Supplementary material for Estimation from Cross-Sectional Samples under Bias and Dependence

Micha Mandel and Yosef Rinott
The Hebrew University of Jerusalem

March 12, 2014

A Proofs

A.1 Proof of Theorem 1

We start by showing that $\{A_j^*, X_j^*\}$ are independence if and only if $(N_1^*, \ldots, N_K^*) \mid M^*$ has a multinomial distribution. Any permutation of the ages is equally likely since the indices of individuals in the sample are non-informative, and therefore

$$\mathrm{pr}(\{A_j^* = a_j^*\}_{j=1}^{M^*} \mid M^* = m^*) = \frac{\prod_{k=1}^K n_k^*!}{m^*!} P\big(N_1^* = n_1^*, \ldots, N_K^* = n_K^* \mid M^* = m^*\big), \qquad \text{(A.1)}$$

where $n_k^* = \sum_{j=1}^{m^*} I(a_j^* = a_k)$. Under our model, and conditionally on $\{A_j^*\}$, the sojourn times $\{X_j^*\}$ are independent having the distribution $\mathrm{pr}(X_j^* \le x \mid A_j^* = a) = G(x)/\bar{G}(a-)$ for $a \le x$. Thus, the joint density of $(\{A_j^*\}, \{X_j^*\})$ conditionally on $\{M^* = m^*\}$ at points satisfying $\{a_j^* \le x_j^*\}$ is

$$\frac{\prod_{k=1}^K n_k^*!}{m^*!} P\big(N_1^* = n_1^*, \ldots, N_K^* = n_K^* \mid M^* = m^*\big) \prod_{j=1}^{m^*} dG(x_j^*)\bar{G}(a_j^*-)$$
$$= \frac{\prod_{k=1}^K n_k^*!}{m^*!} P\big(N_1^* = n_1^*, \ldots, N_K^* = n_K^* \mid M^* = m^*\big) \prod_{j=1}^{m^*} dG(x_j^*) \prod_{k=1}^{K} \{\bar{G}(a_k-)\}^{-n_k^*}. \qquad \text{(A.2)}$$

Independence means that the above density can be written as a product of the joint probabilities of (A_j^*, X_j^*), and by Equation (2) of the paper, this product is proportional to $\prod_{j=1}^{m^*} dG(x_j^*) \prod_{k=1}^K \{\mathrm{pr}(A_j = a_k)\}^{n_k^*}$. By equating the latter product to (A.2) we see that $(N_1^*, \ldots, N_K^*) \mid M^* = m^*$ has the distribution $\mathrm{Mult}(m^*, p_1^*, \ldots, p_K^*)$, where $p_k^* \propto P(A = a_k)\bar{G}(a_k-)$. On the other hand, if $(N_1^*, \ldots, N_K^*) \mid M^*$ has a multinomial distribution, it is readily seen that (A.2) is a product of terms depending on j, hence the pairs are independent. This completes the first step of the proof.

The main part of the theorem now follows from the following lemma that is of interest of its own

Lemma 1. *Let $(N_1, \ldots, N_K)$ be a vector of random variables taking values in $\mathbb{N}^K$, where $\mathbb{N}$ denotes the non-negative integers, and assume that*

$$N_1^*, \ldots N_K^* \mid N_1, \ldots, N_K \sim \perp\!\!\!\perp_{k=1}^K \mathrm{Bin}(N_k, \pi_k), \qquad \text{(A.3)}$$

that is, $(N_1^*, \ldots, N_K^*)$ are obtained by coordinate-wise thinning as follows: conditioned on $(N_1, \ldots, N_K)$ the variables $N_1^*, \ldots, N_K^*$ are independent with $N_k^* \mid N_1, \ldots, N_K \sim Bin\ (N_k, \pi_k)$ for some parameters $\pi_k \in [0,1]$, $k = 1, \ldots, K$.

The following two conditions are equivalent
Condition A: $(N_1^*, \ldots, N_K^* \mid \sum_{k=1}^K N_k^* = m^*) \sim \mathrm{Mult}(m^*, p^* = (p_1^*, \ldots, p_K^*))$ *for each m^*, for some probability vector p^*.*
Condition B: $(N_1, \ldots, N_K \mid \sum_{k=1}^K N_k = m) \sim \mathrm{Mult}(m, p = (p_1, \ldots, p_K))$ *for each m for some probability vector p.*

Remark 1. *Let $M = \sum_{k=1}^K N_k$, and Let $M^* = \sum_{k=1}^K N_k^*$. If $\mu_k = E(N_k)$ then $E(N_k^*) = \mu_k \pi_k$. Now if Condition A holds then $E(N_k^* \mid M^*) = M^* p_k^*$ and therefore $E(N_k^*) = E(M^*) p_k^*$, and it follows that $p_k^* = \mu_k \pi_k / \sum_i \mu_i \pi_i$. Also, $\mu_k = E(M) p_k$, and we conclude that*

$$p_k^* = p_k \pi_k / \sum_i p_i \pi_i. \tag{A.4}$$

Proof of Lemma 1. We first prove that Condition B $\Rightarrow$ Condition A. Let $(N_1, \ldots, N_K) \sim P_0$ be some vector that satisfies Condition B. Given $M = \sum_{k=1}^K N_k$, construct new variables on a common probability space, $(N_{1i}, \ldots, N_{Ki}) \sim \mathrm{Mult}(1, (p_1, \ldots, p_K))$ independent for $i = 1, \ldots, M$. Obviously, $\sum_{i=1}^M (N_{1i}, \ldots, N_{Ki}) \sim P_0$. Further, construct $(N_{1i}^*, \ldots, N_{Ki}^*) \mid (N_{1i}, \ldots, N_{Ki})$ independently such that $N_{ki}^* \mid (N_{1i}, \ldots, N_{Ki}) \sim \mathrm{Bin}(N_{ki}, \pi_k)$ and set $(N_1^*, \ldots, N_K^*) = \sum_{i=1}^M (N_{1i}^*, \ldots, N_{Ki}^*)$. Obviously, $(N_1^*, \ldots, N_K^*)$ satisfies (A.3) when conditioning is on $\sum_{i=1}^M (N_{1i}, \ldots, N_{Ki})$. Now

$$(N_{1i}^*, \ldots, N_{Ki}^*, N_{1i} - N_{1i}^*, \ldots, N_{Ki} - N_{Ki}^*) \sim \mathrm{Mult}(1, (p_1\pi_1, \ldots, p_K\pi_K, p1(1-\pi_1), \ldots, p_K(1-\pi_K)))$$

are independent for $i = 1, \ldots, M$ so their sum is multinomial, and hence its marginals (given their totals), with probabilities given by (A.4).

We now prove that Condition A $\Rightarrow$ Condition B.

Setting $N^* = (N_1^*, \ldots, N_K^*)$ and $t = (t_1, \ldots, t_K)$, we compute the generating function $g_{N^*}(t) = E(\prod_k t_k^{N_k^*})$ in two ways. First, using condition A we have

$$E\left(\prod_k t_k^{N_k^*} \mid M^*\right) = \sum_{i_1 + \ldots + i_K = M^*} \binom{M^*}{i_1 \ldots i_K} \prod_k (p_k^* t_k)^{i_k} = \left(\sum_{k=1}^K p_k^* t_k\right)^{M^*} = \left(\sum_{k=1}^K p_k^* t_k\right)^{\sum_k N_k^*}.$$

We compute the expectation of the latter expression to obtain $g_{N^*}(t)$ by first conditioning on $(N_1, \ldots, N_K)$ using (A.3) and then taking expectation with respect to $(N_1, \ldots, N_K)$. We readily obtain $E\left\{\left(\sum_{k=1}^K p_k^* t_k\right)^{\sum_k N_k^*} \mid (N_1, \ldots, N_K)\right\} = \prod_k E\left(s^{X_k}\right)$, where $s = \sum_{k=1}^K p_k^* t_k$ and $X_k \sim$ Bin (N_k, π_k). The latter binomial generating function is given by $E(s^{X_k}) = (1 - \pi_k + \pi_k s)^{N_k}$ and we conclude that

$$g_{N^*}(t) = E\{\prod_k (1 - \pi_k + \pi_k s)^{N_k}\} = g_N(1 - \pi_1 + \pi_1 s, \ldots, 1 - \pi_K + \pi_K s). \tag{A.5}$$

We compute g_{N^*} again by first conditioning on $(N_1, \ldots, N_K)$ as follows:

$$\begin{aligned} g_{N^*}(t) &= E\left\{E\left(\prod_k t_k^{N_k^*} \mid N_1, \ldots, N_K\right)\right\} = E\left\{\prod_k (1 - \pi_k + \pi_k t_k)^{N_k}\right\} \\ &= g_N(1 - \pi_1 + \pi_1 t_1, \ldots, 1 - \pi_K + \pi_K t_K). \end{aligned} \tag{A.6}$$

From (A.5) and (A.6) it follows that $g_N(1-\pi_1+\pi_1 t_1,\ldots,1-\pi_K+\pi_K t_K)$ is constant over the set $T_s=\{t : \sum_{k=1}^K p_k^* t_k = s\}$ for any s, and for fixed $\pi_1,\ldots,\pi_K$ we can write

$$g_N(1-\pi_1+\pi_1 t_1,\ldots,1-\pi_K+\pi_K t_K) = g_N(1-\pi_1+\pi_1 s,\ldots,1-\pi_K+\pi_K s) = h(s). \quad \text{(A.7)}$$

Recall

$$\frac{\partial^{n_1}\cdots\partial^{n_K}}{\partial t_1^{n_1}\cdots\partial t_K^{n_K}} g_N(t)\Big|_{t=0} = \prod_k n_k! P(N_1=n_1,\ldots,N_K=n_K).$$

In the same way

$$\frac{\partial^{n_1}\cdots\partial^{n_K}}{\partial t_1^{n_1}\cdots\partial t_K^{n_K}} g_N(1-\pi_1+\pi_1 t_1,\ldots,1-\pi_K+\pi_K t_K)\Big|_{\{t_k=(\pi_k-1)/\pi_k,\ k=1,\ldots,K\}}$$
$$= \prod_k (\pi_k^{n_k} n_k!) P(N_1=n_1,\ldots,N_K=n_K). \quad \text{(A.8)}$$

On the other hand from (A.7), and recalling $s=\sum_{k=1}^K p_k^* t_k$ we have

$$\frac{\partial^{n_1}\cdots\partial^{n_K}}{\partial t_1^{n_1}\cdots\partial t_K^{n_K}} g_N(1-\pi_1+\pi_1 t_1,\ldots,1-\pi_K+\pi_K t_K)\Big|_{\{t_k=(\pi_k-1)/\pi_k,\ k=1,\ldots,K\}}$$
$$= \prod_k (p_k^*)^{n_k}\, h^{(m)}\left(\sum_k p_k^*(\pi_k-1)/\pi_k\right) \quad \text{(A.9)}$$

where $h^{(m)}$ denotes the mth derivative of h defined in (A.7), and $m=\sum_k n_k$. Set $\sum_k p_k^*(\pi_k - 1)/\pi_k = \eta$. From (A.8) and (A.9) it follows that

$$P(N_1=n_1,\ldots,N_K=n_K) = h^{(m)}(\eta)\prod_k\left(\frac{p_k^*}{\pi_k}\right)^{n_k}\prod_k\frac{1}{n_k!}, \quad \text{(A.10)}$$

and

$$P(N_1+\ldots+N_K=m) = h^{(m)}(\eta)\frac{1}{m!}\left(\sum_k\frac{p_k^*}{\pi_k}\right)^m, \quad \text{(A.11)}$$

so we conclude that

$$P(N_1=n_1,\ldots,N_K=n_K \mid N_1+\ldots+N_K=m) = \binom{m}{n_1\ldots n_k}\prod_k\left(\frac{p_k^*}{\pi_k}\Big/\sum_k\frac{p_k^*}{\pi_k}\right)^{n_k}, \quad \text{(A.12)}$$

which is multinomial as required, with parameters that agree with (A.4). □

Finally, the independent case is established by the following lemma, which is similar to a characterization of the Poisson distribution given by Chatterji (1963).

Lemma 2. *Let $N_1,\ldots,N_K$ be independent random variables taking values in $\mathbb{N}$. Then, $(N_1,\ldots,N_K \mid \sum_{k=1}^K N_k = m) \sim \text{Mult}(m, p=(p_1,\ldots,p_K))$ for each m for some probability vector p, if and only if $N_k \sim \text{Poisson}(cp_k)$ for some positive c.*

Proof. The fact that for $N_k \sim \text{Poisson}(cp_k)$ independent $(N_1,\ldots,N_K \mid \sum_{k=1}^K N_k = m) \sim \text{Mult}(m,p)$ is well known. Assume $(N_1,\ldots,N_K \mid \sum_{k=1}^K N_k = m) \sim \text{Mult}(m,p)$. For $K=2$, $1\le a\le m$, and $k=1,2$,

$$\frac{p_k}{1-p_k}\frac{m-a+1}{a} = \frac{\text{pr}(N_k=a \mid N_1+N_2=m)}{\text{pr}(N_k=a-1 \mid N_1+N_2=m)} = \frac{\text{pr}(N_k=a)\text{pr}(N_{3-k}=m-a)}{\text{pr}(N_k=a-1)\text{pr}(N_{3-k}=m-a+1)}$$

which for $a = m$ gives for any m

$$\frac{\mathrm{pr}(N_k = m)}{\mathrm{pr}(N_k = m-1)} = \frac{p_k}{1-p_k}\frac{\mathrm{pr}(N_{3-k} = 1)}{\mathrm{pr}(N_{3-k} = 0)}\frac{1}{m} = \frac{c_k}{m},$$

a ratio which implies $N_k \sim \mathrm{Poisson}(c_k)$. It is easy to see directly that $c_k = E(N_k) = p_k E(N_1 + N_2)$, hence $c = E(N_1 + N_2)$. For $K > 2$ one can prove $N_k \sim \mathrm{Poisson}(cp_k)$ by writing $N_{-k} = \sum_{j\neq k} N_j$ and using $(N_k, N_{-k}) \mid \{\sum_{j=1}^K N_j = m\} \sim \mathrm{Mult}(m, p_k, 1-p_k)$.

□

A.2 Proofs of asymptotic results

Sketch of proof of Theorem 2. Recall that $\{X_{ki}\}$ are independent and identically distributed and are independent of $\{N_k\}$. Denote the true parameter value by θ_0. By (5), with $X^* \sim G^*(\cdot\,;\theta_0)$,

$$\frac{1}{M^*}\ell(\theta) = \frac{\frac{1}{\nu}\sum_{k=1}^K\sum_{i=1}^{N_k} I(a_k \le X_{ki}) \log\frac{w(X_{ki})g(X_{ki};\theta)}{\beta_\theta}}{\frac{1}{\nu}\sum_{k=1}^K\sum_{j=1}^{N_k} I(a_k \le X_{ki})} \to E_{\theta_0}[\log\{g^*(X^*;\theta)\}] \tag{A.13}$$

in probability, where the limit is obtained as follows. Starting with the denominator and recalling that $N_k/(\nu\eta_k) \to 1$ in probability, the law of large numbers implies

$$\sum_{k=1}^K \frac{1}{\nu}\sum_{i=1}^{N_k} I(a_k \le X_{ki}) = \sum_{k=1}^K \eta_k \frac{N_k}{\nu\eta_k}\frac{1}{N_k}\sum_{i=1}^{N_k} I(a_k \le X_{ki}) \to \sum_{k=1}^K \eta_k \bar{G}_{\theta_0}(a_k-) = \beta_{\theta_0}. \tag{A.14}$$

The same reasoning applied to the numerator of (A.13) yields

$$\begin{aligned}
&\sum_{k=1}^K \eta_k \frac{N_k}{\nu\eta_k}\frac{1}{N_k}\sum_{i=1}^{N_k} I(a_k \le X_{ki}) \log\frac{w(X_{ki})g(X_{ki};\theta)}{\beta_\theta} \to \sum_{k=1}^K \int_0^\infty \eta_k I(a_k \le x)\log\{g^*(x;\theta)\}dG(x;\theta_0) \\
&= \int_0^\infty \sum_{k=1}^K \eta_k I(a_k \le x)\log\{g^*(x;\theta)\}dG(x;\theta_0) = \beta_{\theta_0}\int_0^\infty \log\{g^*(x;\theta)\}\frac{w(x)dG(x;\theta_0)}{\beta_{\theta_0}} \\
&= \beta_{\theta_0}\int_0^\infty \log\{g^*(x;\theta)\}dG^*(x;\theta_0) = \beta_{\theta_0}E_{\theta_0}[\log\{g^*(X^*,\theta)\}].
\end{aligned} \tag{A.15}$$

Equations (A.15) and (A.14) imply (A.13). Identifiability and the information inequality assert that $E_{\theta_0}[\log\{g^*(X^*,\theta)\}]$ obtains its maximum at $\theta = \theta_0$; standard arguments guarantee the existence of a consistent sequence of roots (e.g., Lehmann and Casella 1998). □

Example A.1 (Inconsistency of the independence likelihood estimator)**.** *Consider the model* $X_{ki} \sim \mathrm{Exp}(\theta)$ *with* $K = 2$, $a_1 = 0$, $a_2 = 1$ *and* $\eta_1 = \eta_2$, *the exchangeable case. It is easy to see that* $X^*_{ki} - k + 1 \sim \mathrm{Exp}(\theta)$. *Simple calculations show that the independence likelihood estimator,* $\hat\theta$, *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* $\nu \to \infty$, $\bar{X}_k^* \to k - 1 + \theta^{-1}$, $k = 1,2$, *and* $N_k^*/N_k \to e^{-\theta a_k}$ *in probability, 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 independence likelihood estimator is consistent if $N_1/N_2 \to \eta_1/\eta_2 = 1$ in probability, but not otherwise. As a concrete example, let N_k be independent and $N_k = \nu/4$ or $3\nu/4$ with probability $1/2$ each so that N_1/N_2 takes the values $1/3$, 1, and 3 with corresponding probabilities $1/4$, $1/2$, and $1/4$, and the estimator converges to a non-degenerate random variable, and therefore is inconsistent.

Proof of Theorem 3. Using consistency, Taylor expansion of $0 = \partial\ell(\hat{\theta}_\nu)/\partial\theta$ around θ_0, and standard arguments yield the approximation

$$M^{1/2}(\hat{\theta}_\nu - \theta_0) \approx \frac{M^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{N_k}\frac{\partial}{\partial\theta}h_k(X_{ki};\theta_0)}{-M^{-1}\sum_{k=1}^{K}\sum_{i=1}^{N_k}\frac{\partial^2}{\partial\theta^2}h_k(X_{ki};\theta_0)}. \tag{A.16}$$

The conditions on $N_1,\ldots,N_K$ imply $N_k/M \to \eta_k$ in probability and the denominator of (A.16) converges to $-\sum_{k=1}^{K}\eta_k E_{\theta_0}[\partial^2 h_k(X,\theta_0)/\partial\theta^2]$. The analysis of the numerator is more complicated:

$$M^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{N_k}\frac{\partial}{\partial\theta}h_k(X_{ki};\theta_0) = \sum_{k=1}^{K}M^{-1/2}\sum_{i=1}^{N_k}\Big(\frac{\partial}{\partial\theta}h_k(X_{ki};\theta_0) - c_k\Big) + M^{-1/2}\sum_{k=1}^{K}c_kN_k. \tag{A.17}$$

A multivariate version of the proof in Rényi (1957) of Anscombe's Theorem and $N_k/(\eta_k M) \to 1$ in probability imply that

$$\frac{{N_k}^{1/2}}{(\eta_k M)^{1/2}}\frac{1}{{N_k}^{1/2}}\sum_{i=1}^{N_k}\Big(\frac{\partial}{\partial\theta}h_k(X_{ki};\theta_0) - c_k\Big)$$

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

Therefore, in (A.17),

$$\sum_{k=1}^{K}\frac{\eta_k^{1/2}}{(\eta_k M)^{1/2}}\sum_{i=1}^{N_k}\Big(\frac{\partial}{\partial\theta}h_k(X_{ki};\theta_0) - c_k\Big) \to W \tag{A.18}$$

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

Turning to the second term in (A.17), we have $M^{-1/2}\sum_{k=1}^{K}c_kN_k = M^{-1/2}\sum_{k=1}^{K}c_k(N_k - \eta_k\nu) \to V$ in distribution, where we used the facts that $\sum_k \eta_k c_k = 0$, proved below, and $\nu^{-1}M \to 1$ in probability. Now, (7) is obtained from (A.16) by

$$\frac{M^*}{M} = \sum_{k=1}^{K}\frac{N_k}{M}\frac{1}{N_k}\sum_{j=1}^{N_k}I(a_k \le X_{kj}) \to \sum_{k=1}^{K}\eta_k\bar{G}(a_k-;\theta_0) = \beta_{\theta_0}. \tag{A.19}$$

It remains to show $\sum_k \eta_k c_k = 0$ and independence of V and W. Interchanging the order of integration and differentiation and recalling $\sum_{k=1}^{K}\eta_k I(a_k \le x) = w(x)$, we obtain as in (A.15)

$$\sum_{k=1}^{K}\eta_k c_k = \frac{\partial}{\partial\theta}\int_0^\infty\sum_{k=1}^{K}\eta_k I(a_k \le x)\log\{g^*(x;\theta)\}dG(x;\theta_0)\Big|_{\theta=\theta_0} = \beta_{\theta_0}\frac{\partial}{\partial\theta}E_{\theta_0}[\log\{g^*(X^*,\theta)\}]\Big|_{\theta=\theta_0},$$

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

To prove independence of W and V, note that the assumptions on the entrance process imply $N_k/(\eta_k M) \to 1$ in probability, and by (A.18), it suffices to prove asymptotic independence of $U^{(\nu)}$ and

$$W^{(\nu)} = \sum_{k=1}^{K}\Big(\frac{\eta_k}{N_k}\Big)^{1/2}\sum_{j=1}^{N_k}\Big(\frac{\partial}{\partial\theta}h_k(X_{kj};\theta_0) - c_k\Big).$$

Given $\epsilon > 0$ let n_0 be such that $n_k > n_0$ for all k implies $|\mathrm{pr}(W^{(\nu)}/\sigma \le t \mid \{N_k = n_k\}) - \Phi(t)| < \epsilon$, where $\sigma^2 = \sum_{k=1}^K \eta_k \mathrm{var}_{\theta_0}\{\partial h_k(X;\theta_0)/\partial\theta\}$, and let ν be such that $\mathrm{pr}(N_k^{(\nu)} > n_0$ for all $k) > 1-\epsilon$. For n_k's $> n_0$ we have

$$\mathrm{pr}(W^{(\nu)}/\sigma \le t, U^{(\nu)} \le u \mid \{N_k = n_k\}) \le (\Phi(t)+\epsilon) I\left(\sum_{k=1}^K c_k \frac{n_k - \eta_k \nu}{\nu^{1/2}} \le u\right).$$

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

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

A similar lower bound completes the proof. □

Proof of Theorem 4. Since $N_k \to \infty$ in probability, the weak law of large numbers yields

$$\frac{1}{N_k}\sum_{i=1}^{N_k} w(X_{ki})^{-1} I(a_k \le X_{ki} \le x) \to \gamma_k(x)$$

in probability, which holds also for $x = \infty$. In addition, the assumptions imply $N_k/M \to \eta_k$ in probability, and by (9) and (10), $\hat{G}(x) \to \sum_{k=1}^K \eta_k \gamma_k(x) \Big/ \sum_{k=1}^K \eta_k \gamma_k(\infty) = G(x)$. □

Proof of Theorem 5. By (9)

$$M^{1/2}(\hat{G}(x) - G(x)) = \frac{M^{-1/2}\sum_{k=1}^K \sum_{i=1}^{N_k} w(X_{ki})^{-1} I(a_k \le X_{ki})[I(X_{ki} \le x) - G(x)]}{M^{-1}\sum_{k=1}^K \sum_{i=1}^{N_k} w(X_{ki})^{-1} I(a_k \le X_{ki})}. \quad \text{(A.20)}$$

The denominator in (A.20) converges in probability to 1 since $N_k^{-1}\sum_{i=1}^{N_k} w(X_{ki})^{-1} I(a_k \le X_{ki}) \to \gamma_k(\infty)$ by the Law of Large Numbers, $N_k/M \to \eta_k$, and $\sum_{k=1}^K \eta_k \gamma_k(\infty) = 1$, see (10).

Setting

$$S_{ki}(x) = w(X_{ki})^{-1} I(a_k \le X_{ki})[I(X_{ki} \le x) - G(x)],$$

we have $E\{S_{ki}(x)\} = c_k(x)$, and $\sum_k \eta_k c_k(x) = 0$. As in (A.19), $M^*/M \to \beta$ in probability; also, in probability, $M/\nu \to 1$ and we obtain that $M^{*1/2}\{\hat{G}(x) - G(x)\}$ is asymptotically equivalent to

$$\beta^{1/2}\sum_{k=1}^K \frac{1}{M^{1/2}}\sum_{i=1}^{N_k}\{S_{ki}(x) - c_k(x)\} + \beta^{1/2}\sum_{k=1}^K \frac{c_k(x)(N_k - \eta_k\nu)}{\nu^{1/2}}. \quad \text{(A.21)}$$

We have $M^{-1/2}\sum_{i=1}^{N_k}\{S_{ki}(x) - c_k(x)\} \to N(0, \eta_k\sigma_k^2(x))$ in distribution, and therefore,

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

Independence of $W(x)$ and $V(x)$ follows by reasons as in the proof of Theorem 3. □

A.3 Asymptotic normality in the multi-parameter case

Suppose that θ is p-dimensional and that the independence likelihood estimator is consistent. Under standard regularity conditions, Taylor approximation gives

$$M^{1/2}(\hat{\theta} - \theta_0) \approx M H^{-1}(\theta_0)\frac{1}{M^{1/2}} D_\ell(\theta_0), \quad \text{(A.22)}$$

where $D_\ell(\theta_0) = (\partial\ell(\theta_0)/\partial\theta_1, \ldots, \partial\ell(\theta_0)/\partial\theta_p)^t$ and $H(\theta_0) = (\partial^2\ell(\theta_0)/\partial\theta_s\partial\theta_t)$ is the $p \times p$ matrix of second derivatives. Under conditions as in Theorem 3

$$MH^{-1}(\theta_0) = \left(\sum_{k=1}^{K} \frac{N_k}{M} \frac{1}{N_k} \sum_{i=1}^{N_k} \frac{\partial^2}{\partial\theta_s\partial\theta_t} h_k(X_{ki};\theta_0)\right)^{-1} \to \left(\sum_{k=1}^{K} \eta_k E_{\theta_0}\{\frac{\partial^2}{\partial\theta_s\partial\theta_t} h_k(X,\theta_0)\}\right)^{-1}$$

in probability.

Next, write $M^{-1/2}D_\ell(\theta_0)$ as a sum of two vectors:

$$\left(M^{-1/2}\sum_{k=1}^{K}\sum_{i=1}^{N_k} \frac{\partial}{\partial\theta_j} h_k(X_{ki};\theta_0)\right) = \left(\sum_{k=1}^{K} M^{-1/2} \sum_{i=1}^{N_k} \left(\frac{\partial}{\partial\theta_j} h_k(X_{ki};\theta_0) - c_{kj}\right)\right) + M^{-1/2}\left(\sum_{k=1}^{K} N_k c_{kj}\right), \tag{A.23}$$

where $c_{kj} = E_{\theta_0}\{\partial h_k(X;\theta_0)/\partial\theta_j\}$, and note that

$$\left(M^{-1/2}\sum_{i=1}^{N_k}\left(\frac{\partial}{\partial\theta_1} h_k(X_{ki};\theta_0) - c_{k1}\right), \ldots, M^{-1/2}\sum_{i=1}^{N_k}\left(\frac{\partial}{\partial\theta_p} h_k(X_{ki};\theta_0) - c_{kp}\right)\right), \quad k = 1,\ldots,K$$

converge jointly to independent zero mean normal vectors with corresponding covariances $\eta_k \mathrm{cov}\{\partial h_k(X;\theta_0)/\partial\theta_s, \partial h_k(X;\theta_0)/\partial\theta_t\}$. Therefore, the first term on the right hand side of (A.23) satisfies $\left(\sum_{k=1}^{K} M^{-1/2}\sum_{i=1}^{N_k}\left(\partial h_k(X_{ki};\theta_0)/\partial\theta_j - c_{kj}\right)\right) \to W$ in distribution, where

$$W \sim N_p\Big(0, \sum_{k=1}^{K} \eta_k \mathrm{cov}\left\{\frac{\partial}{\partial\theta_s} h_k(X;\theta_0), \frac{\partial}{\partial\theta_t} h_k(X;\theta_0)\right\}\Big). \tag{A.24}$$

The second term in (A.23) vanishes if the N_k's are constant, and otherwise can be treated as in the single parameter case.

B Asymptotics with K

We study the asymptotic properties of the independence likelihood estimator in the following setting. There is a sequence $\{A_k\}_{k=1}^K$ of entrance points, a sequence $\{N_k\}_{k=1}^K$ of non-negative integer numbers, and a sequence $\{X_{ki}\}_{k=1,i=1}^{K,\ N_k}$ of lifetimes. We assume that the sequences are independent, each consisting of independent and identically distributed random variables. Specifically, we assume $A_k \sim W$, $N_k \sim P$, and $X_{ki} \sim G$, with the technical identifiability requirement that $W(x_{\min}) > 0$, where $x_{\min}$ is the left limit of the support of G. We assume that $\nu := E(N_k) < \infty$ and study the independence likelihood estimator when $K \to \infty$. This model assumes exchangeability because the distribution of N_k is independent of k. The analysis is much simpler than in the setting considered in the paper as the likelihood (5) in the paper becomes a sum of K independent and identically distributes random variables and $K \to \infty$.

First recall that the marginal law of the X^*'s is $dG^*(x) = W(x)dG(x)/\beta$, where here the weight function is given by $W = P(A \le x)$, and $\beta = P(A \le X)$. We prove consistency and asymptotic normality for the parametric case. As in the proof of Theorem 2, we show that

$$\frac{1}{M^*}\ell(\theta) = \frac{\frac{1}{\nu K}\sum_{k=1}^{K}\sum_{i=1}^{N_k} I(A_k \le X_{ki}) \log \frac{W(X_{ki})dG(X_{ki};\theta)}{\beta_\theta}}{\frac{1}{\nu K}\sum_{k=1}^{K}\sum_{j=1}^{N_k} I(A_k \le X_{ki})} \to E_{\theta_0}[\log\{dG^*(X^*;\theta)\}] \tag{B.1}$$

in probability. Starting with the denominator, we have that $\sum_{j=1}^{N_k} I(A_k \le X_{ki})$ $k = 1,\ldots,K$ are independent and identically distributed random variables with expectation $\nu\beta$ by Wald's

Lemma, so by the law of large numbers, the denominator converges to β_θ. Similarly,

$$\sum_{i=1}^{N_k} I(A_k \leq X_{ki}) \log \frac{W(X_{ki})dG(X_{ki};\theta)}{\beta_\theta}$$

are independent and identically distributed with expectation

$$\nu \int \log \frac{W(x)dG(x;\theta)}{\beta_\theta} W(x)dG(x) = \beta_\theta E_{\theta_0}[\log\{dG^*(X^*;\theta)\}].$$

Using again the law of large numbers, (B.1) is obtained, and the proof of consistency follows the arguments in Theorem 2.

The asymptotic distribution is simpler than in Theorem 3. Denote

$$h(X, A; \theta) = \frac{\partial}{\partial\theta} I(A \leq X) \log \frac{W(X)dG(X;\theta_0)}{\beta_{\theta_0}}.$$

Starting with a term similar to (A.16), we have:

$$K^{1/2}(\hat{\theta}_K - \theta_0) \approx \frac{\frac{1}{K^{1/2}} \sum_{k=1}^K \sum_{i=1}^{N_k} \frac{\partial}{\partial\theta} h(X_{ki}, A_k; \theta_0)}{\frac{-1}{K} \sum_{k=1}^K \sum_{i=1}^{N_k} \frac{\partial^2}{\partial\theta^2} h(X_{ki}, A_k; \theta_0)}. \tag{B.2}$$

The denominator converges to $-\nu E_{\theta_0}\{\partial^2 h(X, A; \theta_0)/\partial\theta^2\}$. For the numerator, note that

$$\begin{aligned}
E_{\theta_0}\{\tfrac{\partial}{\partial\theta} h(X_{ki}, A_k; \theta_0)\} &= E_{\theta_0}\{\tfrac{\partial}{\partial\theta} I(A \leq X) \log \tfrac{W(X)dG(X;\theta_0)}{\beta_{\theta_0}}\} \\
&= E_{\theta_0}\{\tfrac{\partial}{\partial\theta} \log dG^*(X;\theta_0)W(X)\} \\
&= \beta_{\theta_0} E_{\theta_0}\{\tfrac{\partial}{\partial\theta} \log dG^*(X^*;\theta_0)\} = 0,
\end{aligned}$$

so the numerator converges to a zero mean normal variable with variance

$$\mathrm{var}_{\theta_0}\left(\sum_{i=1}^{N_k} \frac{\partial}{\partial\theta} h(X_{ki}, A_k; \theta_0)\right) = \nu \mathrm{var}_{\theta_0}\left(\frac{\partial}{\partial\theta} h(X, A; \theta_0)\right).$$

Thus, unlike the setting where K is fixed and $\nu \to \infty$, the asymptotic distribution is always normal.

Remark 2. *A similar analysis applies for the case where the number of independent cross-sectional samples increases, that is, K is fixed, N_{kh} is the number of patients in sample h who entered at time $-a_k$, and $h \to \infty$.*

C Asymptotic distribution of V - examples

Example C.1 (Independent N_k's, Normal limit)**.** *Theorem 3 implies that $M^{*1/2}(\hat{\theta}_\nu - \theta_0)$ 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. Suppose that N_k's are independent with $E(N_k) = \eta_k\nu$ and $\mathrm{var}(N_k) = \sigma_k^2$, then*

$$U^{(\nu)} = \sum_{k=1}^K \frac{c_k(N_k - \eta_k\nu)}{\sigma_k} \times \left(\frac{\sigma_k^2}{\nu}\right)^{1/2}.$$

If $(N_k - \eta_k\nu)/\sigma_k \to N(0,1)$ in distribution for $k = 1, \ldots, K$, and $\sigma_k^2/\nu \to b_k$ in probability, then $V \sim N(0, \sum b_k c_k^2)$; this includes the case $N_k \sim \mathrm{Poisson}(\eta_k\nu)$ where $\mathrm{var}(V) = \sum \eta_k c_k^2$. A smaller variance is obtained when $N_k \sim \mathrm{Binomial}(\nu, \eta_k)$ with $\mathrm{var}(V) = \sum \eta_k(1-\eta_k)c_k^2$. By Cramér's Theorem, as $\{c_k(N_k - \eta_k\nu)/\sigma_k\}$ are independent, the condition $b_k^{1/2}(N_k - \eta_k\nu)/\sigma_k \to N(0, b_k)$ is necessary for V and hence for the independence likelihood estimator to have a normal limit. If $b_k = 0$, e.g., for constant $N_k \equiv \eta_k\nu$, then $\nu^{-1/2}\sum_{k=1}^K c_k(N_k - \eta_k\nu) \to 0$, so $V = 0$.

Example C.2 (Dependent N_k's)**.** *As a simple but natural example of a normal limit in the presence of dependence, let $N_k = N_0' + N_k'$, where $N_0', N_1', \ldots, N_K'$ are such that $E(N_k) = \eta_k\nu$. We have $U = \nu^{-1/2}\sum_{k=1}^K c_k(N_k - \eta_k\nu) = \nu^{-1/2}\sum_{k=1}^K c_k\{N_k' - E(N_k')\} + \nu^{-1/2}\sum_{k=1}^K c_k\{N_0' - E(N_0')\}$. It is now easy to construct models having the same marginal distribution of the N_k's, but with different asymptotic distributions of V of Theorem 3, and therefore of the independence likelihood estimator. This is in contrast to consistency, which by Theorem 2 depends only on the marginal distributions of the cohort sizes. For example, the case of equal cohort sizes, $N_1 = \cdots = N_K \sim F$ corresponds to $N_1' = \cdots = N_K' = 0$ and $\eta_k = 1/K$, and recalling $\sum_k c_k\eta_k = 0$ it is easy to see that $V = 0$. On the other hand, if the $N_k' \sim F$'s are independent and $N_0' = 0$, then N_k's are independent having the same distribution F. In this case a non-degenerate normal limit was demonstrated above.*

Example C.3 (Multinomial model)**.** *Another natural model of dependent N_k's that leads to asymptotic normality is the following. See Remark 1 in the paper. Let $M = M^{(\nu)}$ satisfy $E(M) = \nu$, and $M/\nu \to 1$ in probability, corresponding to the assumptions of Theorem 3. If $(N_1, \ldots, N_k) \mid M \sim \mathrm{Mult}(M, (\eta_1, \ldots, \eta_K))$, then V is Gaussian. To see this, write $N_k = \sum_{j=1}^M I(Z_j = e_k)$, where $Z_j \sim \mathrm{Mult}(1, (\eta_1, \ldots, \eta_K))$ independently, and e_k is a vector of K coordinates with the kth being 1 and the rest 0. Then, $\nu^{-1/2}\sum_{k=1}^K c_k(N_k - \eta_k\nu) = \nu^{-1/2}\sum_{j=1}^M\sum_{k=1}^K c_k I(Z_j = e_k)$, which converges to the normal distribution by Anscombe's Theorem, see Rényi (1957).*

Example C.4 (The effect of dependence among the N_k's on $\mathrm{var}(V)$)**.** *If $\mathrm{pr}(N_1 = \ldots = N_K) = 1$ then $\mathrm{pr}(V = 0) = 1$, hence $\mathrm{var}(V) = 0$. The latter condition on the N_k's is the strongest form of dependence. This leads to the question of whether in natural cases $\mathrm{var}(V)$ is decreasing as a function of a suitable measure of dependence among the N_k's. We restrict the discussion to the case that the N_k's have equal expectations and variances, so that $\eta_k \equiv 1/K$, in which case $\sum_k c_k = 0$; however, it is easy to generalize.*

Let R denote the correlation matrix of $(N_1, \ldots, N_K)$. Then $\mathrm{var}(V)$ is proportional to $c^t Rc$, for $c = (c_1, \ldots, c_K)$. We consider two simple models for R with a dependence parameter ρ, the intraclass correlation model, and the autoregressive model. For the intraclass correlation matrix $R = (1-\rho)I + \rho 11^t$, where 1 here denotes a column vector of ones of length K. We have $c^t 11^t c = 0$ because $\sum_k c_k = 0$, and therefore $c^t Rc = (1-\rho)\sum_k c_k^2$, which is clearly decreasing in ρ and hence so is $\mathrm{var}(V)$.

Next consider the first order autoregressive correlation matrix with entries $r_{ij} = \rho^{|i-j|}$. Ignoring a proportionality constant we have $\partial\mathrm{var}(V)/\partial\rho = c^t Bc$, where $B = \partial R/\partial\rho$, having entries $b_{ij} = |i-j|\rho^{|i-j|-1}$. Since $b_{ij} = \lim_{t\downarrow 0} t^{-1}(1 - e^{-tb_{ij}})$ and $\sum_k c_k = 0$, it is easy to see that $c^t Bc \leq 0$, that is, B is conditionally negative definite, provided the matrix with entries $e^{-tb_{ij}}$ is positive definite. For $\rho = 1$ the latter matrix is again a first order autoregressive correlation matrix which is positive definite, and thus $\mathrm{var}(V)$ is decreasing in ρ near 1. However by direct calculations one can see that for $K \geq 4$ we do not have monotonicity of $\mathrm{var}(V)$ for all ρ's.

Example C.5 (Non-Normal limit)**.** *The limit of $\sum_{k=1}^K c_k(N_k - \nu)/\sqrt{\nu}$ may not be Normal, and may not exist. Let N_1, N_2 be independent with $E(N_k) = \nu/2$, and assume that $\mathrm{pr}(N_k = \nu - a) = \mathrm{pr}(N_k = \nu + a) = 1/2$ for some $a = a(\nu)$. In order that $N_k/E(N_k) \to 1$ in probability, a must* *satisfy* $a/\nu \to 0$. *Here* $\eta_1 = \eta_2 = 1/2$ *implying* $c_1 + c_2 = 0$ *and* $\sum_{k=1}^{2} c_k(N_k - \nu/2)/\nu^{1/2} = c_1(N_1 - N_2)/\nu^{1/2}$, *which takes the values 0 or* $\pm 2ac_1/\nu^{1/2}$. *For* $a = \nu^{1/2}$ *the limiting distribution is neither degenerate nor Normal, and for* $a = (2 + (-1)^\nu)\nu^{1/2}$ *the limit does not exist.*

D Parametric models with covariates

Suppose that for each observed sojourn time, X_j^*, we observe covariates denoted by Z_j^*. We aim to estimate the conditional distribution $G(x \mid z; \theta)$. The assumptions we made on the X's now apply to the pairs (X, Z)'s. Conditioning on the observed covariates values z_j^*, the independence likelihood of (4) is replaced by

$$L(\theta) = \prod_{j=1}^{m^*} \frac{w(x_j^*) g(x_j^* \mid z_j^*; \theta)}{\beta_\theta(z_j^*)}, \tag{D.1}$$

where $\beta_\theta(z) = E_\theta\{w(X) \mid Z = z\}$, and the independence likelihood estimator $\hat{\theta}$ is the value of θ that maximizes (D.1). Consistency and asymptotic normality in the sense of Theorems 2 and 3 can be proved in the same way, where the assumptions on $g(x;\theta)$ should hold for $g(x \mid z; \theta)$, and in (6)-(8) $\beta_{\theta_0} = E\{\beta_{\theta_0}(Z)\}$. Also $h_k(X_{ki};\theta)$ as defined in (5) is replaced by $h_k(X_{ki}, Z_{ki}; \theta) = I(a_k \le X_{ki}) \log\{w(X_{ki}) g(X_{ki} \mid Z_{ki}; \theta)/\beta_\theta(Z_{ki})\}$; we shall not repeat the proofs.

E Detailed results of simulation

We conducted a simulation study with $K = 20$ entrance points ($a_k = k - 1$) and a Gamma lifetime distribution with mean 12, variance 48. We considered several models of moderate sample sizes with $E(N_k) = 20$ for all k, $\beta = P(A \le X) = 0{\cdot}5875$ and therefore $E(M^*) = \beta E(M) = 235$, and larger sample sizes with $E(N_k) = 50$ for all k and $E(M^*) = 587$. We also considered a model with $E(N_k)$'s varying between about 15 to 26, and between 40 and 61.

The following models for the distribution of N_k were tested: independent Poisson entrance numbers; mixtures of Poissons: 1/2Pois(15)+1/2Pois(25) in the small sample size scenario, and 1/2Pois(43) + 1/2Pois(57) in the large sample size scenario, which reflect moderate deviation from the Poisson model; 1/2Pois(10) + 1/2Pois(30) and 1/2Pois(35) + 1/2Pois(65), reflecting large deviation from the Poisson model; independent Geometric entrance numbers; a constant number of entrances at each point; a symmetric multinomial model with $M \equiv 400$ and $M \equiv 1000$ for the small and large sample size scenarios, respectively; non-exchangeable N_k's, where entrances are independent following Poisson variables with $N_k \sim \text{Pois}(\exp(3{\cdot}27 - 0{\cdot}027k))$ in the small sample scenario, and $N_k \sim \text{Pois}(\exp(4{\cdot}12 - 0{\cdot}021k))$ 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 independence likelihood estimates of G, and averaged over the 1000 replications to obtain estimates for the MSE at the 10, 25, 50, 75, and 90 percentiles of G. Results are provided in Table E.1. As expected, the results show a clear advantage for the independence likelihood approach when N_k's are Poisson, or when they are relatively stable, such as constant N_k's or mixtures with moderate deviation from the Poisson model, while for more variable N_k's, the conditional approach is preferable.

Figures E.1 shows the ratio MSE(conditional)/MSE(independence likelihood) as a function of the variance of N_k. It shows that the ratio decreases with the variance, where the conditional approach and the independence likelihood approach are equally good when the variance is 2-3 times the expectation. It also reveals that the efficiency of the independence likelihood approach is maximal for the degenerate case, where $N_k \equiv N$.

$E(N_k)$	Model	Method	$G^{-1}(\cdot 10)$	$G^{-1}(\cdot 25)$	$G^{-1}(\cdot 50)$	$G^{-1}(\cdot 75)$	$G^{-1}(\cdot 90)$
20	Pois(20)	parm	1·15	1·17	1·21	1·24	1·20
		non-parm	1·00	1·06	1·17	1·22	1·08
	mix(15,25)	parm	1·10	1·10	1·13	1·18	1·19
		non-parm	1·05	1·06	1·10	1·18	1·15
	mix(10,30)	parm	0·83	0·82	0·81	0·84	0·90
		non-parm	0·88	0·83	0·86	0·86	0·92
	Geo(1/20)	parm	0·53	0·46	0·39	0·37	0·43
		non-parm	1·20	0·70	0·48	0·45	0·56
	Constant=20	parm	1·30	1·34	1·42	1·45	1·33
		non-parm	1·05	1·19	1·33	1·40	1·18
	multinom	parm	1·17	1·18	1·22	1·24	1·19
		non-parm	0·99	1·10	1·19	1·21	1·14
	inhomogeneous	parm	0·67	0·69	0·68	0·67	0·75
		non-parm	0·83	0·83	0·76	0·77	0·93
50	Pois(50)	parm	1·15	1·19	1·24	1·28	1·23
		non-parm	1·07	1·13	1·19	1·28	1·13
	mix(43,57)	parm	1·09	1·10	1·13	1·16	1·16
		non-parm	1·05	1·08	1·08	1·18	1·13
	mix(35,65)	parm	0·89	0·87	0·85	0·87	0·92
		non-parm	0·96	0·88	0·85	0·92	0·94
	Geo(1/50)	parm	0·31	0·26	0·21	0·20	0·23
		non-parm	0·56	0·32	0·24	0·23	0·32
	Constant=50	parm	1·28	1·34	1·43	1·47	1·37
		non-parm	1·09	1·21	1·36	1·42	1·19
	multinom	parm	1·16	1·18	1·22	1·24	1·18
		non-parm	1·04	1·14	1·18	1·22	1·12
	inhomo	parm	0·66	0·65	0·60	0·58	0·66
		non-parm	0·88	0·78	0·66	0·64	0·80

Table E.1: Ratio of MSE of conditional and independent likelihood parametric and non-parametric estimators of G at different percentiles, with $K = 20$ entrance points, $E(N_k) = 20$ or 50, and lifetime distribution G=Gamma with mean 12 and variance 48. Models for the N_k's were (in order of appearance in the table): Poisson, mixture of Poisson variables $\text{mix}(a, b) = 1/2\text{Pois}(a) + 1/2\text{Pois}(b)$, Geometric, Constant, exchangeable multinomial, and inhomogeneous Poisson with $N_k \sim \text{Pois}(\exp(3{\cdot}27 - 0{\cdot}027k))$ and $N_k \sim \text{Pois}(\exp(4{\cdot}12 - 0{\cdot}021k))$.

So far, here and in the paper, we considered the case that the sample comprise all individuals in the cross-sectional population. In the next simulation we study the effect of simple random sampling from a large cross-sectional population. As before, we consider 20 entrance points at times 0,-1,...,-19. The cohort sizes considered are independent negative binomial variables with expectation $\nu = 5000$ and standard deviations varying between $0{\cdot}1\nu$ and $0{\cdot}5\nu$, Poisson(ν) where the standard deviation is $\sqrt{\nu}$, and a degenerate distribution (standard deviation=0). Lifetimes were generated from a Gamma distribution with mean 12 and variance 48. This process generated the cross sectional population of about 50-60 thousand individuals according to the criterion $A \leq X$. From the cross-sectional population at time 0, random samples of $m^* = 400$ and $m^* = 1000$ individuals were selected and the conditional/unconditional parametric/nonparametric estimators were calculated. The MSE ratio of the conditional to the unconditional parametric estimators in the 0·1, 0·5, and 0·9 quantiles are compared in Table E.2 for the various standard deviations. These are based on 1000 replications. The results are similar for non-parametric estimation and for the other simulation studies: the independence likelihood approach is more efficient for cohort sizes that have variance similar to the expectation or smaller, and the conditional approach is more efficient when the variance is much larger than the expectation.

m^*	SD=	0	$\sqrt{\nu}$	$0{\cdot}1\nu$	$0{\cdot}2\nu$	$0{\cdot}3\nu$	$0{\cdot}4\nu$	$0{\cdot}5\nu$
400	$q_{0\cdot10}$	1·20	1·12	1·14	1·05	1·01	0·80	0·63
	$q_{0\cdot50}$	1·33	1·20	1·21	1·10	0·98	0·78	0·55
	$q_{0\cdot90}$	1·29	1·29	1·21	1·15	1·02	0·91	0·69
1000	$q_{0\cdot10}$	1·13	1·17	1·05	0·87	0·66	0·51	0·39
	$q_{0\cdot50}$	1·23	1·26	1·09	0·86	0·61	0·45	0·31
	$q_{0\cdot90}$	1·27	1·23	1·12	0·94	0·74	0·57	0·41

Table E.2: MSE ratio - parametric model. Cohorts sizes, all having expectation $\nu = 5000$, are constant (SD=0), Poisson (SD=$\sqrt{\nu}$) and Negative Binomial with varying standard deviations (SDs). Random sampling of $m^* = 400$ and 1000 individuals from the cross-sectional population

References

Chatterji, S. D. (1963). Some elementary characterizations of the Poisson distribution. *The American Mathematical Monthly*, **70**, 958–964.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Vol 2*. John Wiley & Sons.

Lehmann, E. L. & Casella, G. (1998). *Theory of Point Estimation*, 2nd edition. Springer.

Rényi, A. (1957). On the asymptotic distribution of the sum of a random number of independent random variables. *Acta Mathematica Hungarica*, **8**, 193–199.

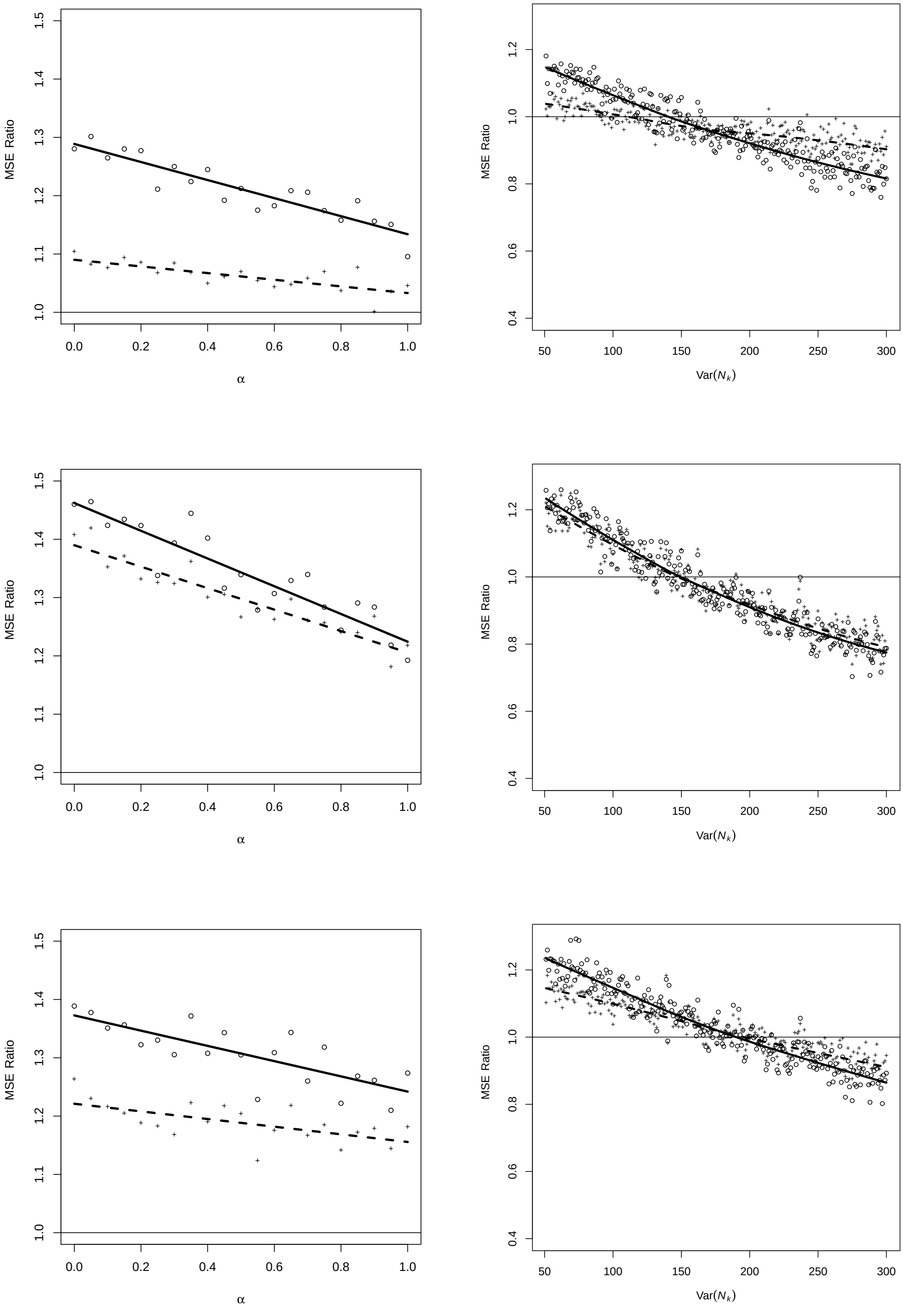

Figure E.1: The effect of variance on the ratio MSE(conditional)/MSE(independent likelihood) at the 0·1, 0·5, and 0·9 quantiles, from top to bottom, calculated from 1000 replications. Entrance process - 20 entrance points, independent cohort sizes with $E(N_k) = 50$. Left - a mixture of a Poisson random variable and a constant: $\alpha\text{Pois}(\nu) + (1-\alpha)\nu$, right - Negative Binomial model with varying variance. Circles and solid curves denote parametric results and a regression fit, and pluses and dashed curves non-parametric results.