Statistica Sinica **27** (2017), 479-499
doi:http://dx.doi.org/10.5705/ss.202015.0209

ESTIMATING COMPONENT RELIABILITY BASED ON FAILURE TIME DATA FROM A SYSTEM OF UNKNOWN DESIGN

Y. Jin[1], Peter Gavin Hall[1,2], Jiming Jiang[1] and Francisco J. Samaniego[1]

[1] *University of California, Davis and* [2] *University of Melbourne*

Abstract: Suppose that N identical systems are tested until failure and that each system is based on n components whose lifetimes are independently and identically distributed with common continuous distribution function $F(t)$ and survival function $\bar{F}(t) = 1 - F(t)$. Under the assumption that the system design is known, Bhattacharya and Samaniego (2010) obtained the nonparametric maximum likelihood estimate of F based on the observed system failure times and characterized its asymptotic behavior. The estimator studied in that paper has the form $\hat{\bar{F}}_0(t) = h^{-1}\left[\hat{\bar{F}}_T(t)\right]$ where $h(\cdot)$ is the system's reliability polynomial (see Barlow and Proshan (1981)) and $\hat{\bar{F}}_T(t)$ is the empirical survival function of the system lifetimes $\{T_1, \ldots, T_N\}$. To treat this estimation problem when the system design is unknown, the design must be estimated from data. In this paper, we assume that auxiliary data in the form of a variable K, the number of failed components at the time of system failure, is available along with the system's lifetime. Such data is typically available from a subsequent autopsy. The problem considered here is motivated by the fact that component reliability under field conditions is often not easily estimated through controlled laboratory tests. The data $(T_1, K_1), (T_2, K_2), \ldots, (T_N, K_N)$ permits the estimation of the reliability polynomial h (through the use of "system signatures" - Samaniego (2007)). Denoting the estimated polynomial as $\hat{h}$, we study the properties of the estimator $\hat{\bar{F}}(t) = \hat{h}^{-1}\left[\hat{\bar{F}}_T(t)\right]$. Our main results include (1) $\hat{\bar{F}}(t)$ is a $\sqrt{n}$-consistent estimator of the component reliability function $\bar{F}(t)$, (2) the asymptotic distribution of $\bar{F}(t)$ is normal and its asymptotic variance is given in closed form, and (3) the asymptotic variance of $\hat{\bar{F}}(t)$, based on the augmented data $\{T_i, K_i\}$, is uniformly no greater than the asymptotic variance of $\hat{\bar{F}}_0(t)$, based on the data $\{T_i\}$ and the assumption that h is known. This latter, perhaps surprising, result is confirmed in a variety of simulations and is illuminated through further heuristic considerations and further analysis.

Key words and phrases: Asymptotic efficiency, asymptotic normality, coherent system, component and system reliability, consistency, nonparametric estimation, NPMLE, nuisance parameter, system signature.

1. Introduction

1.1. Background

In what follows, we make the tacit assumption that the engineered system under study is a coherent system, that is, the system's performance is a monotone function of each component's performance and each component is relevant (see Barlow and Proshan (1981)). Suppose that N identical systems are tested until failure and that each system is based on n components whose lifetimes are independently and identically distributed with common continuous distribution F. Under the standard and generally defensible assumption that the system design is known, Bhattacharya and Samaniego (2010) obtained the nonparametric maximum likelihood estimate of F based on the observed system failure times, and they characterized its asymptotic behavior. Studies on estimating component characteristics from system failure time data are motivated by the following considerations. The laboratory environment in which components are tested often differs in significant ways from the environment in which the components in fielded systems operate. Further, the direct estimation of component characteristics based on data from fielded systems requires extensive autopsy data on the components that is often infeasible or otherwise unavailable. It is often feasible to obtain lifetime data on fielded systems, while estimating the field performance of the system's components is difficult or impossible under laboratory conditions.

The assumption that the lifetimes of the components of the system of interest are independent and identically distributed deserves further comment. Samaniego (2007) discusses a variety of systems to which this assumption applies. A further application, somewhat larger in scale, is the case of computer hardware that is generally referred to as "Redundant Array of Independent Disks (RAID)" — see Patterson, Gibson and Katz (1988). RAID computers are well known for their efficiency and speed, two characteristics that have led to their widespread use. The performance of RAID computers with n independent disks can be designed to perform as a k-out-of-n system that fails upon kth disk failure. With the aid of a randomization device, RAID computers can be used to simulate the performance of an arbitrary coherent system in n components with i.i.d. lifetimes.

Let $T_1, \ldots, T_N \overset{i.i.d.}{\sim} F_T$ be a random sample of N failure times obtained from a system with lifetime distribution F_T. The NPMLE $\hat{\bar{F}}$ of the component reliability is the inverse of the reliability polynomial h applied to the empirical reliability function $\hat{\bar{F}}_T$ of the system's lifetime T. This follows from the well-known link between the reliability function $\bar{F}_T$ of the system's failure time T and that of the component reliability function $\bar{F}$: $\bar{F}_T(t) = h(\bar{F}(t))$. The coefficients of h depend solely on the distribution-free signature vector $\mathbf{s}$ of the system -

see Samaniego (2007). If $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$ are the lifetimes of the components in the system, then the signature vector of the system is the probability vector $\mathbf{s}$ whose ith element is $s_i = P(T = X_{i:n})$, where T is the system's lifetime and $X_{i:n}$ is the ith smallest component failure time in the sample $X_1, \ldots, X_n$. Setting $p = \bar{F}(t)$, Samaniego (1985) proved that

$$h(p) = \sum_{i=1}^{n} s_i \Big[\sum_{j=n-i+1}^{n} \binom{n}{j} p^j (1-p)^{n-j} \Big] \equiv \sum_{i=1}^{n} s_i h_i(p)\,, \tag{1.1}$$

where s_i is the ith element of the signature vector, and h_i is the reliability polynomial associated with the ith ordered component failure time $X_{i:n}$.

When the system design is unknown, it must be estimated from the data. In this paper, we study the estimation of the component reliability function $\bar{F}$ under the condition that auxiliary data is available in the form of the number of failed components in each of the fielded systems. We take advantage of this additional information to estimate the unknown system structure.

Though most engineered systems that are fielded by customers or clients have a specific known design, there are exceptions where the identification of a system structure is not possible. For example, in military operations, it is not uncommon to capture or gain control of a collection of like systems whose precise design is unknown. In addition, the system reliability polynomial h is not always easily obtained, even when the structure is known, because of computational complexity issues. The challenge is to obtain an accurate estimate of the signature vector from system failure-time data, thereby permitting an estimation process of F based on the estimated polynomial h.

A number of authors have studied the estimation of component lifetime distributions from system failure times in the presence of additional information. Notable examples include Moeschberger and David (1971), who treated the estimation problem in a competing risks framework, and Meilijson (1981) and Bueno (1988), both of whom considered the estimation of F based on system failure times together with autopsy statistics on the systems components. Authors who have studied the estimation of component characteristics from masked data include Miyakawa (1984), Usher and Hodgson (1988), and Guess, Usher and Hodgson (1991). Estimation of the component lifetime distribution from system failure time data also arises in other contexts. Boyles and Samaniego (1986) derived the NPMLE of the underlying component distribution F based on nomination sampling, a sampling method that yields data equivalent to the observed lifetimes of parallel systems (see also Boyles and Samaniego (1987)). Inference about the underlying distribution F based on ranked set sampling has been treated by Stokes and Sager (1988) and by Kvam and Samaniego (1993a,b, 1994). A ranked set sample may be viewed as a set of independent lifetimes from k-out-of-n systems with varying k and n. Much of the work cited above makes parametric assumptions; none treats the case of systems of arbitrary design.

The present paper takes a fully nonparametric approach to estimation, treats the problem of estimating component reliability function $\bar{F}$ in a *system of arbitrary design*, and treats (to our knowledge, for the first time) the estimation of $\bar{F}$ in the important special case when the system is assumed to be of *unknown design*. Assuming the availability of some auxiliary data, a consistent, asymptotically normal estimator of $\bar{F}$ is obtained. Quite unexpectedly, this estimator is shown to be superior, asymptotically, to the NPMLE of $\bar{F}$ computed under the assumption that the system's design (that is, its signature vector) is known. An alternative trajectory of research generalizing Bhattacharya and Samaniego (2010) is presented by Hall, Jin and Samaniego (2015), where the nonparametric estimation of component reliability is undertaken under the assumption that failure-time data is available from multiple systems with known designs.

1.2. Outline of the methodology

Our approach to estimating F is to apply the inverse of an estimated reliability polynomial $\hat{h}$ to $\hat{\bar{F}}_T$, imitating the form of the NPMLE $\hat{\bar{F}} = h^{-1} \circ \hat{\bar{F}}_T$ discussed above. Additional information is required to make this estimation possible. Assume that the index of the ordered component failure time of the component that caused the system to fail has been determined. This information is typically available in a subsequent autopsy, as it can be obtained simply by counting the number of failed components in the failed system. The identification of the precise component that caused the system to fail is not required in the analysis below. The count data alluded to above, together with the failure time data from the sample of fielded systems, suffice for obtaining viable estimators of the underlying component lifetime distribution F.

Suppose we have failure time data on a random sample of N fielded systems. We shall assume that each fielded system yields a random pair (T, K), where T is the failure time of the system and K is the number of failed components at the time of system failure. The data $(T_1, K_1), (T_2, K_2), \ldots, (T_N, K_N)$ permits estimation of the signature vector $\mathbf{s}$. Let $X_1, \ldots, X_n$ be the theoretical independently and identically distributed component failure times from the system of interest. Recalling that $s_i = P(T = X_{i:n})$, the observations $K_1, \ldots, K_N$ can be viewed as a random sample from the multinomial distribution $M_n(N, \mathbf{s})$. It follows that $\hat{s}_{N,i}$, defined as the sample proportion of K's that are equal to i, that is:

$$\hat{s}_{N,i} = \frac{\#\{K_j = i, j = 1, \ldots, N\}}{N}, \tag{1.2}$$

is an unbiased estimator of s_i that converges to s_i almost surely as N goes to infinity. We estimate $\mathbf{s}$ by $\hat{\mathbf{s}}_N$, the estimated signature vector based on $K_1, \ldots, K_N$. Since the coefficients in the reliability polynomial h are determined by $\mathbf{s}$, we denote the estimated reliability polynomial based on $\hat{\mathbf{s}}_N$ as $h_{\hat{\mathbf{s}}_N}$, defined by plugging in $\hat{s}_{N,i}$ for s_i in (1.1):

$$h_{\hat{\mathbf{s}}_N}(p) = \sum_{i=1}^{n} \hat{s}_{N,i} h_i(p) \,. \tag{1.3}$$

We denote the underlying reliability polynomial $h(p)$ by $h_{\mathbf{s}}(p)$ hereafter to differentiate it from $h_{\hat{\mathbf{s}}_N}$. This substitution allows for the development of an estimator of F by inverting the approximate relationship $\bar{F}_T(t) \approx h_{\hat{\mathbf{s}}_N}(\bar{F}(t))$ as $\hat{\bar{F}} = {h_{\hat{\mathbf{s}}_N}}^{-1} \circ \hat{\bar{F}}_T$, or

$$\hat{\bar{F}}(t) = {h_{\hat{\mathbf{s}}_N}}^{-1} \left[\hat{\bar{F}}_T(t) \right] \,. \tag{1.4}$$

In section 2, it is shown that the estimator $\hat{\bar{F}}$ in (1.4) of the reliability function $\bar{F}$ of component lifetimes is a consistent estimator of the true $\bar{F}$. Further, the asymptotic normal distribution of the estimator is established, and its asymptotic variance is obtained in a form that permits comparisons with alternative estimators. This derivation is followed by the somewhat surprising result that $\hat{\bar{F}} = {h_{\hat{\mathbf{s}}_N}}^{-1} \circ \hat{\bar{F}}_T$ is as good or better than $\hat{\bar{F}} = {h_{\mathbf{s}}}^{-1} \circ \hat{\bar{F}}_T$, as measured by asymptotic variance. In Section 3, simulation results are presented that confirm that our asymptotic results tend to hold even for small to moderate sample sizes.

2. Estimation Using the Inversion Method

2.1. Main theoretical results

We first establish the consistency of the proposed estimator of $\bar{F}$. The proofs of the results stated in this section have been relegated to the Appendix.

Theorem 1. *Let $(T_1, K_1), (T_2, K_2), \ldots, (T_N, K_N)$ be a random sample of system lifetimes T augmented by the count K of the number of failed components when the system fails. Then*

$$\hat{\bar{F}}(t) = {h_{\hat{\mathbf{s}}_N}}^{-1} \left[\hat{\bar{F}}_T(t) \right]$$

is a consistent estimator of $\bar{F}(t)$ for all $t > 0$, where $\hat{\bar{F}}_T$ is the empirical distribution function of the system failure times and $h_{\hat{\mathbf{s}}_N}(p)$ is defined as in (1.3).

Now we turn to the asymptotic normality of the proposed estimator. We use the following notation: for a fixed $t > 0$,

$$\pi_0 = \bar{F}_T(t), \tag{2.1a}$$

$$\hat{\pi}_N = \hat{\bar{F}}_T(t), \tag{2.1b}$$

$$p_0 = \bar{F}(t) = h_{\mathbf{s}}{}^{-1}(\pi_0), \tag{2.1c}$$

$$\hat{p}_N = \hat{\bar{F}}(t) = h_{\hat{\mathbf{s}}_N}{}^{-1}(\hat{\pi}_N), \tag{2.1d}$$

$$\hat{p}_N^{(0)} = h_{\hat{\mathbf{s}}_N}{}^{-1}(\pi_0). \tag{2.1e}$$

Here $\hat{\pi}_N$ is the sample proportion of system survival times that exceed time t; thus, $N\hat{\pi}_N$ follows the $\mathcal{B}(N, \pi_0)$ distribution. Our first result, proven in the Appendix, establishes the joint asymptotic normality of $\hat{\pi}_N$ and the vector $(\hat{s}_{N,1}, \ldots, \hat{s}_{N,n})$.

Lemma 1. *Let $\hat{\pi}_N$ be the sample proportion of system survival beyond time t, and let $\hat{\mathbf{s}}_N$ be defined as in* (1.2). *Then*

$$\begin{pmatrix} \sqrt{N}\,(\hat{\pi}_N - \pi_0) \\ \sqrt{N}(\hat{\mathbf{s}}_N - s) \end{pmatrix} \overset{\mathcal{D}}{\to} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \Sigma), \tag{2.2}$$

where $\overset{\mathcal{D}}{\to}$ denotes convergence in distribution. The covariance matrix can further be partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where

$$\Sigma_{11} = \pi_0(1 - \pi_0),$$

$$\Sigma_{22} = \begin{pmatrix} s_1(1-s_1) & -s_1 s_2 & -s_1 s_3 & \ldots & -s_1 s_n \\ -s_1 s_2 & s_2(1-s_2) & -s_2 s_3 & \ldots & -s_2 s_n \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -s_1 s_n & -s_2 s_n & -s_3 s_n & \ldots & s_n(1-s_n) \end{pmatrix},$$

$$\Sigma_{12} = \Sigma_{21}^T = \Big(s_1\left[h_1\left(p_0\right) - \pi_0\right], s_2\left[h_2\left(p_0\right) - \pi_0\right], \ldots, s_n\left[h_n\left(p_0\right) - \pi_0\right] \Big).$$

Remark. Note that Σ_{22} is not of full rank; in fact, if we denote the upper-left sub-matrix of Σ_{22} of dimension $k \times k$ as

$$M_k = \begin{pmatrix} s_1(1-s_1) & -s_1 s_2 & -s_1 s_3 & \ldots & -s_1 s_k \\ -s_1 s_2 & s_2(1-s_2) & -s_2 s_3 & \ldots & -s_2 s_k \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -s_1 s_k & -s_2 s_k & -s_3 s_k & \ldots & s_k(1-s_k) \end{pmatrix},$$

we can use the formula $\det(A - uu^T) = \det(A)(1 - u^T A^{-1} u)$ and show that:

$$\det(M_k) = (\prod_{i=1}^{k} s_i)(1 - \sum_{i=1}^{k} s_i) = 0\,, \quad \text{if} \quad \sum_{i=1}^{k} s_i = 1\,.$$

This shows that the rank of Σ_{22} is $r - 1 < n$, where s_r is last non-zero element in $\mathbf{s}$. So the matrix Σ is also necessarily singular, with rank r.

The next result establishes the asymptotic distribution of the estimator $\hat{\bar{F}}(t)$; the proof is given in the Appendix.

Theorem 2. *For fixed* $t > 0$, $\sqrt{N}\left(\hat{\bar{F}}(t) - \bar{F}(t)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2\right)$, *where*

$$\sigma^2 = \left\{h'_{\mathbf{s}}\left[\bar{F}(t)\right]\right\}^{-2} \sum_i s_i h_i\left[\bar{F}(t)\right]\left\{1 - h_i\left[\bar{F}(t)\right]\right\}. \tag{2.3}$$

Upon rewriting h as $h_{\mathbf{s}}$, the NPMLE of the component reliability $\hat{\bar{F}}$ can be written as

$$\hat{\bar{F}}_0(t) = h_{\mathbf{s}}^{-1}\left[\hat{\bar{F}}_T(t)\right]. \tag{2.4}$$

The final result of this section establishes the asymptotic domination of the estimator $\hat{\bar{F}}$ of $\bar{F}$ with estimated signature over the estimator $\hat{\bar{F}}_0$ of $\bar{F}$ when the signature vector of the system is known with certainty. The proof is given in the Appendix; an elementary example of the phenomenon in question is provided as an illustration in Section 2.3.

Theorem 3. *The asymptotic variance of the estimator* $\hat{\bar{F}}(t)$ *is less than or equal to that of* $\hat{\bar{F}}_0(t) = h_{\mathbf{s}}^{-1}\left[\hat{\bar{F}}_T(t)\right]$*, where* $\hat{\bar{F}}_T(t)$ *is the NPMLE of* $\bar{F}_T(t)$ *based on a sample of system failure times when the system signature* $\mathbf{s}$ *is known.*

2.2. Discussion

Remark/Heuristic consideration: If a nuisance parameter is known, one would think that the use of an ancillary statistic to estimate it could not provide an improved estimator of the target parameter. In our context, $\bar{F}$ is the target parameter, while the signature vector $\mathbf{s}$ is the nuisance parameter; system lifetimes are sufficient for estimating $\bar{F}$ when $\mathbf{s}$ is known, and K is an ancillary statistic whose distribution depends only on $\mathbf{s}$. From the proof of the theorem it can be seen that the smaller asymptotic variance comes from the positive correlation between T and K. Actually, if $\mathbf{s}$ is estimated independently, for example if we observe K in a separate experiment, the covariance matrix in (2.2) is a diagonal matrix. Then V_{12} is 0, and the asymptotic variance of the estimator is $V_1 + V_2$, clearly larger than V_1, the asymptotic variance of $\hat{\bar{F}}$. The examples in Sections 2.3 and A.2 shed further light on this phenomenon.

2.3. Example

For $i = 1, \ldots, N$, assume (X_i, Y_i) are independent, with $X_1, \ldots, X_N \overset{i.i.d.}{\sim} \mathcal{B}(1, \theta)$ and for each i, $Y_i | X_i = x \sim \mathcal{B}(x, p)$ where $\mathcal{B}(0, p) \equiv \delta_0$ is a degenerate distribution with mass 1 at the value 0. Marginally, $Y_1, \ldots, Y_N \overset{i.i.d.}{\sim} \theta\mathcal{B}(1, p) + (1-\theta)\delta_0$. In particular, $P(Y_i = 1) = \theta p$ and $P(Y_i = 0) = 1 - \theta p$.

If only $Y_1, Y_2, \ldots, Y_N$ are observed and the parameter θ is known, then the MLE of p is $\hat{p}_1 = \overline{Y}/\theta$. Since $Y_1, Y_2, \ldots, Y_N \overset{i.i.d.}{\sim} F$ with mean θp and variance $\theta p(1-\theta p)$, it follows that $\sqrt{N}(\overline{Y} - \theta p) \overset{\mathcal{D}}{\to} U \sim \mathcal{N}\big(0, \theta p(1-\theta p)\big)$, and hence that

$$\sqrt{N}(\hat{p}_1 - p) \overset{\mathcal{D}}{\to} U^* \sim \mathcal{N}\left(0, V_1\right), \tag{2.5}$$

where $V_1 = p(1-\theta p)/\theta$.

Now assume we have auxiliary data $X_1, X_2, \ldots, X_n$. Since we know that $\overline{X}$ is a good estimator of θ, we might be inclined to construct a plug-in estimator $\overline{Y}/\overline{X}$ of p by replacing θ by $\overline{X}$.

Since $\overline{X}$ can take the value 0 with positive probability, we take

$$\hat{p}_2 = \begin{cases} \frac{\overline{Y}}{\overline{X}} & \text{if } \overline{X} > 0, \\ 0 & \text{if } X_1, \ldots, X_N = 0, \end{cases}$$

to obtain a well-defined estimator for p.

Since $\mathrm{E}X = \theta$, $\mathrm{Var}(X) = \theta(1-\theta)$, $\mathrm{E}Y = \theta p$, $\mathrm{Var}(Y) = \theta p(1-\theta p)$ and $\mathrm{Cov}(X, Y) = \theta p(1-\theta)$, we have, by the CLT, that as $N \to \infty$,

$$\sqrt{N}\left[\begin{pmatrix}\overline{X}\\ \overline{Y}\end{pmatrix} - \begin{pmatrix}\theta\\ \theta p\end{pmatrix}\right] \overset{D}{\to} V \sim \mathcal{N}\left[\begin{pmatrix}0\\ 0\end{pmatrix}, \begin{pmatrix}\theta(1-\theta) & \theta p(1-\theta)\\ \theta p(1-\theta) & \theta p(1-\theta p)\end{pmatrix}\right].$$

If $g(x, y) = y/x$, the bivariate delta method theorem implies that, as $N \to \infty$,

$$\sqrt{N}\left[\frac{\overline{Y}}{\overline{X}} - p\right] \overset{D}{\to} V \sim \mathcal{N}\left[0, V_2\right],$$

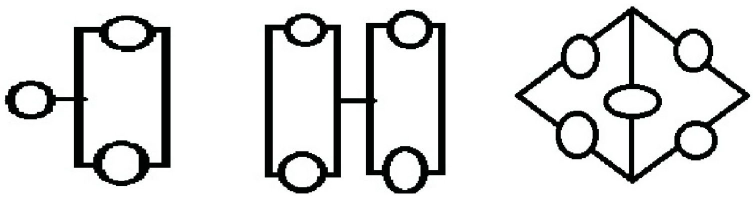
where

$$\begin{aligned} V_2 &= \left(-\frac{p}{\theta}, \frac{1}{\theta}\right)\begin{pmatrix}\theta(1-\theta) & \theta p(1-\theta)\\ \theta p(1-\theta) & \theta p(1-\theta p)\end{pmatrix}\begin{pmatrix}-\frac{p}{\theta}\\ -\frac{1}{\theta}\end{pmatrix} \\ &= \frac{p^2(1-\theta)}{\theta} - \frac{2p \cdot \theta p(1-\theta)}{\theta^2} + \frac{\theta p(1-\theta p)}{\theta^2} \\ &= \frac{p(1-p)}{\theta}. \end{aligned}$$

Since $|\hat{p}_2 - \overline{Y}/\overline{X}| \overset{p}{\to} 0$, if follows that

$$\sqrt{N}\left(\hat{p}_2 - p\right) \overset{\mathcal{D}}{\to} V \sim \mathcal{N}(0, V_2). \tag{2.6}$$

It is clear that $V_2 = p(1-p)/\theta < p(1-\theta p)/\theta = V_1$.



(i) (ii) (iii)

Figure 1. Structure of three systems.

This shows that, when the parameter θ is known, the estimator $\hat{p}_2$ which only utilizes the auxiliary data $X_1, \ldots, X_N$ (and ignores the known value of θ) asymptotically outperforms the estimator $\hat{p}_1$ which utilizes the known value of θ.

3. Numerical Results

Here, we simulate the behavior of $\hat{\hat{\bar{F}}}(t)$ and $\hat{\hat{\bar{F}}}_0(t)$. We are going to assume that the components' lifetimes follow (i) Exp(1), (ii) Weibull$(1, 0.5)$ and (iii) Weibull$(1, 5)$ distributions, which enable us to consider models with constant failure rate, decreasing failure rate (DFR), and increasing failure rate (IFR). Three coherent systems are studied, where the number of components is 3, 4 and 5 respectively. The first two are series-parallel systems, and the third is a bridge system; their structures are shown in Figure 1. The signature vectors for the first two coherent systems are $(1/3, 2/3, 0)$ and $(0, 1/3, 2/3, 0)$, while that of the five-component bridge system is $(0, 1/5, 3/5, 1/5, 0)$.

The asymptotic domination of the estimator $\hat{\hat{\bar{F}}}$ over the estimator $\hat{\hat{\bar{F}}}_0$ is proven in Theorem 3 when the signature vector of the system is estimated rather than known with certainty. In order to reflect asymptotic behavior, the number of samples, N was taken to be 100. We first obtained both estimators, $\hat{\hat{\bar{F}}}(t)$ and $\hat{\hat{\bar{F}}}_0(t)$, for t between the 1 and 99 percentiles of the underlying component lifetime distribution in 1,000 replicated simulation runs. Then we were able to calculate the standard deviations of the 1,000 replications of $\hat{\hat{\bar{F}}}(t)$ and $\hat{\hat{\bar{F}}}_0(t)$, respectively, and plot them against $\bar{F}(t)$. Figure 2 shows the simulation results for system (i), where the solid line represents the standard deviation of $\hat{\hat{\bar{F}}}(t)$ and the dash line is for $\hat{\hat{\bar{F}}}_0(t)$.

As shown in Figure 2, the estimators are eventually identical at the ends of their domains, and have maximum standard deviation when $\bar{F}(t)$ is small. As the underlying distribution of components lifetime varies, the difference between the two estimators changes. For the three models examined above, and for the moderate sample size $N = 100$, the inequality $V\left(\hat{\hat{\bar{F}}}(t)\right) \leq V\left(\hat{\hat{\bar{F}}}_0(t)\right)$ is seen to hold for all nonnegative t.

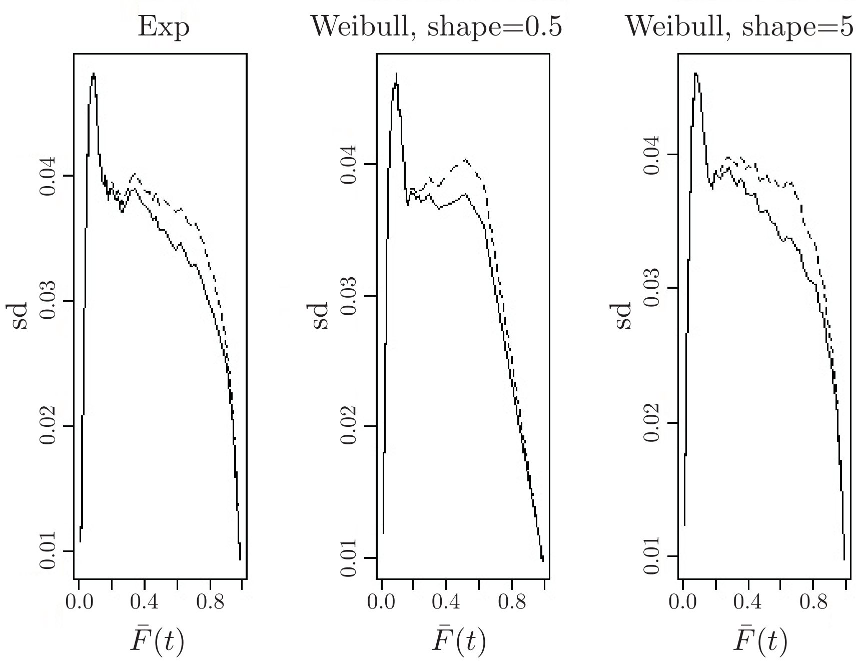
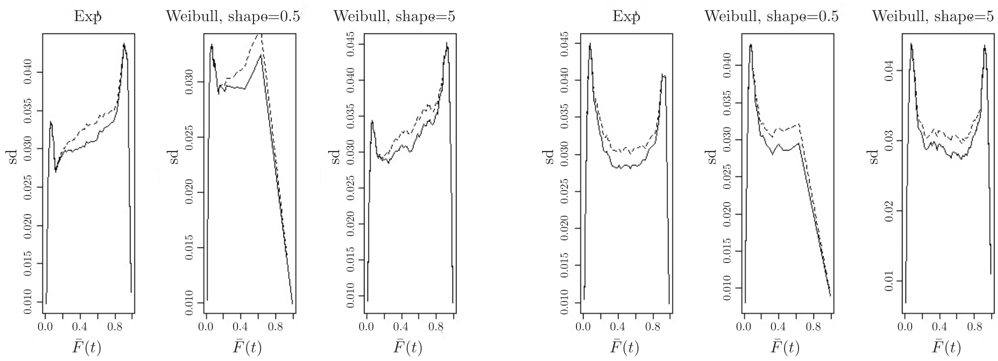


Figure 2. Three component system.

The same study was also undertaken for systems (ii) and (iii). Figure 3a shows results for system (iii), and Figure 3b shows those for system (iii). The pattern of standard deviations is more complicated than that in Figure 2. There are two peaks in the variation: one for small $\bar{F}(t)$ and another for large $\bar{F}(t)$, while the DFR model is different from the other two models. Regardless of the shape, we can still see that the dashed line lies above the solid line so that, the domination also exists here. A similar figure is drawn for the bridge system; see Figure 3b.

The comparison of the two estimators $\hat{\bar{F}}$ and $\hat{\bar{F}}_0$ is of interest as well. Here we turn our attention to global performance measures for the proposed estimators. We calculated, from 1,000 replicated simulation runs, the mean Integrated Squared Error (ISE) of each of the estimators. We do not show the estimator of $\bar{F}$ when t is beyond the 1 and 99 percentiles of the underlying component lifetime distribution, where it is either 0 or 1. Table 1 shows the mean of ISEs over 1,000 replications of the estimators when the sample size N goes from 10 to 50, incremented by 10, when the bridge system has structure (iii). It can be seen that ISE_1 is always smaller than ISE_0, showing that $\hat{\bar{F}}$ has a smaller ISE then $\hat{\bar{F}}_0$ even for small to moderate sample size.



(a) four component system (b) bridge system

Figure 3. Four and five component systems.

	Weibull(1,0.5)		Exp(1)		Weibull(1,5)	
N	ISE$_0$	ISE$_1$	ISE$_0$	ISE$_1$	ISE$_0$	ISE$_1$
10	0.5611	0.5395	0.8548	0.8162	0.9929	0.9434
20	0.3473	0.3390	0.4626	0.4390	0.5220	0.4988
30	0.2628	0.2520	0.3349	0.3211	0.3538	0.3356
40	0.2059	0.1988	0.2549	0.2417	0.2795	0.2632
50	0.1781	0.1744	0.2160	0.2033	0.2227	0.2099

Table 1. Comparison of behavior in a bridge system.

4. Discussion and Summary

In this paper, we have considered an engineering problem which arises quite often in practice. The field performance of the components of an engineered system is not easily estimated through independent experiments of the components themselves. The primary difficulty lies in the construction of a testing environment which effectively simulates the environment in which they will operate when used in the field. This leads us to focus on the inverse problem of inferring component performance from the observed performance of fielded systems. This problem is examined under the assumption that the systems being tested are identical copies of a coherent system and that the components of each system have i.i.d. lifetimes with common reliability function $\bar{F}(t)$ for $t > 0$.

In most applications, the design of the system being tested is known, as the systems themselves would likely have been constructed or acquired (by purchase or commission) by the agency doing the testing. Since the performance of a coherent system in i.i.d. components is fully characterized by its *signature vector* $\mathbf{s}$, we have taken the assumption that the system design is known to mean that the system has a known signature. Under that assumption, Bhattacharya and Samaniego (2010) obtained the nonparametric maximum likelihood estimator of the component reliability $\bar{F}(t)$ from system failure time data and characterized its asymptotic behavior.

In the present work, we have assumed that, in addition to system lifetime data, one is able to observe the number K of failed components in each system that has failed. This additional data, which is typically obtainable through a straightforward autopsy following system failure, provides information from which the systems signature can be consistently and efficiently estimated.

An issue of practical interest is the identification of instances in which one must deal with systems whose design is unknown. The circumstance of interest can occur, as well, in ordinary business or industrial settings in which the precise design of a system developed by a competing firm is unknown. In such cases, the estimation process we have described can serve two important functions in that it provides insight into the unknown design through the system's estimated signature and can also be the basis for inference concerning the performance of the systems components.

The asymptotic behavior of our estimator is given in Theorem 2. In practical applications, it is important to try to determine how large a sample size is required to ensure that the normal distribution of the estimator $\hat{\bar{F}}(t)$ described in Theorem 2 is a reliable approximation. Simulations suggest that, for systems of small to moderate size, $N = 100$ is a sample size for which the result in Theorem 2 can be safely applied.

The estimator of the component reliability $\bar{F}(t)$ developed in this paper is shown to have the unexpected property that it has better asymptotic statistical performance than the estimator one would use (i.e., the NPMLE) if the system design were in fact known. It is tempting to try to explain this dominance as being due to the fact that when the system signature is assumed known, ones estimator depends only on the lifetime data $T_1, \ldots, T_N$, while when the system signature is unknown, the proposed estimator depends on, indeed requires, the bivariate data (T_i, K_i), $i = 1, \ldots, N$. Thus, when the system design is unknown, more data is utilized in estimating $\bar{F}(t)$ and it should not be surprising that a stronger estimator is possible in this latter scenario. But consider the following mind game.

Suppose that there were parallel universes in which experiments were run on separate collections of N identical systems, where the system design was known in the first instance and was unknown in the second. Suppose further that the bivariate data $(T_i, K_i), i = 1, \ldots, N$ were available from both experiments. Now, since the distribution of the discrete variable K depends only on the signature vector **s**, it could justifiably be dismissed as irrelevant in the problem of estimating $\bar{F}(t)$ when **s** is known. But our work shows that it is advantageous (in sufficiently large samples) to ignore the "known" signature s in this scenario and use the estimator developed in the present paper that utilizes the full bivariate data. A proof is provided in Section 3, and is augmented for heuristic reasons, by two simpler examples in which the "surprising" phenomenon is shown to occur. As stated in Section 2, the key to understanding the phenomenon in the reliability problem on which we have focused is the recognition of the fact that T and K are positively correlated and that that correlation can be exploited with profit.

An anonymous referee suggested that the "domination phenomenon" that arises in the problem we have studied here is reminiscent of "a similar result in (the area of) sample survey in estimating the population mean, where estimation of the known population size from the survey data results in an estimator of the population mean with a smaller MSE than the usual estimator which uses the known population size N". We thank the referee for calling our attention to an additional example of this interesting phenomenon.

Dedication and Acknowledgements

Peter Gavin Hall, our dear friend and collaborator, passed away on January 9, 2016, after having bravely fought a series of kidney-related illnesses and, more recently, being besieged by Leukemia. Peter was an inspiration to us and to many students, teachers and researchers in the discipline of Statistics. He was a pleasure to work with, both because his creativity and brilliant technical skill contributed greatly to the scope and depth of the research projects in which he was engaged and because his innate congeniality helped to create an environment in which professional interactions were enjoyable as well as productive. He will be greatly missed, as much for the fine person who he was as for the strong impact that he has had on our profession. We offer our sincere condolences to his loving and devoted wife Jeannie. The present work is dedicated to Peter's memory.

P. G. Hall's research was supported in part by grant 3-SNSFPH3 from the U. S. National Science Foundation and also by an Australian Laureate Fellowship from the Australian Research Council. F. J. Samaniego's and Y. Jin's research was supported in part by grant W911NF-1-11-0428 from the U. S. Army Research Office. "J. Jiang's research was supported in part by NIH grant R01-GM085205A1 and NSF grant DMS-1510219.

Appendix

A.1. Proofs of the theorems and lemmas

Proof of Theorem 1. Since $\hat{s}_{N,i}$ is the sample proportion from a random sample of systems of a fixed design, $\hat{\mathbf{s}}_N$ is clearly a consistent estimator of $\mathbf{s}$. Since the reliability polynomial h of a coherent system is continuous and strictly increasing, $h_{\hat{\mathbf{s}}_N}(x)$ converges to $h_{\mathbf{s}}(x)$ almost surely for any fixed $x \in (0,1)$. Using $\overset{a.s.}{\to}$ to denote almost sure convergence, then

$$h_{\hat{\mathbf{s}}_N}{}^{-1}(x) \overset{a.s.}{\to} h_{\mathbf{s}}^{-1}(x).$$

Moreover, the almost sure convergence of $\hat{\bar{F}}_T(t)$ to $\bar{F}_T(t)$ for any fixed t, along with the continuity of $h_{\hat{\mathbf{s}}_N}$ leads to

$$h_{\hat{\mathbf{s}}_N}{}^{-1}\left[\hat{\bar{F}}_T(t)\right] - h_{\hat{\mathbf{s}}_N}{}^{-1}\left[\bar{F}_T(t)\right] \overset{a.s.}{\to} 0.$$

So

$$h_{\hat{\mathbf{s}}_N}{}^{-1}\left[\hat{\bar{F}}_T(t)\right] \overset{a.s.}{\to} h_{\mathbf{s}}{}^{-1}\left[\bar{F}_T(t)\right] = \bar{F}(t)\,.$$

Proof of Lemma 1. The convergence result follows directly from the Central Limit Theorem. The work that remains is to identify the elements of the covariance matrix Σ. The diagonal terms Σ_{11} and Σ_{22} are derived from the fact that $N \times \hat{\pi}_N \sim \mathcal{B}(N, \pi_0)$ and $N \times \hat{\mathbf{s}}_N \sim M_n(N, \mathbf{s})$. The kth element in the asymptotic covariance matrix is obtained as follows:

$$\begin{aligned}
&\mathrm{Cov}(\hat{\pi}_N - \pi_0, s_{N,k} - s_k)\\
&\quad = E(\hat{\pi}_N s_{N,k}) - E(\hat{\pi}_N)E(s_{N,k})\\
&\quad = N^{-2}E\left(\#\{T>t\}\#\{K=k\}\right) - \pi_0 s_k\\
&\quad = N^{-2}E\Big[\sum_i\sum_j I(T_i>t)I(K_j=k)\Big] - \pi_0 s_k\\
&\quad = N^{-2}\sum_i E\left[I(T_i>t, K_i=k)\right]\\
&\qquad +N^{-2}\sum_{i\neq j} E\left[I(T_i>t)\right]E\left[I(K_j=k)\right] - \pi_0 s_k \qquad\qquad (A.1)\\
&\quad = N^{-1}P(T>t, K=k) + \frac{N-1}{N}\pi_0 s_k - \pi_0 s_k\\
&\quad = N^{-1}P(T>t, K=k) - N^{-1}\pi_0 s_k\\
&\quad = N^{-1}P(K=k)P(T>t|K=k) - N^{-1}\pi_0 s_k = N^{-1}s_k\bar{F}_{(k)}(t) - N^{-1}\pi_0 s_k\\
&\quad = N^{-1}s_k\left[\bar{F}_{(k)}(t) - \pi_0\right] = N^{-1}s_k\left\{h_k\left[\bar{F}(t)\right] - \pi_0\right\}\\
&\quad = N^{-1}s_k\left[h_k\left(p_0\right) - \pi_0\right], \qquad\qquad (A.2)
\end{aligned}$$

where (A.1) follows from the fact that the ith observation is independent of the jth if $i \neq j$. Therefore $\left(\sqrt{N}\,(\hat{\pi}_N - \pi_0)\,, \sqrt{N}(\hat{\mathbf{s}}_N - \mathbf{s})\right)^T$ converges to the multivariate Normal distribution as stated in (2.2).

Proof of Theorem 2. The difference $\hat{\bar{F}}(t) - \bar{F}(t) = {h_{\hat{\mathbf{s}}_N}}^{-1}(\hat{\pi}_N) - {h_{\mathbf{s}}}^{-1}(\pi_0)$ can be decomposed into the sum of two terms by adding and subtracting the term ${h_{\hat{\mathbf{s}}_N}}^{-1}(\hat{\pi}_0)$. This decomposition leads to the representation

$$\begin{aligned}\sqrt{N}\left[\hat{\bar{F}}(t) - \bar{F}(t)\right] &= \sqrt{N}\left[{h_{\hat{\mathbf{s}}_N}}^{-1}(\hat{\pi}_N) - {h_{\mathbf{s}}}^{-1}(\pi_0)\right] \\ &= \sqrt{N}\left[{h_{\hat{\mathbf{s}}_N}}^{-1}(\hat{\pi}_N) - {h_{\hat{\mathbf{s}}_N}}^{-1}(\hat{\pi}_0)\right] + \sqrt{N}\left[{h_{\hat{\mathbf{s}}_N}}^{-1}(\hat{\pi}_0) - {h_{\mathbf{s}}}^{-1}(\hat{\pi}_0)\right] \\ &\equiv \sqrt{N}\left(\hat{p}_N - p_N^{(0)}\right) + \sqrt{N}\left(p_N^{(0)} - p_0\right). \end{aligned} \tag{A.3}$$

We now prove that the two terms in this decomposition can be approximated by two normally distributed random variables. The variance and covariance of these variables will be specified. First, the Taylor series expansion of ${h_{\hat{\mathbf{s}}_N}}^{-1}$ at π_0 maybe written as,

$$\sqrt{N}\left(\hat{p}_N - \hat{p}_N^{(0)}\right) \equiv \sqrt{N}\left[{h_{\hat{\mathbf{s}}_N}}^{-1}(\hat{\pi}_N) - {h_{\hat{\mathbf{s}}_N}}^{-1}(\pi_0)\right] \tag{A.4}$$

$$= \sqrt{N}\left\{\left[\frac{d}{d\pi}{h_{\hat{\mathbf{s}}_N}}^{-1}(\pi)\Big|_{\pi=\pi_0}\right](\hat{\pi}_N - \pi_0) + o(\hat{\pi}_N - \pi_0)\right\}$$

$$\Rightarrow\ \sqrt{N}\left(\hat{p}_N - \hat{p}_N^{(0)}\right) = \left[\frac{d}{d\pi}{h_{\hat{\mathbf{s}}_N}}^{-1}(\pi)\Big|_{\pi=\pi_0}\right]\sqrt{N}(\hat{\pi}_N - \pi_0) + o_P(1)\,. \tag{A.5}$$

In (A.5), we have

$$\frac{d}{d\pi}{h_{\hat{\mathbf{s}}_N}}^{-1}(\pi)\Big|_{\pi=\pi_0} \overset{a.s.}{\to} \frac{d}{d\pi}h_{\mathbf{s}}^{-1}(\pi)\Big|_{\pi=\pi_0} = \frac{1}{h'_{\mathbf{s}}\left[{h_{\mathbf{s}}}^{-1}(\pi_0)\right]} = \frac{1}{h'_{\mathbf{s}}(p_0)}.$$

(using (2.1c) and the property that $\hat{\mathbf{s}}_N$ converges to $\mathbf{s}$ almost surely).

In addition, $\sqrt{N}(\hat{\pi}_N - \pi_0) \overset{\mathcal{D}}{\to} Z_1$ (see Lemma 1). By Slutsky's Theorem,

$$\sqrt{N}\left(\hat{p}_N - \hat{p}_N^{(0)}\right) \overset{\mathcal{D}}{\to} \frac{1}{h'_{\mathbf{s}}(p_0)}Z_1 \sim \mathcal{N}\left(0, V_1\right)\,, \tag{A.6}$$

where

$$V_1 = \left[h'_{\mathbf{s}}\,(p_0)\right]^{-2}\Sigma_{11} = \left[h'_{\mathbf{s}}\,(p_0)\right]^{-2}\pi_0(1-\pi_0)\,. \tag{A.7}$$

On the other hand, if we consider p to be fixed, we may view the reliability polynomial $h_{\mathbf{s}}(p)$ as a function of $\mathbf{s}$. From (1.1), we see that $h_{\mathbf{s}}(p) = \sum_i h_i(p)\,s_i$ is a linear function of the vector $\mathbf{s}$. Define $\nabla_{\mathbf{s}}h_{\mathbf{s}}(p)$ as the gradient of the $n \times 1$ vector $\mathbf{s}$ for fixed p. Then

$$\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(p) = \Big(h_1(p), h_2(p), \ldots, h_n(p)\Big)\,, \tag{A.8}$$

and hence

$$h_{\hat{\mathbf{s}}_N}(p) = h_{\mathbf{s}}(p) + \nabla_{\mathbf{s}}^T h_{\mathbf{s}}(p)\,(\hat{\mathbf{s}}_N - \mathbf{s})\,. \tag{A.9}$$

This linear expansion, along with a Taylor series expansion, helps us to determine the asymptotic behavior of the second term in the decomposition in (A.3). Indeed, setting the fixed argument p in (A.9) to be $\hat{p}_N^{(0)}$, we have

$$h_{\hat{\mathbf{s}}_N}(\hat{p}_N^{(0)}) = h_{\mathbf{s}}(\hat{p}_N^{(0)}) + \nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)})(\hat{\mathbf{s}}_N - \mathbf{s})\,. \tag{A.10}$$

Then, if we apply the inverse function ${h_{\mathbf{s}}}^{-1}$ to both sides of (A.10), and expand the function $h_{\mathbf{s}}^{-1}$ around $h_{\mathbf{s}}(\hat{p}_N^{(0)})$, Taylor's theorem shows that

$$\begin{aligned} h_{\mathbf{s}}^{-1}\left[h_{\hat{\mathbf{s}}_N}(\hat{p}_N^{(0)})\right] &= h_{\mathbf{s}}^{-1}\Big[h_{\mathbf{s}}\left(\hat{p}_N^{(0)}\right) + \nabla_{\mathbf{s}}^T h_{\mathbf{s}}\left(\hat{p}_N^{(0)}\right)(\hat{\mathbf{s}}_N - \mathbf{s})\Big] \\ &= h_{\mathbf{s}}^{-1}\big[h_{\mathbf{s}}(\hat{p}_N^{(0)})\big] + \left[\frac{d}{d\pi} h_{\mathbf{s}}^{-1}(\pi)\bigg|_{\pi = h_{\mathbf{s}}(\hat{p}_N^{(0)})}\right]\Big[\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)})(\hat{\mathbf{s}}_N - \mathbf{s})\Big] \\ &\quad + o\Big[\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)})(\hat{\mathbf{s}}_N - \mathbf{s})\Big] \\ &= \hat{p}_N^{(0)} + \left[\frac{d}{d\pi} h_{\mathbf{s}}^{-1}(\pi)\bigg|_{\pi = h_{\mathbf{s}}(\hat{p}_N^{(0)})}\right]\Big[\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)})(\hat{\mathbf{s}}_N - \mathbf{s})\Big] \\ &\quad + o_P(\|\hat{\mathbf{s}}_N - \mathbf{s}\|). \end{aligned} \tag{A.11}$$

It follows that $h_{\mathbf{s}}^{-1}\left[h_{\hat{\mathbf{s}}_N}(\hat{p}_N^{(0)})\right]$ can be expressed as

$$\hat{p}_N^{(0)} + \left[\frac{d}{d\pi} h_{\mathbf{s}}^{-1}(p)\bigg|_{\pi = h_{\mathbf{s}}(\hat{p}_N^{(0)})}\right]\Big[\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)})(\hat{\mathbf{s}}_N - \mathbf{s})\Big] + o_P(\|\hat{\mathbf{s}}_N - \mathbf{s}\|)\,,$$

while $h_{\mathbf{s}}^{-1}\left[h_{\hat{\mathbf{s}}_N}(\hat{p}_N^{(0)})\right]$ equals to p_0 as well, in view of the definitions in (2.1a)-(2.1e), since

$$p_0 = h_{\mathbf{s}}^{-1}(\pi_0) = h_{\mathbf{s}}^{-1}\left\{h_{\hat{\mathbf{s}}_N}\left[{h_{\hat{\mathbf{s}}_N}}^{-1}(\pi_0)\right]\right\} = h_{\mathbf{s}}^{-1}\left[h_{\hat{\mathbf{s}}_N}(\hat{p}_N^{(0)})\right]\,.$$

Hence

$$p_0 = \hat{p}_N^{(0)} + \left[\frac{d}{d\pi} h_{\mathbf{s}}^{-1}(\pi)\bigg|_{\pi = h_{\mathbf{s}}(\hat{p}_N^{(0)})}\right]\Big[\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)})(\hat{\mathbf{s}}_N - \mathbf{s})\Big] + o_P(\|\hat{\mathbf{s}}_N - \mathbf{s}\|)\,.$$

Furthermore,

$$\begin{aligned} &\hat{p}_N^{(0)} - p_0 \\ &= \hat{p}_N^{(0)} - \left\{\hat{p}_N^{(0)} + \left[\frac{d}{d\pi} h_{\mathbf{s}}^{-1}(\pi)\bigg|_{\pi = h_{\mathbf{s}}(\hat{p}_N^{(0)})}\right]\Big[\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)})(\hat{\mathbf{s}}_N - \mathbf{s})\Big] + o_P(\|\hat{\mathbf{s}}_N - \mathbf{s}\|)\right\} \\ &= \left\{-\left[\frac{d}{d\pi} h_{\mathbf{s}}^{-1}(\pi)\bigg|_{\pi = h_{\mathbf{s}}(\hat{p}_N^{(0)})}\right]\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)})\right\}(\hat{\mathbf{s}}_N - \mathbf{s}) + o_P(\|\hat{\mathbf{s}}_N - \mathbf{s}\|)\,. \end{aligned} \tag{A.12}$$

Since $\hat{p}_N^{(0)} = h_{\hat{\mathbf{s}}_N}{}^{-1}(\pi_0) \overset{a.s.}{\to} h_{\mathbf{s}}^{-1}(\pi_0) = p_0$, we have

(i) $h_{\mathbf{s}}(\hat{p}_N^{(0)}) \overset{a.s.}{\to} h_{\mathbf{s}}(p_0)\,,$

(ii) $\displaystyle \frac{d}{d\pi} h_{\mathbf{s}}^{-1}(\pi)\Big|_{\pi=h_{\mathbf{s}}(\hat{p}_N^{(0)})} \overset{a.s.}{\to} \frac{d}{d\pi} h_{\mathbf{s}}^{-1}(\pi)\Big|_{\pi=h_{\mathbf{s}}(p_0)} = \frac{1}{h'_{\mathbf{s}}\left\{h_{\mathbf{s}}{}^{-1}\left[h_{\mathbf{s}}(p_0)\right]\right\}} = \frac{1}{h'_{\mathbf{s}}(p_0)}\,,$

(iii) $\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)}) \overset{a.s.}{\to} \nabla_{\mathbf{s}}^T h_{\mathbf{s}}(p_0)\,,$

and hence the gradient in (A.12) satisfies

(iv) $\displaystyle -\left[\frac{d}{d\pi} h_{\mathbf{s}}^{-1}(\pi)\Big|_{\pi=h_{\mathbf{s}}(\hat{p}_N^{(0)})}\right] \nabla_{\mathbf{s}}^T h_{\mathbf{s}}(\hat{p}_N^{(0)}) \overset{a.s.}{\to} \frac{1}{h'_{\mathbf{s}}(p_0)} \nabla_{\mathbf{s}}^T h_{\mathbf{s}}(p_0)\,.$

Applying Slusky's Theorem to (A.12), we have:

$$\sqrt{N}\left(\hat{p}_N^{(0)} - p_0\right) \overset{\mathcal{D}}{\to} -\frac{1}{h'_{\mathbf{s}}(p_0)} \nabla_{\mathbf{s}}^T h_{\mathbf{s}}(p_0) Z_2 \sim \mathcal{N}(0, V_2)\,, \tag{A.13}$$

where $V_2 = [h'(p_0)]^{-2}\ \nabla_{\mathbf{s}}^T h_{\mathbf{s}}(p_0)\,\Sigma_{22}\,\nabla_{\mathbf{s}} h_{\mathbf{s}}(p_0)$.

To calculate the value of V_2, recall that from (A.8) that

$$\nabla_{\mathbf{s}} h_{\mathbf{s}}(p_0) = \Big(h_1(p_0), h_2(p_0), \ldots, h_n(p_0)\Big)^T.$$

Thus,

$$\begin{aligned}
V_2 &= \left[h'(p_0)\right]^{-2}\Big(h_1(p_0), \ldots, h_n(p_0)\Big)\begin{pmatrix} s_1(1-s_1) & \ldots & -s_1 s_n \\ -s_1 s_2 & \ldots & -s_2 s_n \\ \ldots & \ldots & \ldots \\ -s_1 s_n & \ldots & s_n(1-s_n)\end{pmatrix}\begin{pmatrix} h_1(p_0) \\ h_2(p_0) \\ \ldots \\ h_n(p_0)\end{pmatrix} \\
&= \left[h'(p_0)\right]^{-2}\Big[\sum_i s_i(1-s_i) h_i^2(p_0) - \sum_{i\neq j} s_i s_j h_i(p_0) h_j(p_0)\Big] \\
&= \left[h'(p_0)\right]^{-2}\Big[\sum_i s_i h_i^2(p_0) - \sum_i s_i^2 h_i^2(p_0) - \sum_{i\neq j} s_i s_j h_i(p_0) h_j(p_0)\Big] \\
&= \left[h'(p_0)\right]^{-2}\Big[\sum_i s_i h_i^2(p_0) - \sum_{i,j} s_i s_j h_i(p_0) h_j(p_0)\Big] \\
&= \left[h'(p_0)\right]^{-2}\Big\{\sum_i s_i h_i^2(p_0) - \Big[\sum_i s_i h_i(p_0)\Big]^2\Big\} \\
&= \left[h'(p_0)\right]^{-2}\Big[\sum_i s_i h_i^2(p_0) - h_{\mathbf{s}}^2(p_0)\Big].
\end{aligned} \tag{A.14}$$

Now, it is apparent from (A.6) and (A.13) that the two summands in the last line of (A.3) have the limits

$$\sqrt{N}\left(\hat{p}_N - \hat{p}_N^{(0)}\right) \overset{\mathcal{D}}{\to} \frac{1}{h'_{\mathbf{s}}(p_0)} Z_1 \sim \mathcal{N}\left(0, V_1\right),$$
$$\sqrt{N}\left(\hat{p}_N^{(0)} - p_0\right) \overset{\mathcal{D}}{\to} -\frac{1}{h'_{\mathbf{s}}(p_0)} \nabla_{\mathbf{s}}^T h_{\mathbf{s}}(p_0) Z_2 \sim \mathcal{N}(0, V_2).$$

The asymptotic covariance between the summands is hence

$$V_{12}=\mathrm{Cov}\Big(\frac{1}{h'_{\mathbf{s}}(p_0)}Z_1,-\frac{1}{h'_{\mathbf{s}}(p_0)}\nabla_{\mathbf{s}}^T h_{\mathbf{s}}(p_0)Z_2\Big)=-\left[h'_{\mathbf{s}}(p_0)\right]^{-2}\mathrm{Cov}(Z_1,Z_2)\,\nabla_{\mathbf{s}}h_{\mathbf{s}}(p_0).$$

It follows from (2.2) in Lemma 1 that

$$\begin{aligned}
V_{12} &= -\left[h'_{\mathbf{s}}(p_0)\right]^{-2}\Sigma_{12}\,\nabla_{\mathbf{s}}h_{\mathbf{s}}(p_0)\\
&= -\left[h'_{\mathbf{s}}(p_0)\right]^{-2}\Big(s_1\left[h_1(p_0)-\pi_0\right],\ldots,s_n\left[h_n(p_0)-\pi_0\right]\Big)\begin{pmatrix}h_1(p_0)\\h_2(p_0)\\\cdots\\h_n(p_0)\end{pmatrix}\\
&= -\left[h'_{\mathbf{s}}(p_0)\right]^{-2}\sum_i\left\{s_i\left[h_i(p_0)-\pi_0\right]h_i(p_0)\right\}\\
&= -\left[h'_{\mathbf{s}}(p_0)\right]^{-2}\Big[\sum_i s_i h_i^2(p_0)-\sum_i s_i h_i(p_0)\pi_0\Big]\\
&= -\left[(h'_{\mathbf{s}}(p_0)\right])^{-2}\Big[\sum_i s_i h_i^2(p_0)-h_{\mathbf{s}}^2(p_0)\Big]=-V_2\,. \qquad \text{(A.15)}
\end{aligned}$$

Finally, the summation in (A.3) has the following asymptotic property: $\sqrt{N}\,(\hat{p}_N-p_0)\overset{\mathcal{D}}{\to}\mathcal{N}(0,\sigma^2)$, where

$$\begin{aligned}
\sigma^2 &= V_1+V_2+2V_{12}=V_1+V_2-2V_2=V_1-V_2 \qquad \text{(A.16)}\\
&= \left[h'_{\mathbf{s}}(p_0)\right]^{-2}\pi_0(1-\pi_0)-\left[h'_{\mathbf{s}}(p_0)\right]^{-2}\Big[\sum_i s_i h_i^2(p_0)-h_{\mathbf{s}}^2(p_0)\Big]\\
&= \left[h'_{\mathbf{s}}(p_0)\right]^{-2}\Big\{h_s(p_0)\left[1-h_s(p_0)\right]-\sum_i s_i h_i^2(p_0)+h_{\mathbf{s}}^2(p_0)\Big\}\\
&= \left[h'_{\mathbf{s}}(p_0)\right]^{-2}\Big[h_s(p_0)-\sum_i s_i h_i^2(p_0)\Big]\\
&= \left[h'_{\mathbf{s}}(p_0)\right]^{-2}\sum_i s_i h_i(p_0)\left[1-h_i(p_0)\right]\,. \qquad \text{(A.17)}
\end{aligned}$$

Proof of Theorem 3. In fact, $\hat{\hat{F}}_0(t)$ can be written as $h_{\mathbf{s}}^{-1}(\hat{\pi}_N)$, then similar to the proof in Theorem 2, we have

$$\sqrt{N}\left[h_{\mathbf{s}}^{-1}(\hat{\pi}_N)-h_{\mathbf{s}}^{-1}(\pi_0)\right]\overset{\mathcal{D}}{\to}\mathcal{N}\left(0,\left[\frac{d}{d\pi}h_{\mathbf{s}}^{-1}(\pi)\Big|_{\pi=\pi_0}\right]^2\mathrm{Var}(Z_1)\right), \qquad \text{(A.18)}$$

by using the delta method. The asymptotic variance is therefore

$$\left[\frac{d}{d\pi}h_{\mathbf{s}}^{-1}(\pi)\Big|_{\pi=\pi_0}\right]^2\mathrm{Var}(Z_1)=V_1\,,$$

and moreover, the asymptotic variance of $\hat{\hat{F}}(t)$ has been proven to be $\sigma^2 = V_1 - V_2 \leq V_1$ in (A.16).

A.2. An additional example for Section 2.3

In Section 2, we provide an elementary example showing that the estimator of a known nuisance parameter can indeed prove efficacious in estimating a given target parameter. Here we provide another example in a linear model setting that shares a similar feature.

In the context of small area estimation (see, e.g., Rao and Molina (2015)), the Fay-Herriot model (see Fay and Herriot (1979)) is well known for estimating the small area means with area-level data. The model can be expressed as

$$y_i = X_i^T \beta + V_i + e_i, \quad i = 1, \ldots, m,$$

where X_i is a vector of known covariates, β is a vector of unknown regression coefficients, V_i is a random effect, and e_i is a sampling error. It is assumed that $V_i, e_i, i = 1, \ldots, m$ are independent with $V_i \sim N(0, A)$ and $e_i \sim N(0, D_i)$, where $D_i, i = 1, \ldots, m$ are known variances, but A is a additional variance.

Under the traditional Fay-Herriot model, the random effects, V_i, are unobserved, and A is unknown. Below we consider two variations from the traditional setting. First, assume that A is known that A is known but that the $V_i^{'}$s are unobserved. In this case, the observations are $Y_i, i = 1, \ldots, m$, and it is well known [see e.g., Jiang (2007, (2.33))] that the MLE of β is given by $\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y$, where $Y = (Y_i)_{1 \leq i \leq m}$, $X = (X_i^T)_{1 \leq i \leq m}$, and $V = AI_m + D$, with I_m being the m-dimensional identity matrix and $D = \text{daig}(D_1, \ldots, D_m)$. It follows that the covariance matrix of $\hat{\beta}$ is $\text{Var}(\hat{\beta}) = (X^T V^{-1} X)^{-1}$.

Now suppose, in addition, that the V_i are actually observed. Then, the observations are $(Y_i, V_i), i = 1, \ldots, m$. To derive the MLE of β, one can write the joint density of (Y, V), with $V = (V_i)_{1 \leq i \leq m}$, as $f(Y, V|\beta, A) = f(Y|V, \beta) f(V|A)$, where $f(Y|V, \beta)$ denotes the conditional density of Y given V, and $f(V|A)$ the density of V. Although A is known, below we ignore this and try to "estimate" A via maximum likelihood. However, because $f(V|A)$ does not involve β, the MLE of β is the same as the maximizer of $f(Y|V, \beta)$, over β (while the MLE of A is the same as the maximizer of $f(V|A)$ over A). Furthermore, we have

$$f(Y|V, \beta) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi D_i}} \exp\left\{ -\frac{(Y_i - V_i - X_i^T \beta)^2}{2D_i} \right\}.$$

The latter is the same as the density function under the above Fay-Herriot model with Y replaced by $Y - V$ and $A = 0$ (hence $V = D$). Thus, by the above results, the MLE of β is $\tilde{\beta} = (X^T D^{-1} X)^{-1} X^T D^{-1}(Y - V)$, whose covariance matrix is $\text{Var}(\tilde{\beta}) = (X^T D^{-1} X)^{-1}$.

Because $V \geq D$ (for symmetric matrices A, B, $A \geq B$ iff $A - B$ is nonnegative definite), it follows that $V^{-1} \leq D^{-1}$ (e.g., Jiang (2010, (i) of Sec. 5.3.1)); hence $X'V^{-1}X \leq X'D^{-1}X$. Thus, by the same argument, we have

$$\text{Var}(\hat{\beta}) = (X'V^{-1}X)^{-1} \geq (X'D^{-1}X)^{-1} = \text{Var}(\tilde{\beta}).$$

The last result shows that, if the V_i's are available, one can actually do better in estimating β by estimating a nuisance parameter, A, which one already knows.

References

Bueno, V. C. (1988). A note on the component lifetime estimation of a multistate monotone system through the system lifetime. *Adv. Appl. Probab.* **20**, 686-689.

Bhattacharya, D. and Samaniego, F. J. (2010). Estimating component characteristics from system failure-time data. *Naval Res. Logist.* **57**, 380-389.

Barlow, R. E. and Proshan, F.(1981). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York.

Boyles, R. A. and Samaniego, F. J. (1986). Estimating a distribution function based on nomination sampling. *J. Amer. Statist. Assoc.* **81**, 1039-1045.

Boyles, R. A. and Samaniego, F. J. (1987). On estimating component reliability for systems with random redundancy levels. *IEEE Trans. Reliab.* **36**, 403-407.

Fay, R. and Herriot, R. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.* **74**, 269-277.

Guess, F. M., Usher, J. S. and Hodgson, T. J. (1991). Estimating system and component reliabilities under partial information on cause of failure. *J. Statist. Plann. Inference* **29**, 75-85.

Hall, P. G., Jin, Y. and Samaniego, F. J. (2015). Estimating component reliability using lifetime data from systems of varying design. *Statist. Sinica* **25**, 1313-1335.

Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and Their Applications*. Springer Science & Business Media, New York.

Jiang, J. (2010). *Large Sample Techniques for Statistics*. Springer Science & Business Media, New York.

Kvam, P. H. and Samaniego, F. J. (1993a). On the inadmissibility of empirical averages as estimators in ranked set sampling. *J. Statist. Plann. Inference* **36**, 39-55.

Kvam, P. H. and Samaniego, F. J. (1993b). On maximum likelihood estimation based on ranked set samples, with applications to reliability. In *Advances in Reliability* (Edited by A. Basu), 215-229. The Netherlands.

Kvam, P. H. and Samaniego, F. J. (1994). Nonparametric maximum likelihood estimation based on ranked set samples. *J. Amer. Statist. Assoc.* **89**, 526-537.

Meilijson, I. (1981). Estimation of the lifetime distribution of the parts from the autopsy statistics of the machine. *J. Appl. Probab.* **13**, 829-838.

Miyakawa, M. (1984). Analysis of incomplete data in competing risks model. *IEEE Trans. Reliab.* **4**, 293-296.

Moeschberger, M. L. and David, H. A. (1971). Life tests under competing cause of failure and the theory of competing risks. *Biometrics* **27**, 909-933.

Patterson, D. A., Gibson, G. and Katz, R. H. (1988). A case for redundant arrays of inexpensive disks (RAID). *J. Assoc. Comput. Machin.* **17**, 109-116.

Rao, J. N. K. and Molina, I. (2015). *Small Area Estimation.* Wiley, New York.

Samaniego, F. J. (1985). On closure of the IFR class under formation of coherent systems. *IEEE Trans. Reliab.* **34**, 60-72.

Samaniego, F. J. (2007). *System Signatures and Their Applications in Engineering Reliability.* Springer Science & Business Media, New York.

Stokes, S. L. and Sager, T. W. (1988). Characterization of a ranked set sample with applications to estimating distribution functions. *J. Amer. Statist. Assoc.* **75**, 908-911.

Usher, J. S. and Hodgson, T. J. (1988). Maximum likelihood analysis of component reliability using masked system life-test data. *IEEE Trans. Reliab.* **37**, 550-555.

Department of Statistics, University of California, 399 Crocker Lane, Davis, CA 95616, USA.

E-mail: golden.ucd@gmail.com

Department of Statistics, University of California, 399 Crocker Lane, Davis, CA 95616, USA.

Department of Mathematics and Statistics, University of Melbourne, Melbourne, VIC 3010, Australia.

E-mail: P.Hall@ms.unimelb.edu.au

Department of Statistics, University of California, 399 Crocker Lane, Davis, CA 95616, USA.

E-mail: jiang@wald.ucdavis.edu

Department of Statistics, University of California, 399 Crocker Lane, Davis, CA 95616, USA.

E-mail: fjsamaniego@ucdavis.edu

(Received June 2015; accepted April 2016)