Harvard University
Harvard University Biostatistics Working Paper Series

Year 2008 *Paper* 94

The Highest Confidence Density Region and Its Usage for Inferences about the Survival Function with Censored Data

Lu Tian* Rui wang†
Tianxi Cai‡ L. J. Wei**

*Stanford University School of Medicine, lutian@stanford.edu
†Massachusetts General Hospital, rwang8@partners.org
‡Harvard University, tcai@hsph.harvard.edu
**Harvard University, wei@hsph.harvard.edu

http://biostats.bepress.com/harvardbiostat/paper94

The highest confidence density region and its usage for inferences about the survival function with censored data

By Lu Tian

Department of Health Policy and Research, Stanford University School of Medicine, Stanford, California 94305, USA.

lutian@stanford.edu

Rui Wang

Department of Biostatistics, Massachusetts General Hospital, 50 Staniford Street, Boston, MA 02115, USA.

rwang8@partners.org

Tianxi Cai and L. J. Wei

Department of Biostatistics, Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115, USA.

tcai@hsph.harvard.edu, wei@hsph.harvard.edu

Summary

Suppose that we are interested in making inferences about a vector of constrained parameters. Confidence regions for these parameters are often constructed via a normal approximation to the distribution of a consistent estimator for a transformation of the parameters. In this article, we utilize the *confidence distribution,* a frequentist counterpart to the posterior distribution in Bayesian statistics, to obtain optimal confidence regions for the parameters. Members of such a region can be efficiently generated via a standard Markov chain Monte Carlo algorithm. We then apply this technique to draw inferences about the temporal profile of the survival function with censored observations. We illustrate the new proposal with the survival data from the well-known Mayo primary biliary cirrhosis study and show that the volume of the new 0.95 confidence region is only one thirty fourth of that of the conventional confidence band.

Some key words: Confidence distribution; Highest posterior density region; Markov chain Monte Carlo; Simultaneous confidence intervals; Survival analysis.

1. Introduction

Let $\boldsymbol{\theta}_0$ be a vector of the unknown true values of p parameters. Suppose that we are interested in constructing a confidence region for $\boldsymbol{\theta}_0$ with a pre-specified confidence level. Often these parameters have certain intrinsic constrains. Conventionally such a confidence region is obtained via an estimator of a transformation of $\boldsymbol{\theta}_0$. For example, to make inferences about the Pearson correlation coefficient $\boldsymbol{\theta}_0$ between two univariate random variables, we customarily utilize a normal approximation to the distribution of

the Fisher transformation

$$0.5\log\{(1+\hat{\boldsymbol{\theta}})/(1-\hat{\boldsymbol{\theta}})\}$$

of the sample correlation coefficient $\hat{\boldsymbol{\theta}}$ to obtain confidence intervals of the transformed parameter. We then apply the inverse of this transformation to the boundaries of such intervals to obtain confidence intervals for $\boldsymbol{\theta}_0$.

As another example, consider the simple one-sample problem with censored data in survival analysis discussed in Chapter 5 of Fleming & Harrington (1991). Let the kth component θ_{k0} of $\boldsymbol{\theta}_0$ be the survival probability at time $t_k, k = 1, \cdots, p$, where $t_1 < \cdots < t_p$. Here, the constrain is $0 \le \theta_{p0} \le \cdots \le \theta_{10} \le 1$. A commonly used transformation for this case is the complementary log-log. Specifically, for the kth component of $\boldsymbol{\theta}_0$, let $\gamma_{k0} = \log\{-\log(\theta_{k0})\}$ and $\boldsymbol{\gamma}_0' = (\gamma_{10}, \cdots, \gamma_{p0})$. One then uses a standard consistent estimator $\hat{\boldsymbol{\gamma}}' = (\hat{\gamma}_1, \cdots, \hat{\gamma}_p)$ for $\boldsymbol{\gamma}_0$, whose distribution can be approximated well by a multivariate normal with mean $\boldsymbol{\gamma}_0$ when the sample size n is large. A conventional $(1-\alpha), 0 < \alpha < 1$, confidence region or simultaneous confidence band for $\boldsymbol{\gamma}_0$ is a p-dimensional rectangular

$$\Pi_{k=1}^{p}(\hat{\gamma}_k - c\hat{\sigma}_k, \hat{\gamma}_k + c\hat{\sigma}_k), \tag{1.1}$$

where $\hat{\sigma}_k$ is the standard error estimate for $\hat{\gamma}_k$ and the cutoff point c is chosen such that

$$\mathrm{pr}(\max_{\{1\le k\le p\}}|\hat{\gamma}_k - \gamma_k|/\hat{\sigma}_k \le c) = 1-\alpha. \tag{1.2}$$

Note that c in (1.2) can be easily obtained via the resampling-perturbation method proposed by (Parzen et al., 1994). It follows that an asymptotic $(1-\alpha)$ confidence region for $\boldsymbol{\theta}_0$ via (1.1) is a p-dimensional rectangular

$$\Pi_{k=1}^{p}(\exp\{-\exp(\hat{\gamma}_k - c\hat{\sigma}_k)\}, \exp\{-\exp(\hat{\gamma}_k + c\hat{\sigma}_k)\}). \tag{1.3}$$

This type of simultaneous confidence intervals has been extensively discussed, for example, in Fleming & Harrington (1991, Chapter 6.3).

As an illustration, we use the well-known Mayo primary biliary cirrhosis mortality data (Fleming & Harrington, 1991, Appendix D) to construct the confidence region (1.3) for the survival function. The data were from a randomized trial in primary biliary cirrhosis of the liver conducted at Mayo Clinic. A detailed description of this data set can be found in Fleming & Harrington (1991, p.2). Here, we used all survival information from 418 patients in the study for illustration. Figure 1 presents the Kaplan-Meier curve (solid curve) for the survival probabilities and the corresponding 0.95 simultaneous confidence band (gray area) obtained via (1.3) for seven time points, $(t_1, \cdots, t_7) = (2, 4, 6, 7, 8, 9, 10)$(years). It is important to note that the complementary log-log transformation for the survival probability does not take into account of the ordered constrain among $\{\theta_{k0}, k = 1, \cdots, p\}$. Therefore, the resulting confidence region in Figure 1 contains functions which are not non-increasing over time and, therefore, are not possible candidates for $\boldsymbol{\theta}_0$.

In this article, under a general setting we utilize the *confidence distribution* to construct classical confidence regions for $\boldsymbol{\theta}_0$. We show that the $(1-\alpha)$ highest confidence density region, HCDR, is a *bona fide* $(1-\alpha)$ asymptotic confidence region and has the smallest volume among a rather large class of $(1-\alpha)$ confidence regions for $\boldsymbol{\theta}_0$. Points in the HCDR can be obtained efficiently via a Markov chain Monte Carlo procedure. The concept of a confidence distribution is purely based on a frequentist interpretation and can be regarded as a counterpart to the posterior distribution in Bayesian statistics. The inference procedures using the confidence distribution have been discussed extensively,

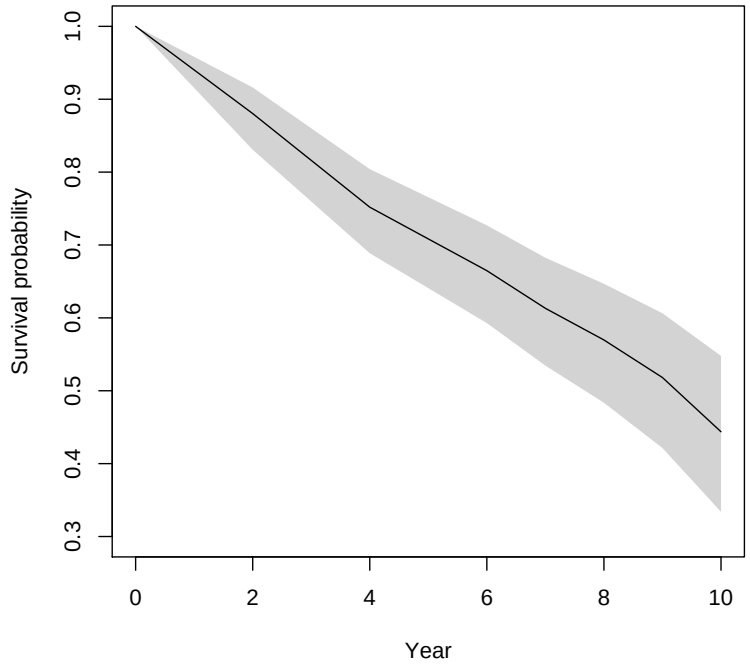

Fig. 1. The traditional 0.95 confidence band (gray region) and the Kaplan-Meier estimator (solid curve) for the survival probabilities at Years 2, 4, 6, 7, 8, 9 and 10 based on the Mayo primary biliary cirrhosis mortality data

for example, by Efron (1993, 1998), Fraser (1991, 1996), Lehmann (1986, 1993), Schweder & Hjort (2002, 2003), and Singh et al. (2007).

We illustrate the new proposal with the data from the above Mayo liver study. We show that the volume of the resulting 0.95 highest confidence density region for the aforementioned seven survival probabilities is one thirty fourth of the volume of the confidence region displayed in Figure 1. After deleting those impossible candidates for the survival function from the band in Figure 1, the resulting region is still 9 times as big as the HCDR with respect to the volume. Lastly, to draw inferences about the temporal profile of the survival function, we show how to generate possible candidates of the survival function from the 0.95 HCDR via a MCMC algorithm.

2. The highest confidence density region

Let $A \in R^p$ be the set of all possible values of $\boldsymbol{\theta}_0$ and X be the data. Assume that there is a smooth, one-to-one transformation $g(\cdot)$ from A to R^p such that any point in R^p is a possible value of $\boldsymbol{\omega}_0 = g(\boldsymbol{\theta}_0)$. Furthermore, assume that there is a consistent estimator $\hat{\boldsymbol{\omega}}$ such that with a large sample size n, the distribution of $(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)$ can be approximated well by a normal with mean 0 and covariance matrix $\hat{\Sigma}$. Moreover, assume that $n\hat{\Sigma}$ converges to a deterministic matrix, as $n \to \infty$.

Let Ω be a random vector whose distribution is $\mathrm{MN}(\hat{\boldsymbol{\omega}}, \hat{\Sigma})$, which can be interpreted as a confidence distribution of $\boldsymbol{\omega}_0$ (Singh et al., 2007, Chapter 5). From a frequentist point of view, this distribution contains all the information about $\boldsymbol{\omega}_0$. For example, let $D_\alpha \in R^p$ be a region such that

$$\mathrm{pr}(\Omega \in \hat{\boldsymbol{\omega}} + D_\alpha \mid X) = 1 - \alpha,$$

where the probability is with respect to Ω conditional on X. Then, asymptotically the set

$$\hat{D}_\alpha = \hat{\boldsymbol{\omega}} + D_\alpha \tag{2.1}$$

is a $(1-\alpha)$ confidence region for $\boldsymbol{\omega}_0$.

Now, let the random vector $\boldsymbol{\Theta} = g^{-1}(\boldsymbol{\Omega})$ and $\hat{f}(\boldsymbol{\theta})$ be the density function of $\boldsymbol{\Theta}$. Based on Theorem 5.1 of Singh et al. (2007), $\hat{f}(\boldsymbol{\theta})$ is a confidence density function of $\boldsymbol{\theta}_0$. Consider a region $\hat{R}_\alpha \subset A$ such that

$$\text{pr}(\boldsymbol{\Theta} \in \hat{R}_\alpha \mid \ X) \geq 1-\alpha. \tag{2.2}$$

Furthermore, assume that $\hat{R}_\alpha$ is an asymptotic confidence region for $\boldsymbol{\theta}_0$ with at least $(1-\alpha)$ confidence level. The class of such regions is quite large. For instance, any set $g^{-1}(\hat{D}_\alpha)$ from (2.1) is in this class. Given data X, let

$$\hat{P}_\alpha = \{\boldsymbol{\theta} : \ \hat{f}(\boldsymbol{\theta}) \geq d\}, \tag{2.3}$$

where d is obtained such that

$$\text{pr}(\boldsymbol{\Theta} \in \hat{P}_\alpha \mid X) \geq 1-\alpha. \tag{2.4}$$

It follows from the justification for the optimality property of the highest posterior density region in Bayesian statistics (Box & Tiao, 1992) that the $(1-\alpha)$ "credible" region $\hat{P}_\alpha$ has the smallest "volume" among all $(1-\alpha)$ confidence or confidence regions $\hat{R}_\alpha$ in (2.2). However, it is not clear that the credible region $\hat{P}_\alpha$ is an asymptotic $(1-\alpha)$ confidence region. In the Appendix, we show that this region is indeed a $(1-\alpha)$ confidence region.

To obtain an approximation to the cut-off point d in (2.3), one may generate a large number, M, of realizations $\{\boldsymbol{\theta}_{(j)} = g^{-1}(\boldsymbol{\omega}_{(j)}), j = 1, \cdots, M\}$, where $\{\boldsymbol{\omega}_{(j)}\}$ are independent realizations from the multivariate normal vector $\boldsymbol{\Omega}$. Then the 100αth empirical percentile of $\{\hat{f}(\boldsymbol{\theta}_{(j)}) \mid j = 1, \cdots, M\}$ can be treated as an estimate for the cut-off d.

To generate points in the optimal region (2.3) or equivalently realizations from a uniform random vector Ψ with its support being $\hat{P}_\alpha$, one may employ a standard Metropolis MCMC algorithm (Liu, 2001). Note that for any set $B \subset \hat{P}_\alpha$, $\text{pr}(\Psi \in B)$ is the ratio of the volumes of the above two sets. Specifically, one may let $\boldsymbol{\psi}_0 = \hat{\boldsymbol{\theta}} = g^{-1}(\hat{\boldsymbol{\omega}})$. For $j = 1, \cdots, M_0$, we iteratively generate realizations $\boldsymbol{\psi}$ via the Markov chain $\boldsymbol{\psi}_j = \boldsymbol{\psi}_{j-1} I\{g(\boldsymbol{\theta}^*) < d\} + \boldsymbol{\theta}^* I\{g(\boldsymbol{\theta}^*) \geq d\}$, where $I(\cdot)$ is the indicator function and $\boldsymbol{\theta}^*$ is simulated from the proposal multivariate normal distribution $\text{MN}(\boldsymbol{\psi}_{j-1}, \widetilde{\Sigma})$. Here, an obvious candidate for $\widetilde{\Sigma}$ is a matrix which is proportional to the variance-covariance matrix for $g^{-1}(\hat{\boldsymbol{\omega}})$ derived from the standard δ-method with the aforementioned $\hat{\Sigma}$. Now, deleting the first L_0 "burn-in period" realizations from this chain, the resulting $\mathcal{V} = \{\boldsymbol{\psi}_j, L_0 \leq j \leq M_0\}$ are realizations approximately from the uniform vector Ψ on $\hat{P}_\alpha$. Like any other problem handled with MCMC procedures, the efficiency of this Metropolis algorithm depends on the choice of the proposal distribution. The scaling parameter can be tuned to control the rejection rate of the Markov chain.

Note that the volume of the confidence region $\hat{P}_\alpha$ can be easily estimated by

$$M^{-1} \sum_{j=1}^{M} \frac{I(\hat{f}(\boldsymbol{\theta}_{(j)}) \geq d)}{\hat{f}(\boldsymbol{\theta}_{(j)})}. \tag{2.5}$$

3. Application to one sample problem in survival analysis

Let $\tilde{T}$ be a random continuous "survival" time and its survival function be $S(t) = \text{pr}(\tilde{T} \geq t)$ for $t \geq 0$. Let $\boldsymbol{\theta}_0' = (S(t_1), \cdots, S(t_p))$ as discussed in the Introduction section. Let C be the independent censoring variable for $\tilde{T}$. One can only observe $T = \min(\tilde{T}, C)$ and $\Delta = I(T = \tilde{T})$. Furthermore, let the data $\{(T_i, \Delta_i), i = 1, \cdots, n\}$ be n independent copies of (T, Δ). We are interested in obtaining the HCDR for $\boldsymbol{\theta}_0$ with such censored data.

Now, let $\Lambda(t) = -\log\{S(t)\}$, the cumulative hazard function of $\tilde{T}$. Also, let

$$\boldsymbol{\omega}_0 = g(\boldsymbol{\theta}_0) = (\log\{\Lambda(t_1)\}, \log\{\Lambda(t_2) - \Lambda(t_1)\}, \cdots, -\log\{\Lambda(t_p) - \Lambda(t_{p-1})\})'.$$

Note that with this transformation $g(\cdot)$, any point in R^p is a possible value for $\boldsymbol{\omega}_0$. To obtain $\hat{\boldsymbol{\omega}}$, let $\hat{\Lambda}(t) = \sum_{i=1}^n \int_0^t Y^{-1}(s) dN_i(s)$, the Nelson estimator for $\Lambda(t)$, where $Y(t) = \sum_{i=1}^n I(T_i \geq t)$ and $N_i(t) = I(T_i \leq t)\Delta_i$. Then, it follows from the resampling-perturbation method (Lin et al., 1994) that the distribution of

$$\{\hat{\Lambda}(t_k) - \hat{\Lambda}(t_{k-1})\} - \{\Lambda(t_k) - \Lambda(t_{k-1})\}, 1 \leq k \leq p, \tag{3.1}$$

can be approximated well by the conditional distribution (conditional on the data) of

$$\sum_{i=1}^n \int_{t_{k-1}}^{t_k} Y^{-1}(s) dN_i(s) G_i, \tag{3.2}$$

where $t_0 = 0$ and $\{G_i\}$ is a random sample from $\text{N}(0, 1)$, which is independent of the data. The distribution of (3.2) can be approximated easily with a large number of realized samples from $\{G_i\}$. The joint distribution of the standardized (3.1) is normal asymptotically. With the standard δ-method, one can show that $(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)$ is approximately normal with mean 0 and covariance matrix $\hat{\Sigma}$, which is the corresponding consistent estimator for the covariance matrix of $\hat{\boldsymbol{\omega}}$.

Let $\boldsymbol{\Omega}$ be a $\text{MN}(\hat{\boldsymbol{\omega}}, \hat{\Sigma})$ and $\boldsymbol{\Theta} = g^{-1}(\boldsymbol{\Omega})$. Then, $\boldsymbol{\Theta}$ has the density function

$$\hat{f}(\boldsymbol{\theta}) = (|2\pi\hat{\Sigma}|)^{-1/2} \exp(-\frac{\{g(\boldsymbol{\theta}) - \hat{\boldsymbol{\omega}}\}'\hat{\Sigma}^{-1}\{g(\boldsymbol{\theta}) - \hat{\boldsymbol{\omega}}\}}{2}) \left|\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|,$$

where

$$\left|\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right| = \left[\Pi_{k=1}^p \theta_k \Pi_{k=1}^p \{\log(\theta_{k-1}) - \log(\theta_k)\}\right]^{-1},$$

$0 \leq \theta_p \leq \cdots \leq \theta_1$. The $(1 - \alpha)$ HCDR $\hat{P}_\alpha$ (2.3) can then be obtained accordingly.

To illustrate the new proposal, we use the survival data set from the Mayo liver study discussed in the Introduction. First, suppose that we are interested in making inference about the survival probabilities only at Years 9 and 10. Here, $p = 2$ and $\boldsymbol{\theta}_0' = \{S(9), S(10)\}$. The conventional 0.95 confidence band is the rectangular region given in Figure 2. The black dot is the estimate $\hat{\boldsymbol{\theta}} = g^{-1}(\hat{\boldsymbol{\omega}}) = (0.52, 0.44)'$. The light gray area is the region obtained by deleting points that violate the constrain $S(9) \geq S(10)$ in the above rectangular. To obtain the 0.95 HCDR for $\boldsymbol{\theta}_0$, we generated $M = 10^6$ $\boldsymbol{\theta}_{(j)}$ to obtain the cutoff point $d = 8.50$ in (2.3). To locate points in $\hat{P}_{0.05}$, we let the proposal distribution be $\text{MN}(\boldsymbol{\psi}_{j-1}, 3 \times \widetilde{\Sigma})$, where $\widetilde{\Sigma}$ is the estimated variance-covariance matrix of $g^{-1}(\hat{\boldsymbol{\omega}})$. Deleting the first $L_0 = 3000$ realizations of Ψ in the run-in period, we generated 5×10^5 $\boldsymbol{\psi}$'s. The average rejection rate is 54.2%. To examine whether the generated

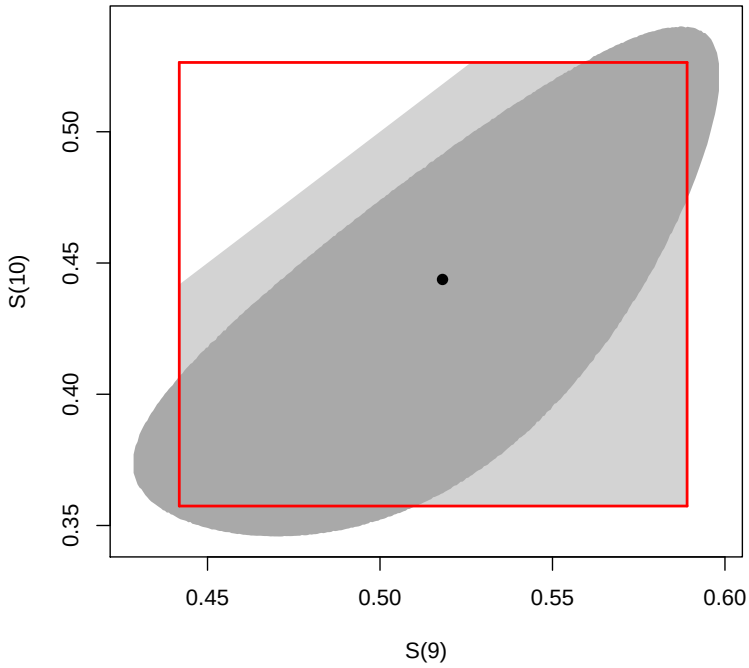

Fig. 2. The 0.95 HCDR (dark gray), the conventional rectangular-shape confidence region (rectangular), and the trimmed conventional confidence region considering order constrains (light gray) for $\boldsymbol{\theta}_0 = \{S(9), S(10)\}'$ based on the Mayo primary biliary cirrhosis mortality data. The black dot is the point estimator for the survival probabilities.

realizations from the Markov chain were stabilized, we continued generating another set of fresh 5×10^5 $\boldsymbol{\psi}$'s. The frequency distributions from these two sets of $\boldsymbol{\psi}$'s are quite similar, indicating that after the run-in period, the Markov chain became stationary. Lastly, we combined these two sets of generated $\boldsymbol{\psi}$'s to estimate the distribution of the uniform random vector over $\hat{P}_{0.05}$. In Figure 2, the dark gray region is the resulting 0.95 HCDR. The conventional rectangular region and its trimmed counterpart (light gray) are 48% and 27% bigger than the HCDR, respectively.

Next, suppose that we are interested in the survival function at the seven time points discussed in the Introduction. Using the same procedure as the above two-dimensional case, for $\alpha = 0.05$, the cutoff point $d = 2.60 \times 10^6$. The 0.95 conventional confidence region in Figure 1 is 34 times as big as the HCDR. By deleting the impossible candidates for the survival function from the above confidence region, the resulting region is still 9 times as big as the HCDR.

The only advantage of using the conventional band (1.3) for the survival function is its ease of the graphical display shown in Figure 1. For visualizing typical members of the 0.95 HCDR, we chose points from the above $M = 10^6$ realizations of the Markov chain sequentially, but separated from each other by a block of 100 realizations of the chain. This may sample up to 10^4 empirically "uncorrelated" sample points from this optimal region. We can then interactively display each member using a survival curve by connecting seven components of the above sample point over time. In Figure 3, we show six such curves with the observed Kaplan-Meier estimate. In contrast to the conventional confidence band in Figure 1, these curves show the possible temporal patterns of the true survival function. Like the curves presented in Figure 3, in general we find that the

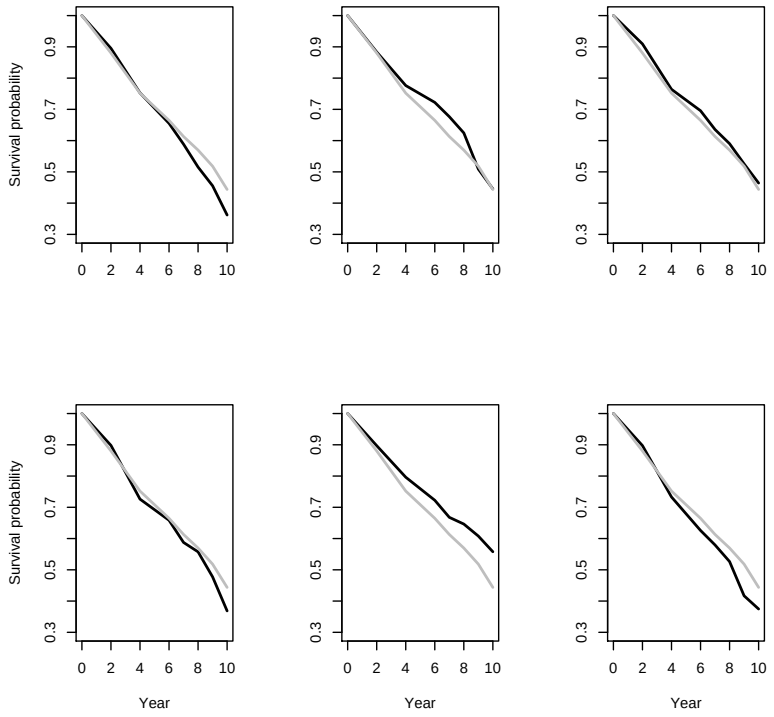

Fig. 3. The independent realizations randomly generated from the 0.95 HCDR (dark curve) with the Kaplan-Meier estimator (gray curve) for the survival probabilities at Years 2, 4, 6, 7, 8, 9 and 10 based on the Mayo primary biliary cirrhosis mortality data.

temporal profiles of members in the optimal region are quite similar except for their right tail parts, which is likely due to relatively low event rates and small numbers of patients in the risk sets beyond seven study years.

We conducted a small simulation study to compare the performance of the proposed HCDR and conventional confidence regions. The simulation study was designed to mimic the underlying stochastic mechanism generating the primary biliary cirrhosis liver data. Specifically, the survival times were generated from a two-parameter Weibull distribution fitted by the Mayo liver study data via the parametric maximum likelihood. The censoring times were generated with the observed Kaplan-Meier estimator constructed from the liver data for the censoring distribution. Here, the sample size $n = 400$. With each simulated data set, we constructed the 0.95 HCDR and the conventional confidence region for the survival function at Years 2, 4, 6, 7, 8, 9 and 10. With 2000 replications, the empirical coverage levels for the 0.95 HCDR and its conventional counterpart are 0.954 and 0.964, respectively. The ratio of the average volumes (conventional/HCDR) is about 34, indicating that on average the new proposal is much smaller than the conventional region.

4. Remarks

For cases with multiple parameters involved, one usually constructs marginal confidence or Bayesian credible intervals for each parameter by treating others as nuisance parameters. In fact, many modern novel statistical inference procedures were developed mainly for eliminating "nuisance parameters" effectively. However, oftentimes it is important and interesting to know the relationships among these parameters and their overall

profile such as the temporal pattern of the survival function discussed in this paper. An obstacle of making efficient joint inference is the difficulty of displaying or visualizing a possibly odd shaped confidence or credible region for a high dimensional vector of parameters. On the other hand, with modern numerical and computing techniques, members of such a joint region can be generated efficiently and displayed interactively as we did in the article.

For the case with a moderate sample size, different transformations $\boldsymbol{\omega}_0 = g(\boldsymbol{\theta}_0)$ may create quite different confidence density functions of $\boldsymbol{\theta}_0$. Moreover, the adequacy of the highest confidence density region estimation procedure depends on the accuracy of the normal approximation to the distribution of $(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}_0)$. It would be interesting and challenging to investigate how to choose a transformation which produces the "optimal" highest confidence density regions with correct coverage levels.

It is not clear how to generalize the HCDR estimation proposal to the case when the dimension of $\boldsymbol{\theta}_0$ is infinity, for example, the parameter of interest is the entire survival function. Further theoretical development along this line is needed.

Appendix

Justification for $\hat{P}_\alpha$ being an asymptotic $(1-\alpha)$ confidence region

To show that $\mathrm{pr}(\boldsymbol{\theta}_0 \in \hat{P}_\alpha) \to 1-\alpha$ as $n \to \infty$, we note that

$$\mathrm{pr}(\boldsymbol{\theta}_0 \in \hat{P}_\alpha) = \mathrm{pr}(\hat{Z}_{\boldsymbol{\theta}_0} \in \hat{S}_\alpha) = \mathrm{pr}(\varphi_{n\hat{\Sigma}}(\hat{Z}_{\boldsymbol{\theta}_0})\jmath\{g(\boldsymbol{\theta}_0)\} \geq n^{1/2}d),$$

where

$$\hat{S}_\alpha = \{\mathbf{z} : \varphi_{n\hat{\Sigma}}(z)\jmath\{g(\hat{\boldsymbol{\theta}}) + n^{-1/2}\mathbf{z}\} \geq n^{1/2}d\},$$

$\hat{Z}_{\boldsymbol{\theta}} = n^{1/2}\{g(\boldsymbol{\theta}) - g(\hat{\boldsymbol{\theta}})\}$, $\varphi_\Gamma(\cdot)$ is the density function of $\mathrm{MN}(0,\Gamma)$, $\jmath(\boldsymbol{\omega}) = J\{g^{-1}(\boldsymbol{\omega})\}$, and $J(\boldsymbol{\theta})$ is the determinant of the matrix of partial derivatives of $g(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. It follows from the convergence of $n\hat{\Sigma} \xrightarrow{\mathcal{P}} \Sigma_0$, a deterministic matrix, and the consistency of $\hat{\boldsymbol{\theta}}$ that $n^{1/2}d \to d_0$ in probability, where d_0 is the solution to

$$\mathrm{pr}\{\varphi_{\Sigma_0}(Z)\jmath\{g(\boldsymbol{\theta}_0)\} \geq d_0\} = 1-\alpha,$$

where the probability is with respect to Z and Z is $\mathrm{MN}(0,\Sigma_0)$. This, coupled with the convergence of $\hat{Z}_{\boldsymbol{\theta}_0} \xrightarrow{\mathcal{D}} \mathrm{MN}(0,\Sigma_0)$, implies that

$$\begin{aligned}\mathrm{pr}(\hat{Z}_{\boldsymbol{\theta}_0} \in \hat{S}_\alpha) &= \mathrm{pr}(\varphi_{\Sigma_0}(\hat{Z}_{\boldsymbol{\theta}_0})\jmath\{g(\boldsymbol{\theta}_0)\} \geq d_0) + o(1)\\ &= \mathrm{pr}(\varphi_{\Sigma_0}(Z)\jmath\{g(\boldsymbol{\theta}_0)\} \geq d_0) + o(1) = 1-\alpha+o(1)\end{aligned}$$

References

Box, G. E. P. & Tiao, G. C. (1992). *Bayesian Inference in Statistical Analysis*. John Wiley & Sons.

Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. *Biometrika* **80**, 3–26.

Efron, B. (1998). R. A. Fisher in the 21st century (invited paper presented at the 1996 R. A. Fisher lecture). *Statistical Science* **13**, 95–122.

Fleming, T. R. & Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons.

Fraser, D. A. S. (1991). Statistical inference: Likelihood to significance. *Journal of the American Statistical Association* **86**, 258–265.

Fraser, D. A. S. (1996). Comment on "Pivotal inference and the fiducial argument" (95V63 p309-323). *International Statistical Review* **64**, 231–235.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. John Wiley & Sons.

Lehmann, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association* **88**, 1242–1249.

Lin, D. Y., Fleming, T. R. & Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73–81.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag Inc.

Parzen, M. I., Wei, L. J. & Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–350.

Schweder, T. & Hjort, N. L. (2002). Confidence and likelihood. *Scandinavian Journal of Statistics* **29**, 309–332.

Schweder, T. & Hjort, N. L. (2003). Frequenstist analogues of priors and posteriors. *Econometrics and the Philosophy of Economics* , Princeton University Press.

Singh, K., Xie, M. & Strawderman, W. E. (2007). Confidence distribution (cd)-distribution estimator of a parameter. *Complex Datasets and Inverse Problems: Tomography, Networks and Beyond*. **IMS Lecture Notes-Monograph Series 54**, Institue of Mathematical Statistics.