

Estimation from Cross-Sectional Samples under Bias and Dependence

BY MICHA MANDEL AND YOSEF RINOTT

Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel

msmic@huji.ac.il yosef.rinott@mail.huji.ac.il

5

SUMMARY

A population can be entered at a known sequence of discrete times; it 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 and often ignored that it may result also in dependence among the observations. We show that observed sojourn times are independent only under a multinomial entrance process. We study asymptotic properties of parametric and nonparametric estimators of the sojourn time distribution using the product of marginals in spite of dependence, and provide conditions under which this approach results in proper or improper and wrong inference. We apply the proposed methods to data on hospitalization time after bowel and hernia surgeries collected by a cross-sectional design.

10

15

Some key words: Discrete entrance process; Length bias; Poisson cohort distribution; Survival analysis; Truncation.

1. INTRODUCTION

Consider a population \mathcal{S} that can be entered at a fixed and known sequence of time-points. In our motivating example \mathcal{S} consists of patients in a hospital and the entrance times are the days on which the relevant treatment is available. 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 function, G , of individuals in \mathcal{S} . Cross-sectioning biases G , and also results in a thinned entrance process. Moreover, 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.

20

25

Let A_j denote the time from entering \mathcal{S} to sampling and let X_j denote the total sojourn time in \mathcal{S} of subject j . We observe (A_j, X_j) for subjects present in \mathcal{S} at the sampling time, that is, those satisfying $A_j \leq X_j$. This is a truncation model where observed pairs have the law of $(A, X) \mid A \leq X$. The standard approach estimates G conditionally on the times from entering \mathcal{S} to sampling, see Woodroffe (1985) and Wang et al. (1986). In general, these times are not ancillary, that is, their distribution depends on G , and therefore, conditioning on them may lead to loss of information.

30

The most common unconditional approach maximizes the likelihood of the observed lifetimes assuming they are independent having a length biased distribution, (Cox, 1969; Vardi, 1985). This assumption is 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 et al. (2002).

35

We show that in our discrete entrance model, observed sojourn times are generally not independent. Moreover, the dependence structure is rather complex, and a full likelihood approach is not feasible. Instead, we study parametric and nonparametric estimators based on the inde-

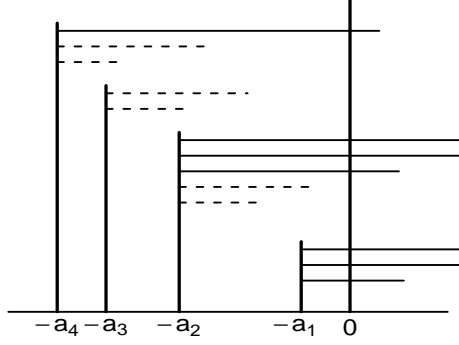


Fig. 1. An example with $K = 4$, and $(N_1, N_2, N_3, N_4) = (3, 5, 2, 3)$. Solid lines - observed sojourn times, dashed lines - unobserved. $(N_1^*, N_2^*, N_3^*, N_4^*) = (3, 3, 0, 1)$.

pendence likelihood, which consists of the product of univariate marginals (Chandler & Bate, 2007). This approach is within the framework of composite likelihood inference (Varin et al., 2011). We provide conditions for consistency and for asymptotic normality and non-normality, and indicate situations where independence likelihood methods are more efficient than the standard conditional approach. We also indicate situations where the independence likelihood fails, showing that the dependence in truncated cross-sectional data cannot be ignored unless certain conditions hold.

2. DISCRETE ENTRANCE PROCESS

Let N_1, \dots, N_K denote the random cohort sizes, that is, the numbers of individuals who enter the population \mathcal{S} at known and fixed time points $-a_K < \dots < -a_1 \leq 0$, where 0 is set to be the cross-sectioning time. The sojourn times of individuals in \mathcal{S} are denoted by X_j , and the times from entrance to \mathcal{S} to 0 are denoted by $A_j \in \{a_1, \dots, a_K\}$, where A_j may be smaller or larger than X_j . We refer to A_j as the truncation time or age of individual j . A standard assumption made throughout is that sojourn times $\{X_j\}$ are independent, with $X_j \sim G$ independently of the ages $\{A_j\}$.

The cross-sectional sample consists of those individuals for whom $A_j \leq X_j$. 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 $X^* \sim G^*$, and denote the corresponding densities or probability functions by g and g^* . The observed number N_k^* 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 , thus $N_k^* | N_k \sim \text{Bin}\{N_k, \bar{G}(a_k-)\}$, where $\bar{G} = 1 - G$. The total population and sample sizes are $M = \sum_k N_k$ and $M^* = \sum_k N_k^*$. See Figure 1.

The cross-sectional sample does not include sojourn times smaller than a_1 and therefore $G(x)$ is not estimable for $x \leq a_1$; and the estimable function is $\text{pr}(X \leq x | X \geq a_1)$. For notational convenience, we henceforth assume that $a_1 = 0$. The sojourn time X_j of individual j is denoted by X_{ki} when we want to emphasize that it is the i th individual in cohort k . The data comprise $\{(a_j^*, x_j^*) : j = 1, \dots, m^*\}$, the ages and sojourn times of the m^* observations; the number of individuals of age a_k in the cross-sectional sample is $n_k^* = \sum_j I(a_j^* = a_k)$, and $m^* = \sum_{k=1}^K n_k^*$.

Under this discrete truncation model, individual sojourn times in the sample are dependent, even asymptotically. For a simple example, let $K = 2$, $a_1 = 1$, $a_2 = 5$, and $(N_1, N_2) =$

(10, 100) or (100, 10) each with probability 0.5. Let the sojourn times be $\text{Unif}(0, 100)$, and suppose $m^* = 100$. Observing $x_1^* = 3$ implies $a_1^* = 1$ and makes the event $\{(N_1, N_2) = (100, 10)\}$ more likely. This suggests that other x_j^* 's will tend to be small. 70

Since conditions for independence of the pairs (A_j^*, X_j^*) are often not stated clearly in the survival literature, it is important to point out exactly when independence holds.

THEOREM 1. *The pairs $(A_1^*, X_1^*), \dots, (A_{M^*}^*, X_{M^*}^*)$ are independent conditionally on M^* if and only if $(N_1, \dots, N_K) \mid M = m$ has a multinomial distribution for all m , which for independent N_1, \dots, N_K is equivalent to $N_k \sim \text{Poisson}(\lambda_k)$, $k = 1, \dots, K$.* 75

All proofs are given in the Supplementary Material. By Theorem 1, strong assumptions are needed for the sojourn times in the cross-sectional sample, $\{X_j^*\}$, to be independent. On the other hand, conditionally on $\{A_j^*\}$, the sojourn times $\{X_j^*\}$ are independent under any distribution of $\{N_k\}$. Hence, the conditional approach, based on the conditional likelihood $\prod_k \prod_i dG(x_{ki}^*)/\bar{G}(a_k -)$, is robust with respect to the entrance process model. 80

Although the multinomial model is a reasonable approximation of many entrance processes, it does not always hold. A natural example is of an infectious disease, where N_1, \dots, N_K denote the number of infected individuals in different months, and X denotes the infection period. As entrance times are dependent, the multinomial model is violated and sojourn times in the cross-sectional sample are dependent. 85

3. INDEPENDENCE LIKELIHOOD

3.1. Likelihood construction

In this section, we study the consequences of using the independence likelihood approach which bases inference on the product of marginals of the (A_j^*, X_j^*) 's. Our goal is to find conditions under which maximizing the independence likelihood provides consistent estimators, and to compute their asymptotic distribution and variance, taking the dependence into account. 90

We start with A_j , the age of subject j , chosen at random among those who entered \mathcal{S} at one of the points $-a_K, \dots, -a_1$. With $0/0 = 0$ by convention, we have

$$\text{pr}(A_j = a_k) = \sum_{\mathcal{N}} \frac{n_k}{\sum_{s=1}^K n_s} \text{pr}(N_1 = n_1, \dots, N_K = n_K) = E(N_k/M), \quad (1)$$

where the sum is over $\mathcal{N} = \{(n_1, \dots, n_K) : \sum_{k=1}^K n_k \geq 1, n_k \in \{0, 1, 2, \dots\} \quad k = 1, \dots, K\}$. 95

In expressions derived from (1), we shall replace $\text{pr}(A_j = a_k) = E(N_k/M)$ by $\eta_k = E(N_k)/E(M)$; they are equal when $(N_1, \dots, N_K) \mid M$ has a multinomial distribution, holding, for example, when N_k 's are independent Poisson, and when N_k 's are exchangeable, in which case $\eta_k = E(N_k/M) = 1/K$. Otherwise, it is an approximation justified by Lemma 1 below.

Recalling $X_j \sim G$, the assumption of independence of A_j and X_j , and that (A_j^*, X_j^*) are distributed as $(A_j, X_j) \mid \{A_j \leq X_j\}$, the joint density of (A_j^*, X_j^*) is 100

$$f_{A_j^*, X_j^*}(a_k, x) = \frac{\text{pr}(A_j = a_k) dG(x)}{\beta} \mathbf{I}(a_k \leq x) = \frac{\eta_k dG(x)}{\beta} \mathbf{I}(a_k \leq x), \quad (2)$$

where $\beta = \text{pr}(A_j \leq X_j) = \sum_{k=1}^K \eta_k \bar{G}(a_k -)$. The marginal density of an observed sojourn time is obtained by summing over a_k :

$$dG^*(x) = \sum_{k: a_k \leq x} \frac{\eta_k dG(x)}{\beta} = \frac{w(x) dG(x)}{\beta}, \quad (3)$$

a weighted version of G , with weights given by the step-function $w(x) = \sum_{k: a_k \leq x} \eta_k$. The distribution obtained in (3) depends on the joint distribution of N_1, \dots, N_K only through the η_k 's. Under exchangeability, the weight function at x reduces to $w(x) = K^{-1} \sum_{k=1}^K \mathbf{I}(a_k \leq x) = K^{-1} \max\{k : a_k \leq x\}$, which is proportional to the potential number of time-points at which an individual having sojourn time x could enter the population and still be included in the sample.

For the exchangeable case and more generally for known η_k 's. e.g., Fluss et al. (2012), the independence likelihood is a product of the marginals of the observed sojourn times as in (3),

$$L(G) = \prod_{j=1}^{m^*} g^*(x_j^*) = \prod_{j=1}^{m^*} \frac{w(x_j^*)g(x_j^*)}{\beta}. \quad (4)$$

Clearly, $L(G)$ is proportional to the product of the joint densities of (A_j^*, X_j^*) , see (2).

3.2. Asymptotic results

Our goal is to study the independence likelihood estimator $\hat{G} = \arg \max L(G)$ in the presence of dependence among $X_1^*, \dots, X_{M^*}^*$. An important device to be used below is a representation of $\ell(G) = \log L(G) = \sum_{j=1}^{M^*} \log g^*(X_j^*)$ in terms of the independent variables $X_{ki} \sim G$:

$$\ell(G) = \sum_{k=1}^K \sum_{i=1}^{N_k} \mathbf{I}(a_k \leq X_{ki}) \log \frac{w(X_{ki})g(X_{ki})}{\beta} = \sum_{k=1}^K \sum_{i=1}^{N_k} h_k(X_{ki}). \quad (5)$$

The $h_k(X_{ki})$ defined in the sum above are independent but not identically distributed.

In our data, and quite typically for truncation models, the relevant asymptotics appear to be associated with increasing the sample size while keeping the marginals of (A_j^*, X_j^*) fixed; see, Woodroffe (1985) and Wang et al. (1986). We consider large entrance numbers N_k , with a fixed number of entrance points K . Asymptotics obtained by considering large K are technically easy, leading to standard consistency and normality results; see Supplementary Material. Set $N_k = N_k^{(\nu)}$ satisfying $\lim_{\nu \rightarrow \infty} \Pr(N_k^{(\nu)} \geq n) = 1$ for all n . For simplicity, we henceforth assume that the sequence is parameterized so that $\nu = E(M^{(\nu)})$ and $E(N_k^{(\nu)}) = \eta_k \nu$. We often omit the superscript ν and expressions like $N_k/E(N_k) \rightarrow 1$ appearing in Lemma 1 or in Theorem 2 below are taken with respect to $\nu \rightarrow \infty$. The following lemma is straightforward; it justifies the approximation described in Section 3.1 of replacing $E(N_k/M)$ by $\eta_k = E(N_k)/E(M)$.

LEMMA 1. *Let $E(N_k) = \eta_k \nu$ with $\sum_k \eta_k = 1$, and assume $N_k/E(N_k) \rightarrow 1$ in probability. Then for $k = 1, \dots, K$, $E(N_k/M) \rightarrow \eta_k$ as $\nu \rightarrow \infty$.*

3.3. Parametric models

Suppose $G = G(\cdot; \theta)$ is indexed by a parameter $\theta \in \Theta \subseteq \mathbb{R}$, taken to be univariate for simplicity; extensions to the multi-parameter case and to models with covariates are discussed in the Supplementary Material. We write $\ell(\theta)$, β_θ , and $h_k(\cdot; \theta)$ instead of $\ell(G)$, β , and $h_k(\cdot)$ of (5). The next theorem deals with consistency.

THEOREM 2. *Let $N_k/E(N_k) \rightarrow 1$ in probability and assume that $g(x; \theta)$ is differentiable with respect to θ for all x , and the standard regularity conditions of identifiability, common support 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 independence likelihood score equation $\partial \ell(\theta)/\partial \theta = 0$.*

The regularity conditions above appear as Conditions (A0), (A1), and (A3) in Lehmann and Casella (1998, p. 444–5); their Condition (A2) is that $\ell(G)$ is a sum of independent and identically distributed random variables. We extend the proof to our case, where $h_k(X_{ki})$ are independent but not identically distributed. 140

When the root is unique, it is the independence likelihood estimator, and the resulting sequence is consistent. In particular, if G belongs to a canonical exponential family, then so does $G^*(\cdot; \theta)$ 145 and if the independence likelihood estimator exists, it is unique. Also, if the parameter space is compact, then any sequence of maximum independent likelihood estimators is consistent and differentiability of $g(x; \theta)$ is not needed (Ferguson, 1996).

By Chebyshev's inequality, $\text{pr}\{|N_k/E(N_k) - 1| > \varepsilon\} \leq \text{var}(N_k)/\{\varepsilon E(N_k)\}^2$ so that for models where the coefficient of variation vanishes, $N_k/E(N_k) \rightarrow 1$ in probability. This holds 150 for Poisson or binomial, but not geometric or uniform variables.

The regularity conditions and parts of the analysis of the asymptotic distribution of the independence likelihood estimator appear similar to those of Theorem 3.10 in Lehmann and Casella (1998, page 449). However, the results are not the same; interestingly, the limiting distribution of the independence likelihood estimator is not necessarily normal. 155

THEOREM 3. *Assume all the conditions of Theorem 2, and additionally that in some neighborhood of θ_0 the following standard regularity conditions hold. (i) $g(x; \theta)$ 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 $-E_\theta\{\partial^2 \log g(X; \theta)/\partial\theta^2\} \in (0, \infty)$; (iv) $|\partial^3 \log g(x; \theta)/\partial\theta^3| < \psi(x)$ for all x , where $E_{\theta_0}\{\psi(X)\} < \infty$. Let* 160

$$U = U^{(\nu)} = \sum_{k=1}^K c_k \frac{N_k - \eta_k \nu}{\nu^{1/2}}, \quad c_k = E_{\theta_0} \left\{ \frac{\partial}{\partial \theta} h_k(X; \theta_0) \right\}. \quad (6)$$

If $U^{(\nu)} \rightarrow V$ in distribution for some random variable V , then for any consistent sequence $\hat{\theta}_\nu$ of roots of the independence likelihood score equation $\partial \ell(\theta)/\partial \theta = 0$,

$$M^{*1/2}(\hat{\theta}_\nu - \theta_0) \rightarrow \frac{\beta_{\theta_0}^{1/2}}{\sum_{k=1}^K \eta_k E_{\theta_0} \left\{ \frac{\partial^2}{\partial \theta^2} h_k(X, \theta_0) \right\}} (W + V) \quad (7)$$

in distribution, where $W \sim N\left[0, \sum_{k=1}^K \eta_k \text{var}_{\theta_0} \left\{ \frac{\partial}{\partial \theta} h_k(X; \theta_0) \right\}\right]$ is independent of V , and the resulting asymptotic variance of $M^{*1/2}(\hat{\theta}_\nu - \theta_0)$ is

$$\beta_{\theta_0} \frac{\sum_{k=1}^K \eta_k \text{var}_{\theta_0} \left\{ \frac{\partial}{\partial \theta} h_k(X; \theta_0) \right\} + \text{var}(V)}{\left[\sum_{k=1}^K \eta_k E_{\theta_0} \left\{ \frac{\partial^2}{\partial \theta^2} h_k(X, \theta_0) \right\} \right]^2}. \quad (8)$$

Remark 1. When $(N_1, \dots, N_K) \mid M$ has a multinomial distribution, holding, for example, 165 when $N_k \sim \text{Poisson}(\eta_k \nu)$ independent, then by (6) $\text{var}(U^{(\nu)}) = \sum \eta_k c_k^2$. In this case, standard calculations show that (8) reduces to $1/E_{\theta_0} \left\{ \partial g^*(X^*, \theta_0)/\partial \theta \right\}^2$, which has the usual interpretation of Fisher's information for independent X_j^* 's. This is expected, since by Theorem 1 the observed sojourn times are indeed independent. It follows that ignoring the dependence in variance calculations leads to anti-conservative inference if $\text{var}(V) > \sum \eta_k c_k^2$, and conversely. Thus, 170 when the Poisson model is reasonable or when the variance of cohort sizes is smaller than their expectations, the observed Fisher information of the independence likelihood can be used as a basis for a conservative variance estimator. However, if the variance of the cohort sizes is much

larger than their expectation, this variance estimator is anti-conservative, and moreover, the estimator itself may be inconsistent. In such cases, one may resort to the conditional approach.

In general, $\text{var}(W)$ can be estimated by plugging $\hat{\theta}$ in (8), but $\text{var}(V)$ depends on the distribution of the N_k 's. If the N_k 's are constant then $\text{var}(V) = 0$; certain pre-scheduled allocations are indeed close to being constant. If the distribution of the N_k 's is known from past experience, then $\text{var}(V)$ can be evaluated.

In the Supplementary Material, we discuss various dependence structures of cohort sizes, leading to different asymptotic distributions of V and hence of $M^{*1/2}(\hat{\theta}_\nu - \theta_0)$, and study the effect of correlations among the cohort sizes on the asymptotic variance of V .

3.4. Nonparametric models

In the nonparametric case, (4) has the form of a likelihood under biased sampling and is maximized by the inverse weighting estimator, e.g., Vardi (1985):

$$\hat{G}(x) = \frac{\sum_{j=1}^{M^*} w(X_j^*)^{-1} I(X_j^* \leq x)}{\sum_{j=1}^{M^*} w(X_j^*)^{-1}}.$$

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

$$\hat{G}(x) = \frac{M^{-1} \sum_{k=1}^K \sum_{i=1}^{N_k} w(X_{ki})^{-1} I(a_k \leq X_{ki} \leq x)}{M^{-1} \sum_{k=1}^K \sum_{i=1}^{N_k} w(X_{ki})^{-1} I(a_k \leq X_{ki})}. \quad (9)$$

We next discuss consistency and the asymptotic distribution of the nonparametric independence likelihood estimator. For this purpose, set $\gamma_k(x) = E\{I(a_k \leq X \leq x)/w(X)\}$, then

$$\sum_{k=1}^K \eta_k \gamma_k(x) = E \left\{ \frac{\sum_{k=1}^K \eta_k I(a_k \leq X)}{w(X)} I(X \leq x) \right\} = G(x). \quad (10)$$

THEOREM 4. *Suppose that $N_k/E(N_k) \rightarrow 1$ in probability. Then for all x , $\hat{G}(x) \rightarrow G(x)$ in probability as $\nu \rightarrow \infty$, i.e., the independence likelihood estimator is consistent.*

THEOREM 5. *Suppose that $N_k/E(N_k) \rightarrow 1$ in probability. For any given x , let $U^{(\nu)}(x) = \sum_{k=1}^K c_k(x) \nu^{1/2} (N_k - \eta_k \nu)$, where $c_k(x) = \gamma_k(x) - \gamma_k(\infty)G(x)$. If $U^{(\nu)}(x) \rightarrow V(x)$ in distribution for some random variable $V(x)$, then*

$$M^{*1/2} \{\hat{G}(x) - G(x)\} \rightarrow \beta^{1/2} \{W(x) + V(x)\} \quad (11)$$

in distribution, where $W(x) \sim N\left[0, \sum_{k=1}^K \eta_k \sigma_k^2(x)\right]$ with $\sigma_k^2(x) = \text{var}[I(a_k \leq X)\{I(X \leq x) - G(x)\}/w(X)]$, and $W(x)$ and $V(x)$ are independent.

4. DATA ANALYSIS AND SIMULATION

As part of a monitoring program in Israel, four cross-sectional studies were 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, see Fluss et al. (2012) for more detail. In this section, we compare the nonparametric independence likelihood and conditional estimators of the distribution of length of hospitalization, trimmed at 30 days.

It is reasonable to assume that the numbers of urgent surgeries on different days of the week are exchangeable, suggesting the assumption $\eta_k = 1/K$. However, the daily schedule of elective

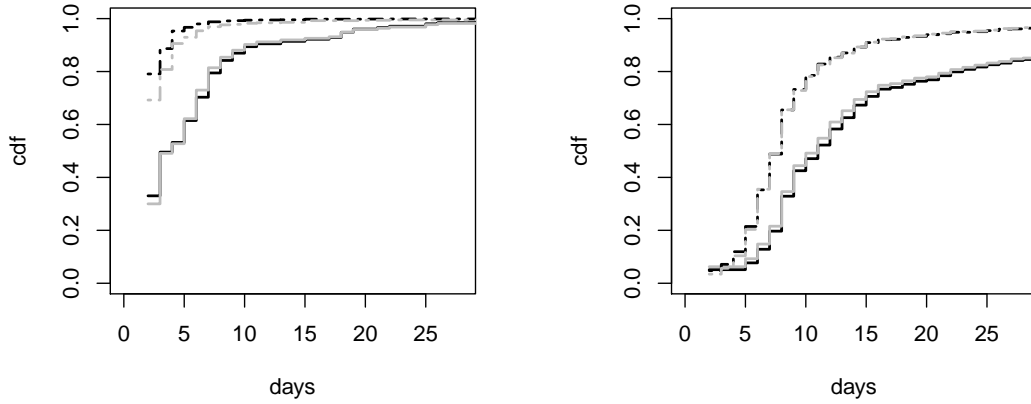


Fig. 2. Hospitalization time after bowel (left) and hernia (right) surgery. Solid lines - urgent surgery, broken line - elective surgery. Black - nonparametric independence likelihood, gray - nonparametric conditional likelihood.

surgeries suggests the possibility of unequal expected cohort sizes, see Fluss et al. (2012). It is therefore expected that the independence likelihood estimator with $\eta_k = 1/K$ will work well for urgent surgeries, but may fail for elective ones. The data comprise 587 bowel and 232 hernia surgeries of which 57% and 81% are elective, respectively.

Figure 2 presents the nonparametric estimates stratified by surgery type and 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 independence likelihood estimates are quite similar except for the group of elective hernia surgeries. In the latter case, the independence likelihood estimate for $G(2)$ is 0.79 with standard error of 0.026, while the conditional likelihood estimate is 0.69 with standard error of 0.039. Elective surgeries schedules in general do not satisfy the exchangeability assumption of cohort sizes which may explain the difference between the estimates. However, in this case the schedule may be known, leading to good estimates of the relevant η_k 's (Fluss et al., 2012). We remark that the data we analyzed were aggregated from several hospitals. This may moderate the problem of varying η_k 's if different hospitals have different schedules.

We conducted a simulation study in order to compare small sample properties of the conditional and the independence likelihood approaches, using K and sample size M^* that are somewhat similar to the above data. The results show a clear advantage of the independence likelihood approach for Poisson or relatively stable N_k 's, while for more variable N_k 's, the conditional approach is preferable. Details are given in the Supplementary Material.

5. DISCUSSION

Our results and simulations show that when the N_k 's are close to multinomial or to being constant, independence likelihood inference is preferable to the conditional approach. Scheduled

processes lead to constant N_k 's, and for many random allocation processes, the multinomial model appears to be reasonable. In certain non-exchangeable problems, the entrance process may be learnt from past data and the independence likelihood approach can be used; if not, one may resort to the conditional approach.

We conjecture that when the entrance process is a continuous renewal process, Theorem 1 can be extended to show that 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, see Gan & Yang (1989).

ACKNOWLEDGEMENT

This research was supported by The Israel Science Foundation. We thank the associate editor and the referees for helpful comments and suggestions.

SUPPLEMENTARY MATERIAL

Supplementary Material available at *Biometrika* online includes proofs, extensions of asymptotic results, examples, and detailed simulation results.

REFERENCES

- ASGHARIAN, M., M'LAN, C. E. & WOLFSON, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *J. Am. Statist. Assoc.*, **97**, 201–209.
- CHANDLER, R. E. & BATE, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, **94**, 167–183.
- COX, D. R. (1969). Some sampling problems in technology. in *New Developments in Survey Sampling*, Eds. N. L. Johnson & H. Smith, 506–527. New York: John Wiley & Sons.
- FERGUSON, T. S. (1996). *A Course in Large Sample Theory*. Chapman & Hall/CRC.
- FLUSS, R., MANDEL, M., FREEDMAN, L. S., SALZ-WEISS, I., EKKA-ZOHAR, A., HAKLAI, Z., GORDON, E.-S. & SIMCHEN, E. (2012). Correction of sampling bias in a cross-sectional study of post-surgical complications. *Statist. Med.*, **32**, 2467–2478.
- GAN, G. & YANG, S.-S. (1989). An elementary proof of the order-statistics characterization of the Poisson process. *Am. Statist.*, **43**, 45–46.
- LASLETT, G. M. (1982). The survival curve under monotone density constraints with applications to two-dimensional line segment processes. *Biometrika*, **69**, 153–160.
- LEHMANN, E. L., & CASELLA, G. (1998). *Theory of Point Estimation*. New York: Springer-Verlag.
- VARDI, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.*, **13**, 178–203.
- VARDI, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*, **76**, 751–761.
- VARIN, C., REID, N., & FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica*, **21** 5–42.
- WANG M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Am. Statist. Assoc.*, **86**, 130–143.
- WANG, M.-C., JEWELL, N. P. & TSAI, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.*, **14**, 1597–1605.
- WOODROOFE, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.*, **13**, 163–177.

[Received April 2012. Revised September 2012]