for which (1) is the hazard function of the solution so in this case there is no gain in conducting the unconditional approach. This is similar to the product-limit estimator being the NPMLE of truncation models when the distribution of the truncation variable is completely unspecified (see Wang, 1991).

Sometimes, auxiliary data are available and the rates $\lambda_1, \ldots, \lambda_K$ are known up to a constant, say $\lambda_k = \lambda\alpha_k$, where $\alpha_1, \ldots, \alpha_K$ are known. In such cases, the estimator (7) can be applied with $\sum_{k=1}^{K} \alpha_k I\{a_k \leq X_j^*\}$ replacing $\lfloor X_j^* \rceil$.

4 Composite Likelihood

When entrances to $\mathcal{S}$ do not follow a Poisson distribution, sojourn times in the cross-sectional sample are not independent and the likelihood (2) becomes complicated. One can always step back and use the conditional approach. Alternatively, the composite likelihood method, which consists of maximizing the product of the marginals in spite of the dependence of the observations, may provide more efficient yet simple estimators.

Composite likelihood has been studied by many authors in different contexts (see the review paper by Varin, Reid and Firth, 2011), but mostly for the case of many independent samples with dependent observations. For example, standard composite likelihood theory would apply easily when the asymptotics is with respect to the number of independent cross-sectional studies. In our data, and quite typically for such studies, the number of cross-sectional samples is small, and the relevant asymptotics appear to be associated with large entrance numbers N_k. In this section, we explore the implications of maximizing the composite likelihood in our scenario where there is only one sample comprising of many dependent observations.

We analyzed above the case of independent Poisson entrance process, which reduces to conditional inference in the inhomogeneous case. We now relax the Poisson assumption, but maintain homogeneity and assume that $EN_1 = \cdots = EN_K$, and $E\left(N_k / \sum_s N_s\right) = 1/K$ for all k. Rather than repeat these conditions, we will make the stronger assumption that the N_k's are exchangeable, although the above weaker conditions suffice. Our goal is to understand the conditions under which composite likelihood works or fails to work in situations where it is expected to perform better than the conditional approach.

The composite likelihood is a product of the marginals of the observed sojourn times as in (6), and is therefore proportional to the product :

$$L(G) = \prod_{j=1}^{m^*} dG^*(x_j^*) = \prod_{j=1}^{m^*} \frac{\lfloor x_j^* \rceil dG(x_j^*)}{\mu}. \tag{9}$$

Our goal is to study the MCompLE (Maximum Composite Likelihood Estimator), that is, $\hat{G} = \arg\max L(G)$, in the presence of dependence among $X_1^*, \ldots, X_{M^*}^*$, where the maximization is over all distributions G or a parametric family. We study consistency and the asymptotic distribution of the MCompLE with respect to sequences of population entrance numbers $N_k^{(\nu)}$ satisfying $N_k^{(\nu)} \to \infty$ in probability as $\nu \to \infty$, that is, $\lim_{\nu\to\infty} P(N_k^{(\nu)} \geq n) = 1$ for all n. Without loss of generality, we assume that the sequences are parameterized so that $EN_k^{(\nu)} = \nu$. For simplicity of notation, we omit the superscript ν in the sequel, and expressions like $N_k/EN_k \xrightarrow{p} 1$ are taken with respect to $\nu \to \infty$.

An important simple device that will be used below is a representation of the log composite likelihood $\ell(G) = \log L(G) = \sum_{j=1}^{M^*} \log dG^*(X_j^*)$ as a sum of independent random variables:

$$\ell(G) = \sum_{k=1}^{K} \sum_{i=1}^{N_k} I_{\{X_{ki} \geq a_k\}} \log \frac{\lfloor X_{ki} \rceil dG(X_{ki})}{\mu} =: \sum_{k=1}^{K} \sum_{i=1}^{N_k} h_k(X_{ki}). \tag{10}$$

Recall that X_{ki} are iid having the law G.

4.1 Parametric Models

Suppose $G = G(\cdot; \theta)$ is indexed by a parameter $\theta \in \Theta \subset \mathbb{R}$ taken to be univariate for simplicity (an extension to the multiparameter case is discussed in the appendix). We write $\ell(\theta)$, μ_θ, and $h_k(\cdot; \theta)$ instead of $\ell(G)$, μ, and $h_k(\cdot)$ of (10). The next theorem deals with consistency of the MCompLE; the proof is given in the Appendix.

Theorem 4.1. *Suppose $N_1, \ldots, N_K$ are exchangeable random variables, independent of $\{X_{kj}\}$, having distribution that depends on a parameter ν and parameterized so that $EN_k = \nu$ and $N_k/EN_k \xrightarrow{p} 1$. Assume $dG(x;\theta)$ is differentiable with respect to θ for all x, and the standard regularity conditions of identifiability, common support of $dG(x;\theta)$ for all θ, and the true parameter θ_0 being an interior point of the parameter space. Then there exists a consistent sequence $\hat{\theta}_\nu$ of roots of the composite likelihood equation $\frac{\partial}{\partial\theta}\ell(\theta) = 0$.*

The regularity conditions above appear as (A0), (A1), and (A3) in Lehmann and Casella (1998, p. 444-5); in Condition (A2), the iid assumption is relaxed, maintaining only independence of sojourn times in the populations; see Assumption 2.1.

When the composite likelihood equation has a unique root $\hat{\theta}$, it is the MCompLE, and the resulting sequence is consistent. In particular, if G belongs to a canonical exponential family, then so does $G^*(\cdot;\theta)$ and if the MCompLE exists, it is unique. Also, if the parameter space is compact, then any sequence of maxima of the composite likelihood, that is, any MCompLE sequence, is consistent. See Ferguson (1996).

Cohort size distributions such as Poisson(ν) or Binomial(ν, θ) satisfy the conditions of Theorem 4.1, while Geometric($1/\nu$) or Uniform($1, \nu$) do not. When there is a positive probability of one cohort being much larger than the others, the MCompLE becomes inconsistent. The following example demonstrates this point.

Example 4.1 (Inconsistency of the MCompLE)**.** *Consider the model $X_{kj} \sim \mathrm{Exp}(\theta)$ with $K = 2$, and $a_1 = 0$, $a_2 = 1$. The MCompLE, $\hat{\theta}$, solves the equation*

$$\frac{1}{\hat{\theta}} + \frac{e^{-\hat{\theta}}}{1 + e^{-\hat{\theta}}} = \frac{N_1^*}{N_1^* + N_2^*}\bar{X}_1^* + \frac{N_2^*}{N_1^* + N_2^*}\bar{X}_2^*,$$

where $\bar{X}_k^* = (N_k^*)^{-1}\sum_{i=1}^{N_k^*} X_{ki}^*$, $(k = 1, 2)$. As $\nu \to \infty$, $\bar{X}_k^* \xrightarrow{p} k - 1 + \theta^{-1}$, $k = 1, 2$, and

$N_k^*/N_k \xrightarrow{p} e^{-\theta a_k}$, so the estimating equation is approximately

$$\frac{1}{\hat{\theta}} + \frac{e^{-\hat{\theta}}}{1 + e^{-\hat{\theta}}} \approx \frac{1}{\theta} + \frac{e^{-\theta}}{\frac{N_1}{N_2} + e^{-\theta}}.$$

The MCompLE is consistent if $N_1/N_2 \xrightarrow{p} 1$, but not otherwise. As a concrete example let $N_k = \nu$ or 2ν with probability 1/2 each so that N_1/N_2 takes the values 0.5, 1, and 2 with corresponding probabilities 1/4, 1/2, and 1/4, and the estimator converges to a non-degenerate random variable, and therefore is inconsistent.

By Markov Inequality, $P(|\frac{N_k}{EN_k} - 1| > \varepsilon) \leq \text{Var}N_k/(\varepsilon EN_k)^2$ so that models where the coefficient of variation vanishes as $\nu \to \infty$, e.g., Poisson(ν), ensure $N_k/EN_k \xrightarrow{p} 1$. This is not the case in Example 4.1, where the coefficient of variation is 1/3.

The analysis of the asymptotic distribution of the MCompLE is similar to that of Theorem 3.10 in Lehmann and Casella (1998, page 449); interestingly, the limiting distribution of the MCompLE is not necessarily normal. Conditions for normality and examples of non normal limits are given below.

Theorem 4.2. *Assume all the conditions of Theorem 4.1, and additionally assume that in some neighborhood of θ_0 : (i) $dG(x;\theta)$ is differentiable three times with respect to θ for all x; (ii) $\int dG(x;\theta)$ can be twice differentiated under the integral; (iii) the Fisher information $-E_\theta[\frac{\partial^2}{\partial\theta^2} \log dG(X;\theta)] \in (0, \infty)$; (iv) $|\frac{\partial^3}{\partial\theta^3} \log dG(x;\theta)| < \psi(x)\ \forall x$, where $E_{\theta_0}\psi(X) < \infty$.*

Letting $U = U^{(\nu)} = \sum_{k=1}^K c_k \frac{N_k - \nu}{\sqrt{\nu}}$, where $c_k = E_{\theta_0}(\frac{\partial}{\partial\theta} h_k(X;\theta_0))$, we have $E_{\theta_0} U^{(\nu)} = 0$. If $U^{(\nu)} \xrightarrow[\nu\to\infty]{\mathcal{D}} V$ for some random variable V, then for any consistent sequence $\hat{\theta}_\nu$ of roots of the composite likelihood equation $\frac{\partial}{\partial\theta}\ell(\theta) = 0$,

$$\sqrt{M^*}(\hat{\theta}_\nu - \theta_0) \xrightarrow[\nu\to\infty]{\mathcal{D}} \frac{\sqrt{\mu_{\theta_0}}}{\sum_{k=1}^K E_{\theta_0} \frac{\partial^2}{\partial\theta^2} h_k(X,\theta_0)} (W + V), \tag{11}$$

where $W \sim N\Big(0, \sum_{k=1}^K \text{Var}_{\theta_0}(\frac{\partial}{\partial\theta} h_k(X;\theta_0))\Big)$ is independent of V, and the resulting asymptotic variance of $\sqrt{M^*}(\hat{\theta}_\nu - \theta_0)$ is

$$\mu_{\theta_0} \frac{\sum_{k=1}^K \text{Var}(\frac{\partial}{\partial\theta} h_k(X;\theta_0)) + \text{Var}(V)}{\{\sum_{k=1}^K E \frac{\partial^2}{\partial\theta^2} h_k(X,\theta_0)\}^2}. \tag{12}$$

Sketch of proof: Using consistency, Taylor expansion of $0 = \frac{\partial}{\partial\theta}\ell(\hat{\theta}_\nu)$ around θ_0, and standard arguments yield the approximation

$$\sqrt{M}(\hat{\theta}_\nu - \theta_0) \approx \sqrt{K} \cdot \frac{\frac{\sqrt{K}}{\sqrt{M}} \sum_{k=1}^{K} \sum_{i=1}^{N_k} \frac{\partial}{\partial\theta} h_k(X_{ki};\theta_0)}{\frac{-K}{M} \sum_{k=1}^{K} \sum_{i=1}^{N_k} \frac{\partial^2}{\partial\theta^2} h_k(X_{ki};\theta_0)}. \tag{13}$$

The conditions on $N_1, \ldots, N_K$ imply $N_k/M \xrightarrow{p} 1/K$ and the denominator of (13) converges to $\sum_{k=1}^{K} E_{\theta_0} \frac{\partial^2}{\partial\theta^2} h_k(X, \theta_0)$. The analysis of the numerator is more complicated:

$$\frac{1}{\sqrt{M/K}} \sum_{k=1}^{K} \sum_{i=1}^{N_k} \frac{\partial}{\partial\theta} h_k(X_{ki};\theta_0) = \sum_{k=1}^{K} \frac{1}{\sqrt{M/K}} \sum_{i=1}^{N_k} \Big(\frac{\partial}{\partial\theta} h_k(X_{ki};\theta_0) - c_k \Big) + \sum_{k=1}^{K} \frac{c_k N_k}{\sqrt{M/K}}. \tag{14}$$

A multivariate version of the proof in Rényi (1957) of Anscombe's theorem implies that

$$\frac{1}{\sqrt{M/K}} \sum_{i=1}^{N_k} \Big(\frac{\partial}{\partial\theta} h_k(X_{ki};\theta_0) - c_k \Big)$$

converge jointly for $k = 1, \ldots, K$ to independent mean zero normal variables.

Therefore, in (14),

$$\sum_{k=1}^{K} \frac{1}{\sqrt{M/K}} \sum_{i=1}^{N_k} \Big(\frac{\partial}{\partial\theta} h_k(X_{ki};\theta_0) - c_k \Big) \xrightarrow{\mathcal{D}} W, \tag{15}$$

where $W \sim N\Big(0, \sum_{k=1}^{K} \mathrm{Var}_{\theta_0}(\frac{\partial}{\partial\theta} h_k(X;\theta_0))\Big)$.

By the facts that $\sum_k c_k = 0$, as proved in the Appendix, and $\nu^{-1}M/K \xrightarrow{p} 1$, we have $\sum_{k=1}^{K} c_k N_k/\sqrt{M/K} - U \xrightarrow{p} 0$, so the last term in (14) converges to V. Now, (11) is obtained from (13) by

$$\frac{M^*}{M} = \sum_{k=1}^{K} \frac{N_k}{M} \frac{1}{N_k} \sum_{j=1}^{N_k} I\{X_{kj} \geq a_k\} \xrightarrow{p} \frac{1}{K} \sum_{k=1}^{K} \bar{G}(a_k-;\theta_0) = \frac{1}{K} \mu_{\theta_0}. \tag{16}$$

Independence of V and W is shown in the Appendix, completing the proof.

We next discuss several models which lead to different results regarding the asymptotic distribution of $\sqrt{M^*}(\hat{\theta}_\nu - \theta_0)$.

Independent N_k's, Normal limit. Theorem 4.2 implies that $\sqrt{M^*}(\hat{\theta}_\nu - \theta_0)$ is asymptotically normal if V is a normal random variable, possibly degenerate. This condition is also necessary by Cramér's Theorem, e.g., Feller (1971) p. 525, which says that a sum of independent random variables has a normal distribution if and only if the summands are normal. Several conditions characterizing such cases are discussed next.

Suppose that N_k's are iid with $E(N_k) = \nu$ and $\mathrm{Var}(N_k) = \sigma^2$, then

$$U^{(\nu)} = \sum_{k=1}^{K} \frac{c_k(N_k - \nu)}{\sigma} \times \sqrt{\frac{\sigma^2}{\nu}} \, .$$

If $(N_k - \nu)/\sigma \xrightarrow{\mathcal{D}} N(0,1)$ for $k = 1, \ldots, K$, and $\sigma^2/\nu \xrightarrow{p} b$, then $V \sim N(0, b \sum c_k^2)$. This includes the cases where N_k has a Poisson or Binomial distribution. By Cramér's Theorem, as $\{c_k(N_k - \nu)/\sigma\}$ are independent, the above condition covers all cases of independent N_k's. If $b = 1$, e.g., $N_k \sim \text{Poisson}(\nu)$, then recalling $c_k = E(\frac{\partial}{\partial\theta} h_k(X;\theta_0))$, the numerator of (12) reduces to $\sum_{k=1}^{K} E\left\{\frac{\partial}{\partial\theta} h_k(X,\theta_0)\right\}^2$. It can be shown that the latter expression equals $-\sum_{k=1}^{K} E\left\{\frac{\partial^2}{\partial\theta^2} h_k(X,\theta_0)\right\}$, where equality holds for the sum, but not term by term. Therefore (12) becomes

$$-1/\sum_{k=1}^{K} E\left\{\frac{\partial^2}{\partial\theta^2} h_k(X,\theta_0)\right\} = 1/\sum_{k=1}^{K} E\left\{\frac{\partial}{\partial\theta} h_k(X,\theta_0)\right\}^2$$

which has interpretation in terms of Fisher's information.

If $b = 0$, e.g., when $N_k \equiv \nu$, then $\sum_{k=1}^{K} c_k(N_k - \nu)/\sqrt{\nu} \xrightarrow{\mathcal{D}} 0$ and therefore $V = 0$.

Dependent N_k's. As a simple but natural example of a normal limit in the presence of dependence, let N_0' be independent of $N_1', \ldots, N_K'$ with $EN_k' = \nu'$, and suppose the entrance numbers are $N_k = N_0' + N_k'$. Since $\sum_k c_k = 0$, we have $U = \sum_{k=1}^{K} c_k(N_k - \nu)/\sqrt{\nu} = \frac{\sqrt{\nu'}}{\sqrt{\nu}} \sum_{k=1}^{K} c_k(N_k' - \nu')/\sqrt{\nu'} := \frac{\sqrt{\nu'}}{\sqrt{\nu}} V'$. If $\nu'/\nu \to c$ as $\nu \to \infty$, then $U \xrightarrow{\mathcal{D}} \sqrt{c}\, V'$, which is normal if $N_1', \ldots, N_K'$ are such that V' is normal.

It is easy to construct models as above having the same marginal distribution of the N_k's, but with different values of c, showing that the dependence structure influences the asymptotic distribution of the MCompLE. This is in contrast to its consistency, which by Theorem 4.1 depends only on the marginal distribution of the cohort sizes.

Another natural model of dependent N_k's that leads to asymptotic normality is the following. Let $M = M^{(\nu)}$ satisfy $EM = K\nu$, and $M/\nu \xrightarrow{p} K$, corresponding to the assumptions of Theorem 4.2. Suppose $(N_1, \ldots, N_k) \mid M \sim \text{Multinomial}(M, (\frac{1}{K}, \ldots, \frac{1}{K}))$, then V is Gaussian. To see this, write $N_k = \sum_{j=1}^{M} I\{Z_j = e_k\}$, where $Z_j \sim \text{Mult}(1, (\frac{1}{K}, \ldots, \frac{1}{K}))$ and e_k is a vector of K coordinates with the kth being 1 and the rest 0. Therefore $\sum_{k=1}^{K} c_k(N_k - \nu)/\sqrt{\nu} = \sum_{j=1}^{M} \sum_{k=1}^{K} c_k I\{Z_j = e_k\}/\sqrt{\nu}$, which converges to the normal distribution by Anscombe's Theorem, see Rényi (1957).

Non-Normal limit. The limit of $\sum_{k=1}^{K} c_k(N_k - \nu)/\sqrt{\nu}$ may not be Normal or 0, and may not exist. Let N_1, N_2 be independent with $E(N_k) = \nu$, where $\nu = 1, 2, \ldots$, and assume that $P(N_k = \nu - a) = P(N_k = \nu + a) = 0.5$ for some $a = a(\nu)$. In order that $N_k/E(N_k) \to 1$, a must satisfy $a/\nu \to 0$. Now, $c1 + c2 = 0$ implies $\sum_{k=1}^{2} c_k(N_k - \nu)/\sqrt{\nu} = c_1(N_1 - N_2)/\sqrt{\nu}$, which takes the values 0 or $\pm 2ac_1/\sqrt{\nu}$. For $a = \sqrt{\nu}$ the limiting distribution is neither degenerate nor Normal, and for $a = (2 + (-1)^{\nu})\sqrt{\nu}$ the limit does not exist.

We complete the section by demonstrating, via a simple example, the utility of using the MCompLE.

Example - efficiency of the MCompLE. Consider the case of a fixed cohort size, $N_k = N$, $k = 1, \ldots, K$, and the model $X_{kj} \sim \text{Exp}(\theta)$. Obviously, this model satisfies the regularity conditions required by Theorem 4.2 and $V \equiv 0$. Here, we study the asymptotic variance of the McompLE and compare it to the asymptotic variances of the maximum conditional likelihood estimator and of the maximum likelihood estimator under the assumption that the number of entrances N is known, see the discussion around Equation (3). The latter estimator cannot be calculated in standard cross-sectional sampling setting as N is not known, but the comparison demonstrates the effect of different kinds of information on the estimators.

Some elementary but tedious calculations of the terms of (12) show that the asymptotic variance of the McompLE is

$$\frac{\sum_{k=1}^{K}\left\{\theta^{-2}e^{-\theta a_k}+(a_k-\bar{a}(\theta))^2e^{-\theta a_k}(1-e^{-\theta a_k})\right\}}{\left[\left\{\theta^{-2}+\mathrm{var}_\theta(a)\right\}\sum_{k=1}^{K}e^{-\theta a_k}\right]^2},$$

where $\bar{a}(\theta)$ and $\mathrm{var}_\theta(a)$ are the mean and variance of a discrete random variable taking the values a_k with probabilities $p(a_k) \propto \exp(-\theta a_k)$ $(k = 1, \ldots, K)$. The variance of the conditional likelihood estimator is $\theta^2\{\sum_k \exp(-\theta a_k)\}^{-1}$. If N is known, the likelihood is of left censored data with $N - N_k^*$ observations censored at a_k with asymptotic variance $[\sum_{k=1}^{K}\{a_k^2/(1-\exp(-\theta a_k)) + \theta^{-2}\}\exp(-\theta a_k)]^{-1}$.

Figure 2 compares the asymptotic standard errors of the three approaches discussed above. It reveals that the composite likelihood approach is more efficient than the conditional likelihood approach for a large range of parameter values. The asymptotic standard deviation is increased for the MCompLE by less than 20% relative to the standard deviation of the estimator based on known N, and by more than 60% for certain values of θ for the conditional estimator. Exploring the asymptotic terms for large values of θ, it is seen that the leading terms are $\theta^2 e^{\theta a_1}$ for both the conditional and the composite likelihood approaches, and $\theta^2 e^{\theta a_1}/(a_1^2\theta^2 + 1)$ for the known N case. This explains the behavior of the two ratios for $a_1 = 0$, where they approach 1 for large θ.

4.2 Nonparametric Models

In the nonparametric case, the composite likelihood (9) has the form of a likelihood under biased sampling and is maximized by the inverse weighting estimator (e.g., Vardi, 1985):

$$\hat{G}(x) = \frac{\sum_{k=1}^{K}\sum_{i=1}^{N_k^*}\frac{1}{\lfloor X_{ki}^*\rceil}I_{\{X_{ki}^*\le x\}}}{\sum_{k=1}^{K}\sum_{i=1}^{N_k^*}\frac{1}{\lfloor X_{ki}^*\rceil}}.$$

This estimator can be represented in terms of the independent and identically distributed variables X_{ki} by

$$\hat{G}(x) = \frac{\frac{1}{M}\sum_{k=1}^{K}\sum_{i=1}^{N_k}\frac{1}{\lfloor X_{ki}\rceil}I_{\{a_k\le X_{ki}\le x\}}}{\frac{1}{M}\sum_{k=1}^{K}\sum_{i=1}^{N_k}\frac{1}{\lfloor X_{ki}\rceil}I_{\{a_k\le X_{ki}\}}}. \tag{17}$$

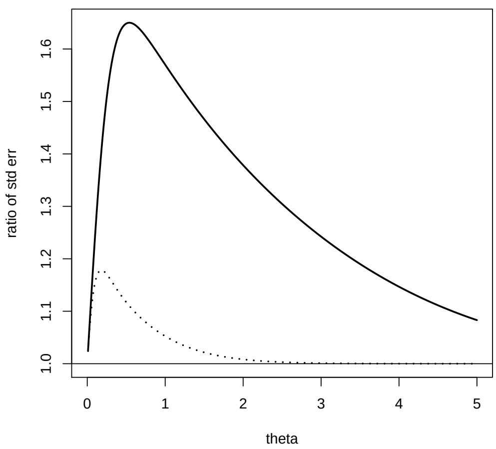

Figure 2: The ratio of the asymptotic standard errors of the conditional (solid) and composite (dotted) estimators to the MLE based on a known fixed N, for $G(t) = 1 - e^{\theta t}$ and $a_k = k$, $k = 0, \ldots, 10$.

Theorem 4.3. *Suppose $N_1, \ldots, N_K$ are exchangeable, independent of $\{X_{ki}\}$, having distribution that depends on a parameter ν such that $EN_k = \nu$ and $N_k/\nu \xrightarrow{p} 1$ as $\nu \to \infty$. Then for all x, $\hat{G}(x) \xrightarrow{p} G(x)$ as $\nu \to \infty$.*

Next we study the asymptotic distribution of $\hat{G}(x)$. Recall that for $a_k \le x < a_{k+1}$, $\lfloor x \rceil = k$, where $a_{K+1} := \infty$. For $k = 1, \ldots, \lfloor x \rceil$, define

$$\gamma_k(x) := E \frac{1}{\lfloor X \rceil} I_{\{a_k \le X \le x\}} = \sum_{\ell=k}^{\lfloor x \rceil - 1} \frac{1}{\ell} [G(a_{\ell+1}-) - G(a_\ell -)] + \frac{1}{\lfloor x \rceil} [G(x) - G(a_{\lfloor x \rceil}-)],$$

and let $\gamma_k(x) := 0$ for $k > \lfloor x \rceil$. It follows that $\sum_{k=1}^K \gamma_k(x) = G(x)$.

Theorem 4.4. *Suppose $N_1, \ldots, N_K$ are exchangeable, independent of $\{X_{ki}\}$, having distribution that depends on a parameter ν such that $EN_k = \nu$ and $N_k/EN_k \xrightarrow{p} 1$ as $\nu \to \infty$. For any given x, let $U^{(\nu)}(x) = \sum_{k=1}^K c_k(x) \frac{N_k - \nu}{\sqrt{\nu}}$, where $c_k(x) = \gamma_k(x) - \gamma_k(\infty) G(x)$, we have $EU^{(\nu)}(x) = 0$. If $U^{(\nu)}(x) \xrightarrow[\nu \to \infty]{\mathcal{D}} V(x)$ for some random variable $V(x)$,*

$$\sqrt{M^*}\{\hat{G}(x) - G(x)\} \xrightarrow[\nu \to \infty]{\mathcal{D}} \sqrt{\mu}(W(x) + V(x)), \tag{18}$$

where $W(x) \sim N\left(0, \sum_{k=1}^K \sigma_k^2(x)\right)$ is independent of $V(x)$, and $\sigma_k^2(x) := Var(\frac{1}{\lfloor X \rceil} I_{\{a_k \le X\}} [I_{\{X \le x\}} - G(x)])$.

5 Data Analysis

5.1 Age of Full Professorship

Sojourn times in academic positions is of interest to universities for planning purposes, see for example, Kaminski and Geisler (2012). For the purpose of estimating the service time distributions in the different ranks, and comparing the distribution of age at promotion between men and women and between different faculties, the adminstration of the Hebrew University of Jerusalem (HUJI) collected data on all professors employed in 1998 with a followup of three years during which new promotions were recorded. Cross-sectioning the population in 1998 causes biases which became apparent to the administrators after looking at the average service time. Here we analyze the full professorship rank based on faculty members who were promoted to this rank from 1980 onward, and were at an age ≤ 65 in 1998. We assume that any such person will stay at HUJI until 65 because very few full professors leave the university before that age, and there are no promotions after. We estimate the distribution G of service time at the rank of full professor, from which we extract the distribution of chronological age at promotion, defined as the difference between 65 and the service time. The entrance process here is discrete since in HUJI the date of promotion is October 1st of the academic year the promotion procedure started. Note that one should distinguish between chronological age, whose distribution we are estimating, and age at sampling as defined in this paper, which is the time from promotion to full professor, to 1998. The data comprise 370 individuals.

Figure 3 presents the data. The N_k^*'s on the left panel tend to increase due to thinning caused by cross-sectioning. The right panel presents the age at sampling against the projected service time (under the assumption of service until the age of 65). There are no points below the diagonal dashed line due to truncation; all professors promoted during

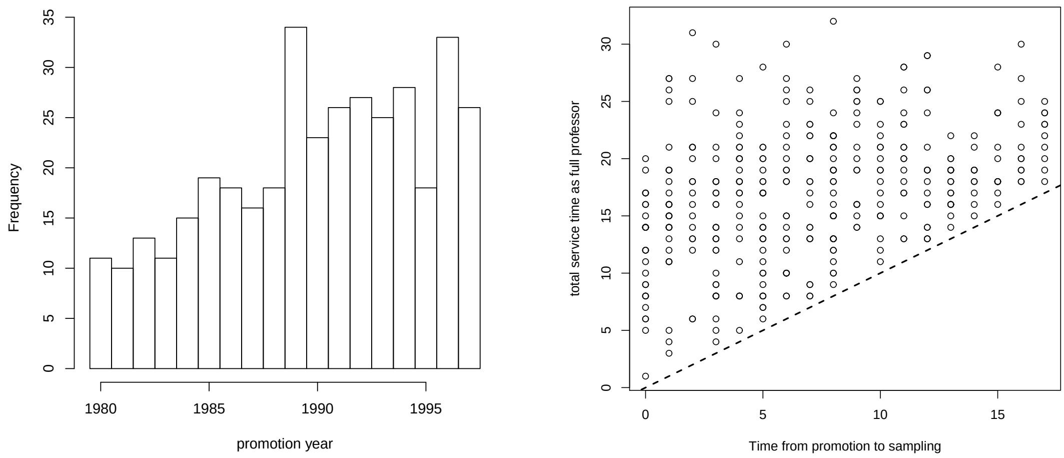

Figure 3: Left: entrance cohort sizes N_k^*. Right: time from promotion to sampling A_j^* and to 65 X_j^*.

1980-1997 who reached 65 before sampling (points below the dashed line) are not included in the sample. Obviously the service times of the selected professors are length-biased.

The left panel of Figure 4 presents estimates for the distribution of age at promotion to full professor obtained as the distribution of $65 - X$ with $X \sim G$. The solid step line is the nonparametric MCompLE of G and the circles present the nonparametric estimate using the conditional approach. The 95% pointwise confidence intervals (gray lines) were calculated assuming a Poisson entrance model and using the equation at the end of section 4.1. The smooth line is a parametric fit of a $Weibull(\alpha, \beta)$ model with survival function $\bar{G}(x) = e^{-(x/\beta)^\alpha}$. We obtained the estimates $\hat{\alpha} = 2.66$ and $\hat{\beta} = 16.91$, implying a mean close to 15 years. The sample standard errors of $\hat{\alpha}$ and $\hat{\beta}$ obtained assuming homogenous Poisson entrance are 0.09 and 0.28. The entrance process to HUJI is rather stable over time. If Constant N_k's are assumed, these standard deviations, obtained from (30), decrease only slightly. Typically to this quite large sample size, the conditional and composite nonparametric estimates are in a very good agreement. The Weibull estimate agrees quite well with the two nonparametric models, suggesting a reasonable fit to the data, with a moderate disagreement for ages larger than 55. All these estimates differ markedly from

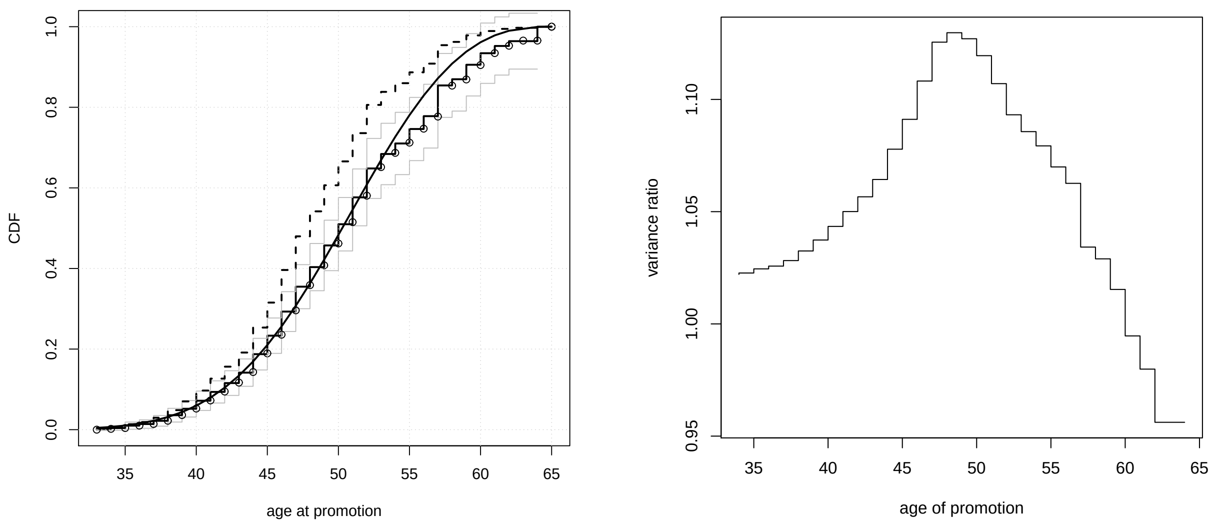

Figure 4: Left: estimates of the age at promotion distribution. Solid step lines - nonparametric unconditional estimate and the corresponding 95% pointwise confidence intervals (gray). Circles - nonparametric conditional estimate. Dashed - empirical distribution. Smooth line - parametric Weibull model. Right: ratio of variances of nonparametric estimators: conditional/unconditional.

the empirical distribution (dashed line) that fails to account for the sampling bias. The estimates indicate that about 60% of all promotions occur between the ages of 45 and 55, with about 90% occurring before the age of 60.

The right panel shows the ratio of variances of nonparametric estimators between the conditional and unconditional approaches. The latter is more efficient in most of the range by about 5%-10%.

A stratified analysis revealed that men are promoted at a younger age than women, with median ages of 46 and 50, respectively, and a similar comparison between experimental and non-experimental sciences shows medians of 45 versus 49. However, 53% of the 330 men are in the experimental sciences compared to only 33% of the 40 women; this may explain part of the differences between men and women or between faculties. Further study is required in order to test the significance of such differences.

5.2 Hospitalization Time Following Surgeries

As part of a monitoring program in Israel, four cross-sectional studies have been conducted in all hospitals in the country. On each survey day, data on all patients who had undergone surgery during the past 30 days were collected from the surgery day to 30 days afterward. The main goal was to monitor the incidence of complications after surgery, such as surgical site infection, repeated surgery, and death, see Fluss et al (2012) for more details. In this section, we apply the nonparametric composite and conditional estimation approaches to the distribution of length of hospitalization, trimmed at 30 days.

The number of urgent surgeries is reasonably constant over different days of the week, suggesting that the exchangeability assumption may hold. However, since most, but not all, elective surgeries are conducted on specific days of the week, the assumption that the N_k's are identically distributed is somewhat questionable, see Fluss et al (2012). It is therefore expected that the composite likelihood estimator will work well for urgent surgeries, but may fail for elective ones. We concentrate on data on bowel and hernia surgeries, the largest groups in the survey (587 bowel and 232 hernia surgeries of which 57% and 81% are elective, respectively).

Figure 5 presents the nonparametric estimators stratified by urgency status. Urgent surgeries typically require longer hospitalization compared to elective operations, and the same holds for bowel versus hernia surgeries (median=11 versus 4 days for urgent surgeries, median=8 versus 2 days for elective surgeries). The conditional and unconditional composite likelihood estimators are quite similar except for the group of elective hernia surgeries. In the latter case, the composite likelihood estimate for $G(2)$ is 0.79 (se=0.026) while the conditional likelihood estimate is 0.69 (se=0.039). It is easy to see that length-bias affects distributions supported by short durations more than by longer ones. The combination of short hospitalization times and inhomogeneous entrance process may explain the difference

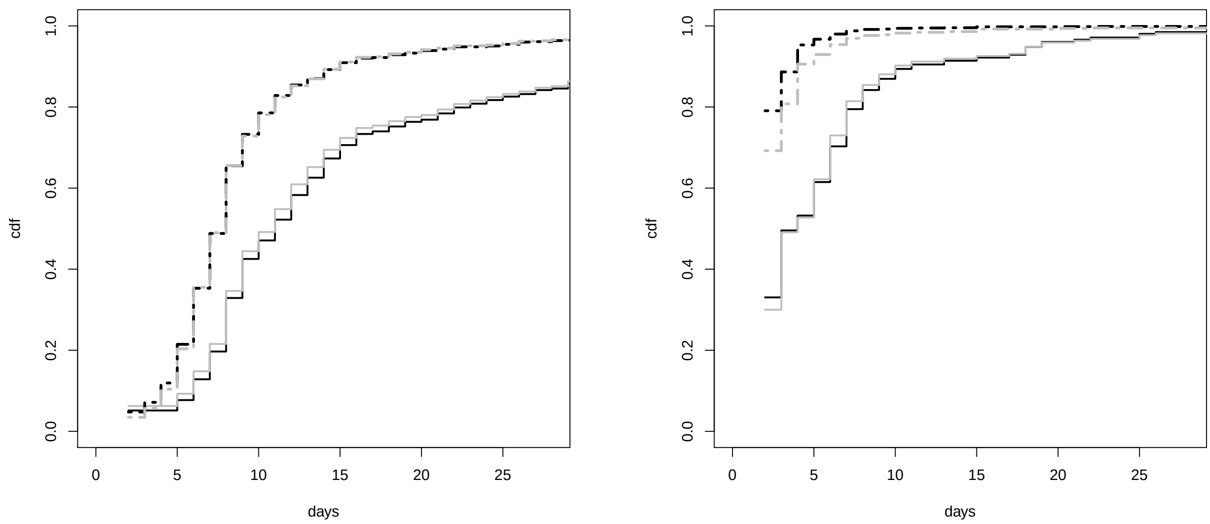

Figure 5: Hospitalization time after bowel (left) and hernia (right) surgery. Solid lines - urgent surgery, broken line - elective surgery. Black - nonparametric maximum composite likelihood, gray - nonparametric maximum conditional likelihood.

between the estimates, and therefore the use of the conditional approach is recommended in such cases. Composite likelihood methods involving auxiliary information that are appropriate for such inhomogeneous models are a topic of current research.

We remark that the data we analyzed were aggregated from several hospitals. This may moderate the problem of inhomogeneous entrance process in each hospital, if different hospitals have different schedules.

6 Simulation

To compare small sample properties of the conditional and composite likelihood approaches, we conducted a simulation study with $K = 20$ entrance points ($a_k = k - 1$) and a Gamma lifetime distribution with mean 12 and variance 48 ($\mu = E\lfloor X \rceil = 11.75$). We considered small/moderate sample sizes with $EN_k = 20$ (correspondingly, $EM^* = 235$) and larger sample sizes with $EN_k = 50$ ($EM^* = 587$). The following models for the distribution of N_k

were tested: Pois - independent Poisson entrance numbers; Mix1 - $\frac{1}{2}$Pois(15) + $\frac{1}{2}$Pois(25) in the small sample size scenario, and $\frac{1}{2}$Pois(43) + $\frac{1}{2}$Pois(57) in the large sample size scenario, which reflect moderate deviation from the Poisson model; Mix2 - $\frac{1}{2}$Pois(10) + $\frac{1}{2}$Pois(30) and $\frac{1}{2}$Pois(35) + $\frac{1}{2}$Pois(65), which reflect large deviation from the Poisson model; Geo - independent Geometric entrance numbers; Const - A constant number of entrances at each point; Mult - symmetric multinomial models with $M \equiv 400$ and $M \equiv 1000$ for the small and large sample size scenarios, respectively; Inhomo - entrances are independent following inhomogeneous Poisson variables with $N_k \sim \text{Pois}(\exp(3.27 - 0.027k))$ in the small sample scenario, and $N_k \sim \text{Pois}(\exp(4.12 - 0.021k))$ in the large sample scenario. These number were chosen so that the means of the N_k's are around 20 and 50 respectively.

For each model, we simulated 1000 samples and estimated G nonparametrically and parametrically in the Gamma(α, β) family. In each framework and for each simulated sample, we calculated the conditional and MCompLE estimates of G, and averaged over the 1000 replications to obtain estimates for the bias and MSE at the 10, 25, 50, 75, and 90 percentiles of G. Table 1 lists the ratio MSE(conditional)/MSE(MCompLE) at the median of G; complete results are provided in the Web-Appendix, along with simulations on further models with various entrance processes. As expected, the results show a clear advantage for the composite approach when N_k's are Poisson, or when they are relatively stable, such as Constant or Mix1, while for more variable N_k's, the conditional approach is preferable.

7 Discussion

The fact that entrance processes other than Poisson may cause dependence among observed sojourn times has been overlooked before. In part of the literature independence was assumed without clearly modeling the entrance process. Here, we clarify this issue and

	$E(N_k = 20)$		$E(N_k = 50)$	
Model	parm	nonparm	parm	nonparm
Pois	1.21	1.17	1.24	1.19
Mix1	1.13	1.10	1.13	1.08
Mix2	0.81	0.86	0.85	0.85
Geo	0.39	0.48	0.21	0.24
Const	1.42	1.33	1.43	1.36
Multinom	1.22	1.19	1.22	1.18
Inhomo	0.68	0.76	0.60	0.66

Table 1: Ratio of MSE of conditional and composite estimators for G at the median. parm - parametric estimation, nonparm - nonparametric estimation.

discuss conditions on the entrance process under which composite likelihood inference yields consistent and asymptotically normal estimators. These conditions allow a non-Poisson entrances (with certain restrictions on their variances), and hence dependent observations. We demonstrate that composite likelihood estimators can be more efficient than standard product-limit estimators for well-behaved entrance processes, but may fail for others.

A relevant and important question is therefore testing goodness-of-fit of entrance models in general, and the homogeneous Poisson in particular. This seems a difficult task that requires further study because of the thinning of the sampling process. However, following initial estimation of G, the thinning process can be estimated, and can be used to test models for the entrance process.

Goodness-of-fit of parametric models for the sojourn time distribution G may be based on a comparison to nonparametric estimates using generalized Chi-Square tests (Li and Doss 1993). This requires the study of the limiting behavior of the continuous process

$\sqrt{M^*}\{\hat{G}(t) - G(t)\}$ of the nonparametric MCompLE, which is beyond the scope of the present paper.

The independence result (Corollary 2.1) and the composite likelihood approach may be extended to continuous entrance processes. We conjecture that when the entrance process is a renewal process, sojourn times in a cross-sectional sample are independent if and only if entrances are governed by a Poisson process. The proof may use a characterization of the Poisson process given by Liberman (1985) and by Gan and Yang (1989).

References

[1] Asgharian, M., M'Lan, C. E., and Wolfson, D. B. (2002), "Length-Biased Sampling with Right Censoring: An Unconditional Approach," *Journal of the American Statistical Association*, 97, 201-209.

[2] Chatterji, S. D. (1963), "Some Elementary Characterizations of the Poisson Distribution," *The American Mathematical Monthly*, 70, 958-964.

[3] Cox, D.R. (1969), "Some Sampling Problems in Technology," in New Developments in Survey Sampling, Eds. N. L. Johnson and H. Smith, 506-527, Wiley (New York).

[4] Feller, W. (1971), *An Introduction to Probability Theory and its Applications*, Vol 2, Wiley.

[5] Gan, G., and Yang, S.-S. (1989), "An Elementary Proof of the Order-Statistics Characterization of the Poisson Process," *The American Statistician*, 43, 45-46.

[6] Fluss, R., Mandel, M., Freedman, L. S., Salz-Weiss, I., Ekka-Zohar, A., Haklai, Z., Gordon, E.-S., and Simchen, E. (2012), "Correction of Sampling Bias in a Cross Sectional Study of Post Surgical Complications," submitted.

[7] Kaminski, D., and geisler, C. (2012), "Survival Analysis of Faculty Retention in Science and Engineering by Gender," *Science*, 335, 864-866.

[8] Kaplan, E. L., and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 53, 457481.

[9] Laslett, G. M. (1982), "The Survival Curve Under Monotone Density Constraints with Applications to Two-Dimensional Line Segment Processes," *Biometrika*, 69, 153-160.

[10] Lehmann, E. L., and Casella, G. (1998), *Theory of Point Estimation*, 2nd edition, Springer.

[11] Li, G., and Doss, H. (1993), "Generalized Pearson-Fisher Chi-Square Goodness-of-Fit Tests, with Applications to Models with Life History Data," *The Annals of Statistics* 21, 772-797.

[12] Liberman, U. (1985), "An Order Statistic Characterization of the Poisson Renewal Process," *Journal of Applied Probability*, 22, 717-722.

[13] Lynden-Bell, D. (1971), "A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars," *Monograph National Royal Astronomical Society*, 155, 95-118.

[14] Mandel, M., and Fluss, R. (2009), "Nonparametric Estimation of the Probability of Illness in the Illness-Death Model Under Cross-Sectional Sampling," *Biometrika*, 96, 861-872.

[15] Rényi, A. (1957), "On the Asymptotic Distribution of the Sum of a Random Number of Independent Random Variables," *Acta Mathematica Hungarica*, 8, 193-199.

[16] Vardi, Y. (1985), "Empirical Distributions in Selection Bias Models," *The Annals of Statistics*, 13, 178-203.

[17] Vardi, Y. (1989), "Multiplicative Censoring, Renewal Processes, Deconvolution and Decreasing Density: Nonparametric Estimation," *Biometrika*, 76, 751-761.

[18] Varin, C., Reid, N., and Firth, D. (2011), "An Overview of Composite Likelihood Methods," *Statistica Sinica*, 21, 542.

[19] Wang M.-C. (1991), "Nonparametric Estimation from Cross-Sectional Survival Data," *Journal of the American Statistical Association*, 86, 130-143.

[20] Wang, M.-C., Jewell, N. P., and Tsai, W.-Y. (1986), "Asymptotic Properties of the Product Limit Estimate Under Random Truncation," *The Annals of Statistics*, 14, 1597-1605.

[21] Woodroofe, M. (1985), "Estimating a Distribution Function with Truncated Data," *The Annals of Statistics*, 13, 163-177 (Corr: V15 p883).

A Appendix: Proofs of Theorems

For the proof of Theorem 2.1 we need the following:

Lemma A.1. *1. Let $V_1, \dots, V_n$ be independent random variables with support $\{0, 1, 2, \dots\}$. If $(V_1, \dots, V_n | \sum_{i=1}^n V_i = m) \sim \text{Mult}(m, \pi_1, \dots, \pi_n)$ for all $m \geq 1$, then $V_i \sim \text{Poisson}(c\pi_i)$ for some positive c.*

2. Let $\{Z_i\}_{i\geq 1}$ be independent Bernoulli random variables with parameter $0 < \pi \leq 1$, and let W be an integer valued variable independent of the Z_is. Let $V = \sum_{i=1}^W Z_i$, then $V \sim \text{Poisson}(\pi\mu)$ if and only if $W \sim \text{Poisson}(\mu)$.

Part 1 of the lemma is proved by utilizing a special case of a characterization of the Poisson distribution given by Chatterji (1963). Part 2 says that a thinned variable is Poisson if and only if so is the original variable. For completeness, we provide the proofs.

Proof of lemma A.1. Part 1. For $n = 2$, $1 \leq a \leq m$, and $i = 1, 2$,

$$\frac{\pi_i}{1-\pi_i}\frac{m-a+1}{a} = \frac{P(V_i = a|V_1 + V_2 = m)}{P(V_i = a-1|V_1 + V_2 = m)} = \frac{P(V_i = a)P(V_{3-i} = m-a)}{P(V_i = a-1)P(V_{3-i} = m-a+1)}$$

which for $a = m$ gives for any m

$$\frac{P(V_i = m)}{P(V_i = m-1)} = \frac{\pi_i}{1-\pi_i}\frac{P(V_{3-i} = 1)}{P(V_{3-i} = 0)}\frac{1}{m} = \frac{c_i}{m},$$

a ratio which implies $V_i \sim \text{Poisson}(c_i)$. It is easy to see directly that $EV_i = \pi_i E(V_1 + V_2)$, hence $c = E(V_1 + V_2)$. For $n > 2$ one can prove $V_i \sim \text{Poisson}(c\pi_i)$ by writing $V_{-i} = \sum_{j\neq i} V_j$ and using the Multinomial property $(V_1, \dots, V_n)|\{\sum_{i=1}^n V_i = m\} \sim \text{Mult}(m, \pi_1, \dots, \pi_n)$ implies $(V_i, V_{-i})|\{\sum_{j=1}^n V_j = m\} \sim \text{Mult}(m, \pi_i, 1 - \pi_i)$.

Part 2. Let $\phi_Y(t) = Et^Y$ denote the probability generating function of a random variable Y, then $\phi_V(t) = E[t\pi + 1 - \pi]^W = \phi_W(t\pi + 1 - \pi)$ which gives at once $V \sim \text{Poisson}(\pi\mu) \Leftrightarrow \phi_V(t) = e^{\pi\mu(t-1)} \Leftrightarrow \phi_W(t) = e^{\mu(t-1)} \Leftrightarrow W \sim \text{Poisson}(\mu)$. □

Proof of theorem 2.1. Since any permutation of the ages is equally likely,

$$P(\{A_j = a_j\}_{j=1}^M | M = m) = \frac{\prod_{k=1}^K n_k!}{m!} P\big(N_1 = n_1, \ldots, N_K = n_K | M = m\big), \tag{19}$$

where $n_k = \sum_{j=1}^m I\{a_j = a_k\}$. Assumption 2.1 asserts that the ages at sampling and sojourn times are independent, and together with (19) the joint density of ages and sojourn times at $\{a_j, x_j\}$, conditionally on $M = m$, is

$$\frac{\prod_{k=1}^K n_k!}{m!} P\big(N_1 = n_1, \ldots, N_K = n_K | M = m\big) \prod_{j=1}^m dG(x_j). \tag{20}$$

If $N_k \sim \text{Poisson}(\lambda_k)$ then $(N_1, \ldots, N_K | M = m) \sim \text{Mult}\,(m, \lambda_1/\sum \lambda_k, \ldots, \lambda_K/\sum \lambda_k)$. Using this in (20), the joint density of $(A_1, \ldots, A_m, X_1, \ldots, X_m)|M = m$ reduces to

$$\prod_{j=1}^m \left\{ dG(x_j) \prod_{k=1}^K \left(\frac{\lambda_k}{\sum_\ell \lambda_\ell} \right)^{I\{a_j = a_k\}} \right\}, \tag{21}$$

which proves conditional (on m) independence of the pairs (A_j, X_j), $j = 1, \ldots, m$.

Now assume that given m, the above pairs are independent. Then the joint distribution of ages and sojourn times (20) must equal the product of its marginals:

$$\frac{\prod_{k=1}^K n_k!}{m!} P\big(N_1 = n_1, \ldots, N_K = n_K | M = m\big) \prod_{j=1}^m dG(x_j) = \prod_{k=1}^K \{P(A = a_k)\}^{n_k} \prod_{j=1}^m dG(x_j),$$

which implies that $(N_1, \ldots, N_K | M = m) \sim \text{Mult}\,(m, P(A = a_1), \ldots, P(A = a_K))$. Part 1 of Lemma A.1 establishes that N_k has a Poisson distribution. □

Proof of Corollary 2.1. If $N_k \sim \text{Poisson}(\lambda_k)$, $k = 1, \ldots, K$, then the claim follows from Theorem 2.1 and from the general fact that if any random variables $V_1, \ldots, V_n$ are independent and if $B_1, \ldots, B_n$ are sets, then $\{V_j\}_{j: V_j \in B_j}$ are independent conditionally on the set of indices $\{j : V_j \in B_j\}$. In our case, $V_j = (A_j, X_j)$ and $B_j = \{(a, x) \in \mathbb{R}^{+2} : a \le x\}$.

It remains to prove the other direction. Assumption 2.1 implies that conditionally on $\{A_j^*\}$, the sojourn times $\{X_j^*\}$ are independent having the distribution $P(X_j^* \le x | A_j^* = a) = G(x)/\bar{G}(a-)$ for $x \ge a$. The arguments leading to (20) show that the joint density of

$(\{A_j^*\},\{X_j^*\})$ conditionally on $\{M^*=m^*\}$ is

$$\frac{\prod_{k=1}^K n_k^*!}{m^*!}P\big(N_1^*=n_1^*,\ldots,N_K^*=n_K^*|M^*=m^*\big)\prod_{j=1}^{m^*}dG(x_j^*)\prod_{k=1}^K\{\bar{G}(a_k^*-)\}^{-n_k^*}. \tag{22}$$

Under independence, (22) is proportional to $\prod_{j=1}^{m^*}dG(x_j^*)\prod_{k=1}^K\{P(A_j=a_k)\}^{n_k^*}$, see (4), so that $(N_1^*,\ldots,N_K^*)|M^*=m^*\sim \text{Mult}(m^*,P(A_1=a_1)\bar{G}(a_1^*-),\ldots,P(A_k=a_k^*)\bar{G}(a_K^*-))$. Part 1 of Lemma A.1 guarantees that the N_k^*'s have a Poisson law, and are independent due to Assumption 2.1, and Part 2 of the lemma guarantees the same for N_k, $k=1,\ldots,K$. □

Proof of Theorem 4.1 First note that $\{X_{ki}\}$ are iid and are independent of $\{N_k\}$. Denote by θ_0 the true parameter value, then by (10)

$$\frac{1}{M^*}\ell(\theta)=\frac{\frac{1}{\nu}\sum_{k=1}^K\sum_{i=1}^{N_k}I_{\{X_{ki}\geq a_k\}}\log\frac{\lfloor X_{ki}\rceil dG(X_{ki};\theta)}{\mu_\theta}}{\frac{1}{\nu}\sum_{k=1}^K\sum_{j=1}^{N_k}I_{\{X_{ki}\geq a_k\}}}\xrightarrow{p}E_{\theta_0}\log\{dG^*(X^*;\theta)\}, \tag{23}$$

where the limit is obtained as follows. Starting with the denominator and recalling that $\nu/N_k\xrightarrow{p}1$, the law of large numbers implies

$$\sum_{k=1}^K\frac{1}{\nu}\sum_{i=1}^{N_k}I_{\{X_{ki}\geq a_k\}}\xrightarrow{p}\sum_{k=1}^K E_{\theta_0}I_{\{X\geq a_k\}}=\mu_{\theta_0}. \tag{24}$$

The same reasoning applied to the numerator of (23) yields

$$\begin{aligned}&\sum_{k=1}^K\frac{1}{\nu}\sum_{i=1}^{N_k}I_{\{X_{ki}\geq a_k\}}\log\frac{\lfloor X_{ki}\rceil dG(X_{ki};\theta)}{\mu_\theta}\xrightarrow{p}\sum_{k=1}^K\int_0^\infty I_{\{x\geq a_k\}}\log\{dG^*(x;\theta)\}dG(x;\theta_0)\\&=\int_0^\infty\sum_{k=1}^K I_{\{x\geq a_k\}}\log\{dG^*(x;\theta)\}dG(x;\theta_0)=\mu_{\theta_0}\int_0^\infty\log\{dG^*(x;\theta)\}\frac{\lfloor x\rceil dG(x;\theta_0)}{\mu_{\theta_0}}\\&=\mu_{\theta_0}\int_0^\infty\log\{dG^*(x;\theta)\}dG^*(x;\theta_0)=\mu_{\theta_0}E_{\theta_0}\log\{dG^*(X^*,\theta)\},\end{aligned} \tag{25}$$

which together with (24) implies (23). Identifiability and the information inequality assert that $E_{\theta_0}\log\{dG^*(X^*,\theta)\}$ obtains its maximum at θ_0; standard arguments guarantee the existence of a consistent sequence of roots (e.g., Lehmann and Casella 1998). □

Proof of Theorem 4.2. Interchanging the order of integration and differentiation we obtain as in (25)

$$\sum_{k=1}^K c_k=\frac{\partial}{\partial\theta}\sum_{k=1}^K\int_0^\infty I_{\{x\geq a_k\}}\log\{dG^*(x;\theta)\}dG(x;\theta_0)\Big|_{\theta=\theta_0}=\mu_{\theta_0}\frac{\partial}{\partial\theta}E_{\theta_0}\log\{dG^*(X^*,\theta)\}\Big|_{\theta=\theta_0},$$

which vanishes since the maximum of $E_{\theta_0} \log\{dG^*(X^*, \theta)\}$ is attained at θ_0.

We now prove independence of W and V. The assumptions on the entrance process imply $M/(KN_k) \xrightarrow{p} 1$, and by (15), it suffices to prove asymptotic independence of $U^{(\nu)}$ and $W^{(\nu)} := \sum_{k=1}^K \frac{1}{\sqrt{N_k}} \sum_{j=1}^{N_k} \left(\frac{\partial}{\partial\theta} h_k(X_{kj}; \theta_0) - c_k\right)$. Given $\epsilon > 0$ let n_0 be such that $n_k > n_0$ for all k implies $|P(W^{(\nu)}/\sigma \le t \mid \{N_k = n_k\}) - \Phi(t)| < \epsilon$, where $\sigma^2 = \sum_{k=1}^K \mathrm{Var}_{\theta_0}\left(\frac{\partial}{\partial\theta} h_k(X; \theta_0)\right)$, and let ν be such that $P(N_k^{(\nu)} > n_0 \text{ for all } k) > 1 - \epsilon$. For $n_k > n_0$ we have

$$P(W^{(\nu)}/\sigma \le t, U^{(\nu)} \le u \mid \{N_k = n_k\}) \le (\Phi(t) + \epsilon) I\{\sum_{k=1}^K c_k \frac{n_k - \nu}{\sqrt{\nu}} \le u\}.$$

Unconditioning by summing over all $\{n_k\}$ readily yields

$$P(W^{(\nu)}/\sigma \le t, U^{(\nu)} \le u) \le (\Phi(t) + \epsilon) P(U^{(\nu)} \le u) + \epsilon.$$

A similar lower bound completes the proof. □

Proof of Theorem 4.3 Since $N_k \xrightarrow{p} \infty$, the weak law of large numbers yields

$$\frac{1}{N_k} \sum_{i=1}^{N_k} \frac{1}{\lfloor X_{ki} \rceil} I_{\{a_k \le X_{ki} \le x\}} \xrightarrow{p} E\left(\frac{1}{\lfloor X \rceil} I_{\{a_k \le X \le x\}}\right),$$

which holds also for $x = \infty$. In addition, $N_k/M \xrightarrow{p} 1/K$ by assumption, and by (17)

$$\hat{G}(x) \xrightarrow{p} E \sum_{k=1}^K \frac{1}{\lfloor X \rceil} I_{\{a_k \le X \le x\}} \Big/ E \sum_{k=1}^K \frac{1}{\lfloor X \rceil} I_{\{a_k \le X\}}.$$

By definition, $\sum_{k=1}^K \frac{1}{\lfloor X \rceil} I_{\{a_k \le X \le x\}} = I_{\{X \le x\}}$ therefore its expectation is $G(x)$. □

Proof of Theorem 4.4 By (17)

$$\sqrt{M}(\hat{G}(x) - G(x)) = \frac{\frac{1}{\sqrt{M}} \sum_{k=1}^K \sum_{i=1}^{N_k} \frac{1}{\lfloor X_{ki} \rceil} I_{\{a_k \le X_{ki}\}} [I_{\{X_{ki} \le x\}} - G(x)]}{\frac{1}{M} \sum_{k=1}^K \sum_{i=1}^{N_k} \frac{1}{\lfloor X_{ki} \rceil} I_{\{a_k \le X_{ki}\}}}. \tag{26}$$

The denominator in (26) converges to $1/K$ since $\frac{1}{N_k} \sum_{i=1}^{N_k} \frac{1}{\lfloor X_{ki} \rceil} I_{\{a_k \le X_{ki}\}} \xrightarrow{p} \gamma_k(\infty)$ by the Law of Large Numbers, $\sum_{k=1}^K \gamma_k(\infty) = 1$, and $N_k/M \xrightarrow{p} 1/K$.

Setting $S_{ki}(x) = \frac{1}{\lfloor X_{ki} \rceil} I_{\{a_k \le X_{ki}\}} [I_{\{X_{ki} \le x\}} - G(x)]$ we have $ES_{ki}(x) = c_k(x)$, and $\sum_k c_k = 0$. Using (16) we obtain that $\sqrt{M^*}\{\hat{G}(x) - G(x)\}$ is asymptotically distributed as

$$(\mu K)^{1/2} \sum_{k=1}^K \frac{1}{\sqrt{M}} \sum_{i=1}^{N_k} \{S_{ki}(x) - c_k(x)\} + \mu^{1/2} \sum_{k=1}^K \frac{c_k(x)(N_k - \nu)}{\sqrt{\nu}}. \tag{27}$$

We have $\frac{1}{\sqrt{M}}\sum_{i=1}^{N_k}\{S_{ki}(x)-c_k(x)\}\xrightarrow{\mathcal{D}}N(0,\sigma_k^2(x)/K)$, and therefore,

$$(\mu K)^{1/2}\sum_{k=1}^{K}\frac{1}{\sqrt{M}}\sum_{i=1}^{N_k}\{S_{ki}(x)-c_k(x)\}\xrightarrow{\mathcal{D}}\mu^{1/2}N\left(0,\sum_{k=1}^{K}\sigma_k^2(x)\right). \qquad \square$$

Independence of $W(x)$ and $V(x)$ follows by reasons as in the proof of Theorem 4.2.

Asymptotic normality in the multi-parametric case

Suppose that $\boldsymbol{\theta}$ is p-dimensional and that the MCompLE is consistent. Under standard regularity conditions, Taylor approximation gives

$$\sqrt{M}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\approx MH^{-1}(\boldsymbol{\theta}_0)\frac{1}{\sqrt{M}}D_\ell(\boldsymbol{\theta}_0), \tag{28}$$

where $D_\ell(\boldsymbol{\theta}_0)=\left(\frac{\partial}{\partial\theta_1}\ell(\boldsymbol{\theta}_0),\ldots,\frac{\partial}{\partial\theta_p}\ell(\boldsymbol{\theta}_0)\right)^t$ and $H(\boldsymbol{\theta}_0)=\left(\frac{\partial^2}{\partial\theta_s\partial\theta_t}\ell(\boldsymbol{\theta}_0)\right)$ is the $p\times p$ matrix of second derivatives. Under conditions as in Theorem 4.2

$$MH^{-1}(\boldsymbol{\theta}_0)=\left(\sum_{k=1}^{K}\frac{N_k}{M}\frac{1}{N_k}\sum_{i=1}^{N_k}\frac{\partial^2}{\partial\theta_s\partial\theta_t}h_k(X_{ki};\boldsymbol{\theta}_0)\right)^{-1}\xrightarrow{p}\left(\frac{1}{K}\sum_{k=1}^{K}E_{\boldsymbol{\theta}_0}\frac{\partial^2}{\partial\theta_s\partial\theta_t}h_k(X,\boldsymbol{\theta}_0)\right)^{-1}.$$

Next, write $\frac{1}{\sqrt{M}}D_\ell(\boldsymbol{\theta}_0)$ as a sum of two vectors:

$$\left(\frac{1}{\sqrt{M}}\sum_{k=1}^{K}\sum_{i=1}^{N_k}\frac{\partial}{\partial\theta_j}h_k(X_{ki};\boldsymbol{\theta}_0)\right)=\left(\sum_{k=1}^{K}\frac{1}{\sqrt{M}}\sum_{i=1}^{N_k}\left(\frac{\partial}{\partial\theta_j}h_k(X_{ki};\boldsymbol{\theta}_0)-c_{kj}\right)\right)+\left(\sum_{k=1}^{K}\frac{N_kc_{kj}}{\sqrt{M}}\right), \tag{29}$$

where $c_{kj}=E(\frac{\partial}{\partial\theta_j}h_k(X;\boldsymbol{\theta}_0))$, and note that

$$\left(\frac{1}{\sqrt{M}}\sum_{i=1}^{N_k}\left(\frac{\partial}{\partial\theta_1}h_k(X_{ki};\boldsymbol{\theta}_0)-c_{k1}\right),\ldots,\frac{1}{\sqrt{M}}\sum_{i=1}^{N_k}\left(\frac{\partial}{\partial\theta_p}h_k(X_{ki};\boldsymbol{\theta}_0)-c_{kp}\right)\right),\quad k=1,\ldots,K$$

converge jointly to independent zero mean normal vectors with corresponding covariances $\mathrm{Cov}(\frac{\partial}{\partial\theta_s}h_k(X;\boldsymbol{\theta}_0),\frac{\partial}{\partial\theta_t}h_k(X;\boldsymbol{\theta}_0))/K$. Therefore, the first term on the right hand side of (29) satisfies $\left(\sum_{k=1}^{K}\frac{1}{\sqrt{M}}\sum_{i=1}^{N_k}\left(\frac{\partial}{\partial\theta_j}h_k(X_{ki};\boldsymbol{\theta}_0)-c_{kj}\right)\right)\xrightarrow{\mathcal{D}}W$, where

$$W\sim N_p\Big(0,\frac{1}{K}\sum_{k=1}^{K}\mathrm{Cov}\left\{\frac{\partial}{\partial\theta_s}h_k(X;\boldsymbol{\theta}_0),\frac{\partial}{\partial\theta_t}h_k(X;\boldsymbol{\theta}_0)\right\}\Big). \tag{30}$$

The second term in (29) vanishes if the N_k's are constant, and otherwise can be treated as in single parameter case.