Supplementary Materials for Pointwise Confidence Intervals for a Survival Distribution with Small Samples or Heavy Censoring

MICHAEL P. FAY, ERICA H. BRITTAIN, MICHAEL A. PROSCHAN
National Institute of Allergy and Infectious Diseases,
Bethesda, MD, 20892-7630, USA
correspondence to M.P. Fay: mfay@niaid.nih.gov

Summary

This supplement provides additional mathematical details and simulations for the paper.

A Distribution of $S(T_j)$ with Progressive Type II Censoring

Let $Y_i = Y(T_i)$ be the number of patients at risk just prior to the ith event time T_i. With progressive type II censoring, $Y_i - Y_{i+1} - 1$ patients are selected randomly to be censored at time T_i^+ (infinitesimally after time T_i) among the $Y_i - 1$ people who did not die and were not censored by time T_i. We compute the conditional distribution of the time of the next event, T_{i+1}, given $\mathbf{T}_i = (T_1, \ldots, T_i)$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)$. (Note: we could condition on only the Ys up to the ith event, and this gives the same answer.) Each of the remaining Y_{i+1} patients just after time T_i has amassed followup time T_i without an event. Therefore, the conditional probability that $T_{i+1} > t_{i+1}$, given $\mathbf{T}_i$, and $\mathbf{Y}$ is $[\{S(t_{i+1})\}/\{S(T_i)\}]^{Y_{i+1}}$. But $\{T_{i+1} > t_{i+1}\} = \{S(T_{i+1}) < S(t_{i+1})\}$, so

$$\begin{aligned}
\Pr\{S(T_{i+1}) < S(t_{i+1}) \,|\, \mathbf{T}_i, \mathbf{Y}\} &= \left[\frac{S(t_{i+1})}{S(T_i)}\right]^{Y_{i+1}} \\
\Pr\{S(T_{i+1}) < u_1 \,|\, \mathbf{T}_i, \mathbf{Y}\} &= \left[\frac{u_1}{S(T_i)}\right]^{Y_{i+1}} \\
\Pr\{S(T_{i+1}) < u_2 S(T_i) \,|\, \mathbf{T}_i, \mathbf{Y}\} &= u_2^{Y_{i+1}} \\
\Pr\left\{\frac{S(T_{i+1})}{S(T_i)} < u_2 \,|\, \mathbf{T}_i, \mathbf{Y}\right\} &= u_2^{Y_{i+1}}.
\end{aligned}$$

This shows that the conditional distribution of $S(T_{i+1})/S(T_i)$ given $\mathbf{T}_i$ and $\mathbf{Y}$ is beta$(Y_{i+1}, 1)$. Because this distribution is the same for all values of $\mathbf{T}_i$, $S(T_{i+1})/S(T_i)$ is conditionally independent of $\mathbf{T}_i$, given $\mathbf{Y}$. This holds for all i, so conditioned on $\mathbf{Y}$, $S(T_2)/S(T_1)$ is independent of $S(T_1)$; $S(T_3)/S(T_2)$ is independent of $\{S(T_1), S(T_2)/S(T_1)\}$, etc. It follows that, conditioned on $\mathbf{Y}$,

$$S(T_i) = S(T_1)\frac{S(T_2)}{S(T_1)} \cdots \frac{S(T_i)}{S(T_{i-1})}$$

is the product of independent betas with respective parameters $(Y_1 = n, 1), (Y_2, 1), \ldots (Y_i, 1)$.

B Proof of Theorem 1

The confidence interval method of this paper can be viewed as a special case of the following technique (formally shown in Casella and Berger (2002), Theorem 9.2.14). Suppose that W has distribution $F_\theta(w)$, where F is a decreasing continuous function of θ. Determine $\theta_L = \theta_L(w_{\text{obs}})$ as the solution to $\Pr(W \geq w_{\text{obs}} \,|\, \theta_L) = \alpha/2$, where w_{obs} is the observed value of W. It is helpful to regard this technique in the following way. Fix w_{obs} and determine θ_L such that if it is the true value of θ and we repeat the experiment by drawing a new value W, the new value has probability $\alpha/2$ of being $\leq w_{\text{obs}}$. Notice that the event $\theta < \theta_L(w_{\text{obs}})$ implies that $\Pr(W \geq w_{\text{obs}} \,|\, \theta) \leq \alpha/2$. By definition, the probability that this occurs under true parameter value θ is $\leq \alpha/2$. Similarly, define θ_U to be the solution to $\Pr(W \leq w_{\text{obs}} \,|\, \theta_U) = \alpha/2$.

By the same reasoning as above, The event $\theta > \theta_U$ has probability $\leq \alpha/2$. Therefore, the probability that the random interval (θ_L, θ_U) covers θ is at least $1-\alpha$.

Now apply the above technique to the estimated survival probability $\hat{S}(t) = S_{KM}\{t; \mathbf{Z}(t)\}$ using the Kaplan-Meier estimator with progressive type 2 censoring. Fix the observed number of events, i, by time t, and the observed censoring pattern $\mathbf{Y} = \mathbf{y}_{obs}$. Imagine repeating the experiment with the proviso that the censoring pattern of the replicate is the same as in the original experiment up to the number of events observed in the replicate. Let $\hat{S}_{\text{obs}}(t)$ be the estimated survival at time t in the original experiment and $\hat{S}(t)$ be the corresponding survival estimate in the replicate experiment. Then $\hat{S}(t) \leq \hat{S}_{\text{obs}}(t)$ if and only if at least i events in the replicate experiment occur by time t. Equivalently, $T_i \leq t$, where T_i is the time of the ith observed event in the replicate experiment. This in turn is equivalent to $S(T_i) \geq S(t)$. This is equivalent to

$$S(T_1)\frac{S(T_2)}{S(T_1)}\cdots\frac{S(T_i)}{S(T_{i-1})} \geq S(t). \tag{B.1}$$

Likewise, $\hat{S}(t) \geq \hat{S}_{\text{obs}}(t)$ if and only if at most i events in the replicate experiment occur by time t, which is equivalent to $T_{i+1} > t$, which is equivalent to

$$S(T_1)\frac{S(T_2)}{S(T_1)}\cdots\frac{S(T_{i+1})}{S(T_i)} < S(t). \tag{B.2}$$

Therefore, we can determine a $1-\alpha$ confidence interval for $S(t)$ that has the correct conditional coverage probability given the observed censoring pattern $\mathbf{Y} = \mathbf{y}_{\text{obs}}$ as follows. The upper limit is the upper $\alpha/2$ quantile (i.e., the value x exceeded with probability $\alpha/2$) of the distribution of the product of the i independent betas on the left side of (B.1). The lower limit of a $1-\alpha$ confidence interval for $S(t)$ is the lower $\alpha/2$ quantile of the distribution of the product of the $i+1$ independent betas on the left side of (B.2). Because this confidence interval has the correct conditional coverage probability given that $\mathbf{Y} = \mathbf{y}_{\text{obs}}$, it has the correct unconditional coverage probability also.

C Proof of Theorem 2

The BPCP for $S(t)$ with $t = T_j$ uses a beta product random variable with parameters $\mathbf{Y}(t) = [y_1, y_2, \ldots, y_j]$ and $\mathbf{1} = [1, \ldots, 1]$, where $y_i = Y(T_i)$. Let

$$V_j \quad = \quad -\log(BP\{\mathbf{Y}(t), \mathbf{1}\}) = \sum_{i=1}^{j} -\log(B(y_i, 1))$$

Note that if $B_i \sim beta(y_i, 1)$ with y_i fixed then $W_i^* = -\log(B_i)$ is exponential with mean $1/y_i$. Therefore,

$$V_j \quad = \quad \sum_{i=1}^{j} \frac{1}{y_i} W_i,$$

where W_i is exponential with mean 1.

Consider first the case where G, the censoring distribution is fixed, so that the number of observed failures before t goes to infinity. Since $E(W_i) = Var(W_i) = 1$ and

$$\frac{\max_{i=1,\ldots,j} 1/y_i^2}{\sum_{i=1}^{j} 1/y_i^2} \leq \frac{1/y_j^2}{j/n^2} = (1/j)(n/y_j)^2 \to 0,$$

we can use the Liapounov central limit theorem to see that

$$Z_j = \frac{\sum \frac{W_i - 1}{y_i}}{\sqrt{\sum 1/y_i^2}} = \frac{V_j - \hat{A}(t)}{\hat{\sigma}(t)} \xrightarrow{L} N(0,1).$$

where $N(0,1)$ denotes the standard normal distribution function (see e.g., Lehmann (1999), p. 102). Treating Z_j as asymptotically standard normal we get

$$\begin{aligned} 1-\alpha &\approx Pr\left[-\Phi^{-1}(1-\alpha/2) \le \frac{V_j - \hat{A}(t)}{\hat{\sigma}(t)} \le \Phi^{-1}(1-\alpha/2)\right] \\ &= Pr\left[\hat{A}(t) - \Phi^{-1}(1-\alpha/2)\hat{\sigma}(t) \le -\log\left\{BP\left(\mathbf{Y},1\right)\right\} \le \hat{A}(t) + \Phi^{-1}(1-\alpha/2)\hat{\sigma}(t)\right] \\ &= Pr\left[\exp\left\{-\hat{A}(t) - \Phi^{-1}(1-\alpha/2)\hat{\sigma}(t)\right\} \le BP\left(\mathbf{Y},1\right) \le \exp\left\{-\hat{A}(t) + \Phi^{-1}(1-\alpha/2)\hat{\sigma}(t)\right\}\right]. \end{aligned}$$

We see that the BPCP asymptotically matches (3). The asymptotic equivalence of (3.3) and (3.4) has been shown (see e.g., Andersen *and others* (1993); Aalen *and others* (2008)).

D Alternative Notation for Grouped Data

To write the collapsed beta product we introduce new notation. Let $N(t)$ be the number who are known to have failed at or before t. For $t \in (g_{i-1}, g_i]$, let $D(t) = N(g_i) - N(g_{i-1})$, $Y_g^+(t) = Y(g_i)$ and $Y_g^-(t) = Y(g_{i-1})$. Suppose there are k_g assessment times, g_i, where $D(g_i) > 0$, and let the associated indices be $i_1, \ldots, i_{k_g}$. Let

$$\begin{aligned} \mathbf{Y}_g(t) &= [Y_g^-(g_{i_1}), Y_g^-(g_{i_2}), \ldots, Y_g^-(g_{i_h}), Y_g^+(t)] \quad \text{if } t \in (g_{i_h-1}, g_{i_{h+1}-1}], \quad \text{and} \\ \mathbf{D}_a(t) &= [D(g_{i_1}), D(g_{i_2}), \ldots, D(g_{i_h}), a] \quad \text{if } t \in (g_{i_h-1}, g_{i_{h+1}-1}] \end{aligned}$$

and let $\mathbf{1}_g(t)$ be a vector of 1s the same length as $\mathbf{Y}_g(t)$. Then repeatedly using (2.1), we can show that for $t \in \{g_1, \ldots, g_m\}$,

$$BP\left\{\mathbf{Y}(t), \mathbf{1}_a(t)\right\} \sim BP\left\{\mathbf{Y}_g(t) - \mathbf{D}_a(t) + \mathbf{1}_g(t), \mathbf{D}_a(t)\right\}.$$

So an equivalent definition of the BPCP for grouped data is: for $t \in (g_{i-1}, g_i]$,

$$\begin{aligned} L_t(\mathbf{Z}_g(t), 1-\alpha/2) &= Q\left\{\alpha/2, \mathbf{Y}_g(g_i) - \mathbf{D}_1(g_i) + \mathbf{1}_g(g_i), \mathbf{D}_1(g_i)\right\} \\ &\quad \textit{and} \\ U_t(\mathbf{Z}(t), 1-\alpha/2) &= Q\left\{1-\alpha/2, \mathbf{Y}_g(g_{i-1}) - \mathbf{D}_0(g_{i-1}) + \mathbf{1}_g(g_{i-1}), \mathbf{D}_0(g_{i-1})\right\}. \end{aligned}$$

E Simulation on Median Unbiased Estimator

Table E.1: 1000 times mean squared error of the survival estimates at the true values given by the top row.

	0.99	0.9	0.75	0.5	0.25	0.1	0.01
KML	0.39	3.66	7.61	11.00	9.45	6.11	0.30
KMM	0.39	3.66	7.61	11.00	9.43	5.73	0.80
KMH	0.39	3.66	7.61	11.00	9.41	5.60	2.87
$BPCP_{MM}$	0.44	3.49	7.41	10.66	8.89	4.37	1.49

F Simulation Modeled After Nash et al (2007) Data

Table F.2: Simulated Size, with the simulation modeled after Nash *and others* (2007) Data; $n = 34$ with X a mixture of two exponentials, one exponential with mean 0.227 sampled with probability p=0.187 and the second exponential with mean 22.44 sampled with probability 1-p=0.813; censoring, C, is uniform on $(2, 8)$. Simulation had 100,000 replications. Percent Error on Each Side of 95% Interval (nominal is 2.5%).

	$t=3$		$t=4$		$t=5$		$t=6$	
	$S(t)=0.71$		$S(t)=0.68$		$S(t)=0.65$		$S(t)=0.62$	
	$\hat{E}(N(t))=9.8$		$\hat{E}(N(t))=10.7$		$\hat{E}(N(t))=11.2$		$\hat{E}(N(t))=11.6$	
	$\hat{E}(Y(t))=20.1$		$\hat{E}(Y(t))=15.3$		$\hat{E}(Y(t))=11.0$		$\hat{E}(Y(t))=7.1$	
	low	high	low	high	low	high	low	high
Greenwood (log)	4.3	0.5	5.2	0.6	5.4	0.3	7.0	0.0
Modified Lower	4.3	0.5	4.1	0.6	2.7	0.3	2.5	0.0
Borkowf (log)	4.3	0.5	3.7	0.3	2.5	0.3	2.7	0.0
Borkwof (log, shrink)	4.0	0.5	2.3	0.3	1.7	0.3	1.9	0.0
Strawderman-Wells	3.4	1.9	2.0	1.6	2.6	1.6	4.5	1.7
Thomas-Grunkemeier	1.4	2.2	2.0	2.2	2.4	1.8	3.8	2.4
Constrained Beta	3.4	2.0	2.0	2.0	2.5	1.8	3.8	2.5
Bootstrap	4.0	2.7	3.2	2.6	3.4	2.4	4.2	2.3
Constrained Bootstrap	2.6	2.3	2.0	2.7	2.0	2.1	3.8	2.1
Binomial (Censor>t)	1.2	1.4	1.4	1.7	1.2	1.2	1.0	1.6
BPCP (MM)	1.2	1.9	0.5	1.4	0.4	1.4	0.3	1.1
BPCP (MC)	1.1	1.9	0.5	1.4	0.4	1.5	0.2	1.6

References

AALEN, O.O., BORGAN, O. AND GJESSING, H.K. (2008). *Survival and Event History Analysis*. New York: Springer.

ANDERSEN, P.K., BORGAN, O., GILL, R.D. AND KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

CASELLA, GEORGE AND BERGER, ROGER L. (2002). *Statistical Inference, second edition*. Duxbury Press.

LEHMANN, E.L. (1999). *Elements of Large Sample Theory*. Springer.

NASH, R.A., MCSWEENEY, P.A., CROFFORD, L.J., ABIDI, M., CHEN, C.S., GODWIN, J.D., GOOLEY, T.A., HOLMBERG, L., HENSTORF, G., LEMAISTRE, C.F. *and others*. (2007). High-dose immunosuppressive therapy and autologous hematopoietic cell transplantation for severe systemic sclerosis: long-term follow-up of the us multicenter pilot study. *Blood* **110**(4), 1388.