OptBand: (Approximately) minimal-area confidence bands for failure time data

Tom Chen, Sam Tracy and Hajime Uno

2017-10-23

Abstract

Classical simultaneous confidence bands for survival functions (i.e. Hall-Wellner and Equal Precision) are derived from transformations of the asymptotic Brownian nature of the Nelson-Aalen or Kaplan-Meier estimators. Due to the properties of Brownian motion, a theoretical derivation of the highest confidence density region cannot be obtained in closed form. Instead, we provide confidence bands derived from a related optimization problem with local times. These bands can be applied for the one-sample problem regarding both cumulative hazard and survival functions, and the two-sample problem regarding cumulative hazard only. The finite-sample performance of the proposed method is assessed by Monte Carlo simulation studies. The proposed bands are applied to clinical trial data to assess the cardiovascular efficacy and safety of saxaglipitin, a DPP-4 inhibitor, when added to standard of care in patients with type 2 diabetes mellitus.

Introduction

In biomedical studies, it is often of interest to construct confidence bands (CB) around the survival function on a given time interval. The two classical simultaneous CB's based on the Kaplan-Meier estimator were originally developed by Hall and Wellner (1984) and Nair et al (1984), which are now called the Hall-Wellner (HW) and Equal Precision (EP) bands. Both of these aforementioned papers make use of some transformation of the estimated survival function weakly converging to a Brownian motion, and then sample from the transformed distribution to obtain critical values for the coverage levels. Hollander et al (1997) introduced simultaneous CB's based on the likelihood ratio approach for confidence intervals first introduced by Thomas and Grunkemeier (1975). Parzen et al (1994) introduced the resampling-perturbation method as a generalized approach in calculating critical values for transformed distributions, and notably applied this in the two-sample problem Parzen et al (1997). Finally, Tian et al (2011) derived bands that targeted the highest density confidence region (HCDR), but their approach requires the standard specifications and tuning procedures that accompany a Markov chain Monte Carlo.

The approach taken in this paper is based upon the analytic, approximately minimal-area (AMA) confidence bands around a Brownian motion derived in Kendall et al (2007). If $W(\sigma^2(t))$ is a mean-zero Brownian motion with strictly-increasing variance process $\sigma^2(t)$, then the AMA bands are $(-u^*_\kappa(t), u^*_\kappa(t))$, where $u^*_\kappa(t) = \psi(\kappa\sigma^2(t))\sqrt{\sigma^2(t)}$, κ is a variational parameter related to desired coverage level γ, $\psi(x) = \sqrt{-\mathcal{W}_{-1}(-x^2)}\mathbb{I}(x \leq e^{-1/2})$, $\mathcal{W}_{-1}$ is the Lambert W function on the -1 branch, and $\mathbb{I}$ is the indicator function (see Supplementary Materials A for more details). Another, more natural interpretation for κ is the required critical value in order to achieve a coverage level of γ; therefore, one adjusts κ such that $|W(\sigma^2(t))| \leq u^*_\kappa(t)$ for all $t \in [t_L, t_U]$ occurs with probability γ. A general adjustment process for κ was not explored by Kendall et al (2007), and hence the AMA method was quite limited in use. We propose a functional approximation in calculating κ, and use this method to design AMA confidence bands for the cumulative hazard, survival, and difference of cumulative hazard functions. The derived equations for this method are shown in the Appendix at the end of this paper. The performance of the AMA bands against that of Hall-Wellner and Equal Precision are evaluated and illustrated by an application to the SAVOR-TIMI 53 clinical trial data.

Methods

For simulation study, we compared the AMA survival band against the HW and EP bands. We generated 10,000 data sets defined by $X_i \sim \text{Exp}(1)$ and $C_i \sim \text{Exp}(0), \text{Exp}(0.25)$ to represent 0% and 20% censoring, respectively, at sample sizes of $n = 500, 1000$. The effect of a restricted time interval was evaluated by considering the four possible restricted quantile intervals (0.01, 0.99), (0.2, 0.8), (0.01, 0.8), (0.2, 0.99) in addition to the entirety of the observed times. I.e. we trimmed 1% and/or 20% of the starting and ending times. To save space, the areas are normalized by the HW are under the same conditions (e.g. area of EP region / area of HW area). The results are shown in Tables 1 and 2 and the simulation file is included with the online version of this paper.

The SAVOR-TIMI 53 was a double-blind, placebo-controlled trial examining the safety and efficacy of the DPP-4 inhibitor, saxagli3ptin published by Scirica et al (2013). A total of 16,492 patients with type 2 diabetes mellitus and a history of, or at risk of, cardiovascular events were randomized 1:1 to saxagliptin or placebo with mean follow-up of 2.1 years. The primary endpoint was a composite of CV death, nonfatal MI, or nonfatal ischemic stroke. The secondary endpoint was the primary endpoint plus hospitalization for heart failure, coronary revascularization, or unstable angina. We constructed the HW, EP, and AMA bands for time to endpoint on both saxagliptin and placebo groups in an unadjusted analysis, as displayed in Figure 1. Furthermore, we constructed AMA bands for the unadjusted difference in cumulative hazards, as seen in Figure 2.

Results

Table 1: (coverage, area) with $n = 500$ and 0% censoring

α	Method	Restriction None	(0.01,0.99)	(0.2,0.8)	(0.01,0.8)	(0.2,0.99)
90%	HW	0.93, 1.00	0.92, 1.00	0.92, 1.00	0.92, 1.00	0.92, 1.00
	EP	0.80, 0.62	0.90, 0.74	0.92, 1.00	0.91, 1.03	0.91, 0.71
	AMA	0.83, 0.61	0.89, 0.71	0.91, 0.97	0.92, 0.96	0.90, 0.70
95%	HW	0.96, 1.00	0.96, 1.00	0.97, 1.00	0.96, 1.00	0.96, 1.00
	EP	0.86, 0.60	0.93, 0.72	0.96, 0.99	0.95, 1.01	0.95, 0.70
	AMA	0.88, 0.59	0.94, 0.70	0.96, 0.97	0.96, 0.95	0.94, 0.69
99%	HW	0.99, 1.00	0.99, 1.00	0.99, 1.00	0.99, 1.00	0.99, 1.00
	EP	0.92, 0.58	0.98, 0.70	0.99, 0.99	0.99, 0.99	0.98, 0.68
	AMA	0.93, 0.56	0.98, 0.67	0.99, 0.97	0.99, 0.94	0.98, 0.67

As seen in the tables, and also originally noted by Hall and Wellner, the HW band is overly conservative with a significantly larger total regional area within the confidence band than the other two methods. EP and AMA bands do not achieve the nominal level of coverage in the no restriction case. Nair noted that EP bands are valid only in a restricted range, and we make the same observations as well for AMA band. Considering the other four restriction cases, we now see that all three bands attain the targeted coverage levels for the $(n = 500, n = 100) \times (0\% \text{ censor}, 20\% \text{ censor})$ combinations. Among the restriction cases, in terms of area between bands, EP outperforms HW for all except the $(0.01, 0.8)$ truncation case. AMA outperforms both EP and HW bands for all restriction cases. Notably, AMA and EP outperform HW significantly when there is little (1%) restriction at the upper time point of the interval. When we consider significant (20%) restriction

on the upper time point, HW becomes far more competitive relative to AMA and EP, although AMA still outperforms with 3% – 6% savings in area compared to HW. Further simulations with 50% censoring showed that AMA retains the nominal level of coverage, while EP suffers approximately 0.05 below the nominal coverage levels of 0.90, 0.95, and 0.99.

Table 2: (coverage, area) with $n = 500$ and 20% censoring

α	Method	Restriction				
		None	(0.01,0.99)	(0.2,0.8)	(0.01,0.8)	(0.2,0.99)
90%	HW	0.92, 1.00	0.92, 1.00	0.92, 1.00	0.92, 1.00	0.92, 1.00
	EP	0.82, 0.56	0.90, 0.73	0.92, 1.01	0.90, 1.03	0.91, 0.70
	AMA	0.86, 0.56	0.91, 0.71	0.91, 0.99	0.91, 0.97	0.91, 0.70
95%	HW	0.96, 1.00	0.96, 1.00	0.96, 1.00	0.96, 1.00	0.96, 1.00
	EP	0.87, 0.55	0.94, 0.71	0.96, 1.01	0.95, 1.02	0.95, 0.69
	AMA	0.90, 0.54	0.94, 0.69	0.96, 0.98	0.95, 0.96	0.95, 0.68
99%	HW	0.99, 1.00	0.99, 1.00	0.99, 1.00	0.99, 1.00	0.99, 1.00
	EP	0.93, 0.53	0.98, 0.68	0.99, 1.00	0.98, 0.99	0.99, 0.67
	AMA	0.94, 0.52	0.98, 0.67	0.99, 0.98	0.99, 0.95	0.99, 0.66

Although the AMA band clearly outperforms the HW band, the comparison to the EP band is less obvious. In fact, the EP band actually achieves a smaller area than either the AMA or HW bands, outperforming the AMA band by 3.7% and 3.0%, relative to the HW band for the placebo and treatment groups, respectively. However, this reduction in area comes at a significant price, as the clinical trial data was heavily censored and we showed in simulation that the EP band fails to achieve the targeted coverage under much censoring. We note that the estimated bands for the cumulative hazard difference contain the null value of 0 and confirm the results in the original paper by Scirica et al (2013).

Discussion

In this article, we presented analytical confidence bands for the one-sample cumulative hazard, two-sample cumulative hazard, and one-sample survival functions that approximately minimize the area between bands, and hence approximates the HCDR. Classic bands such as HW and EP simply require a table of critical values and their extraction and hence are fast to compute. The method proposed by Tian et al (2011), while exactly targeting the HCDR, requires conducting a MCMC and is more time consuming. AMA bands strike a delicate balance between simplicity with an analytical solution while also roughly targeting the HCDR.

AMA bands also present a very interpretable result. Intuitively, one would expect a confidence band to take the form of an estimate $\widehat{\theta}(t)$ plus or minus some variation $\alpha(t) \cdot \text{se}(\widehat{\theta}(t))$. The AMA bands take exactly that form, with $\alpha(\cdot)$ encompassing the $\psi(\cdot)$ function within. Hence, the $\psi(\cdot)$ function can be viewed as the appropriate weighting function that shapes the bands to have minimal area. These methods have been published for public use in the R software package `optband` on CRAN. We recommend the use of AMA bands with at least minor restriction of the event times, as used in the simulations.

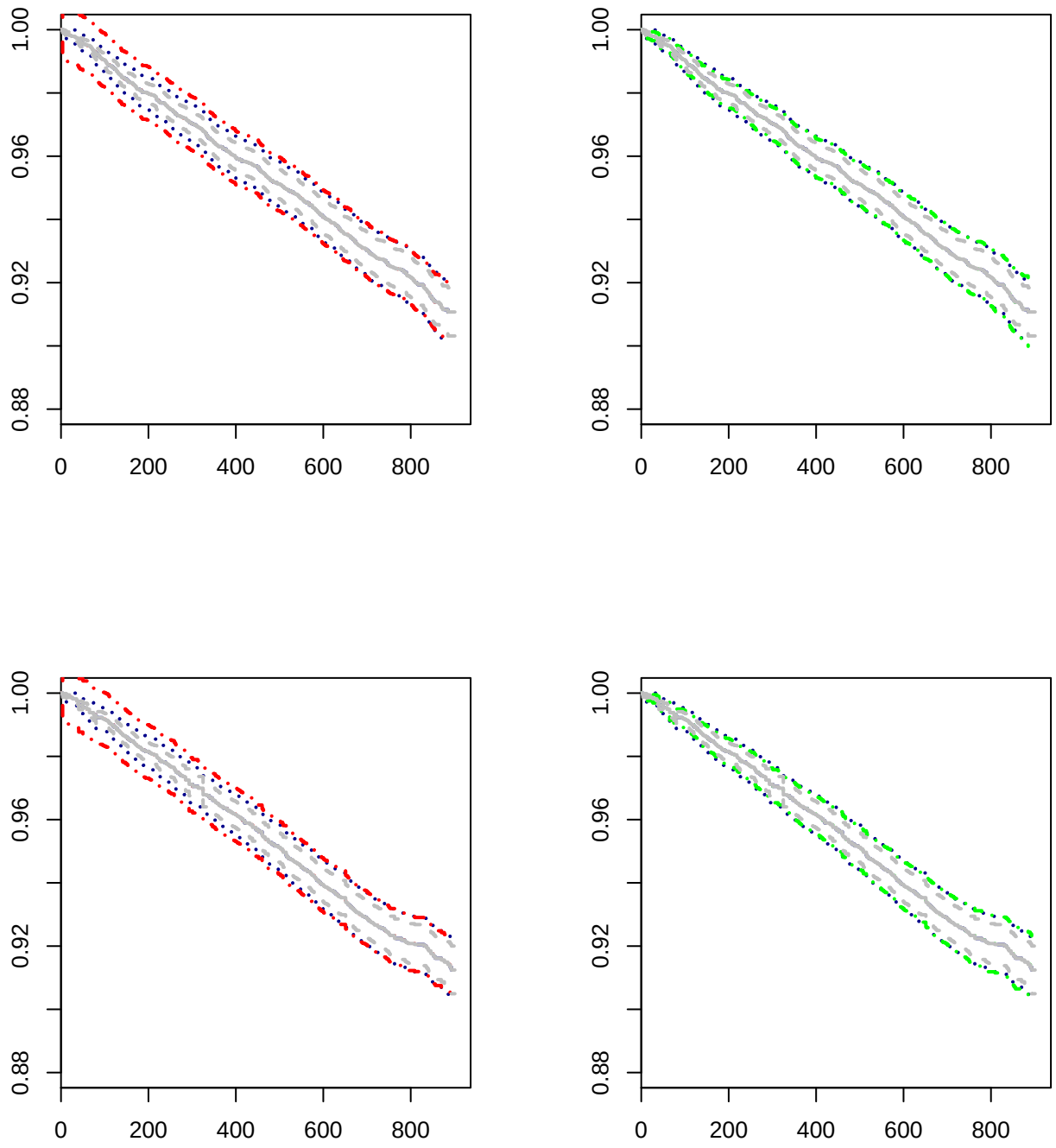

Figure 1: confidence band comparison for the treatment arm in the first row and placebo arm in the second row of the SAVOR clinical trial. The estimated survival curve is in grey, the fitted AMA band in blue, the EP band in green and the HW band in red.

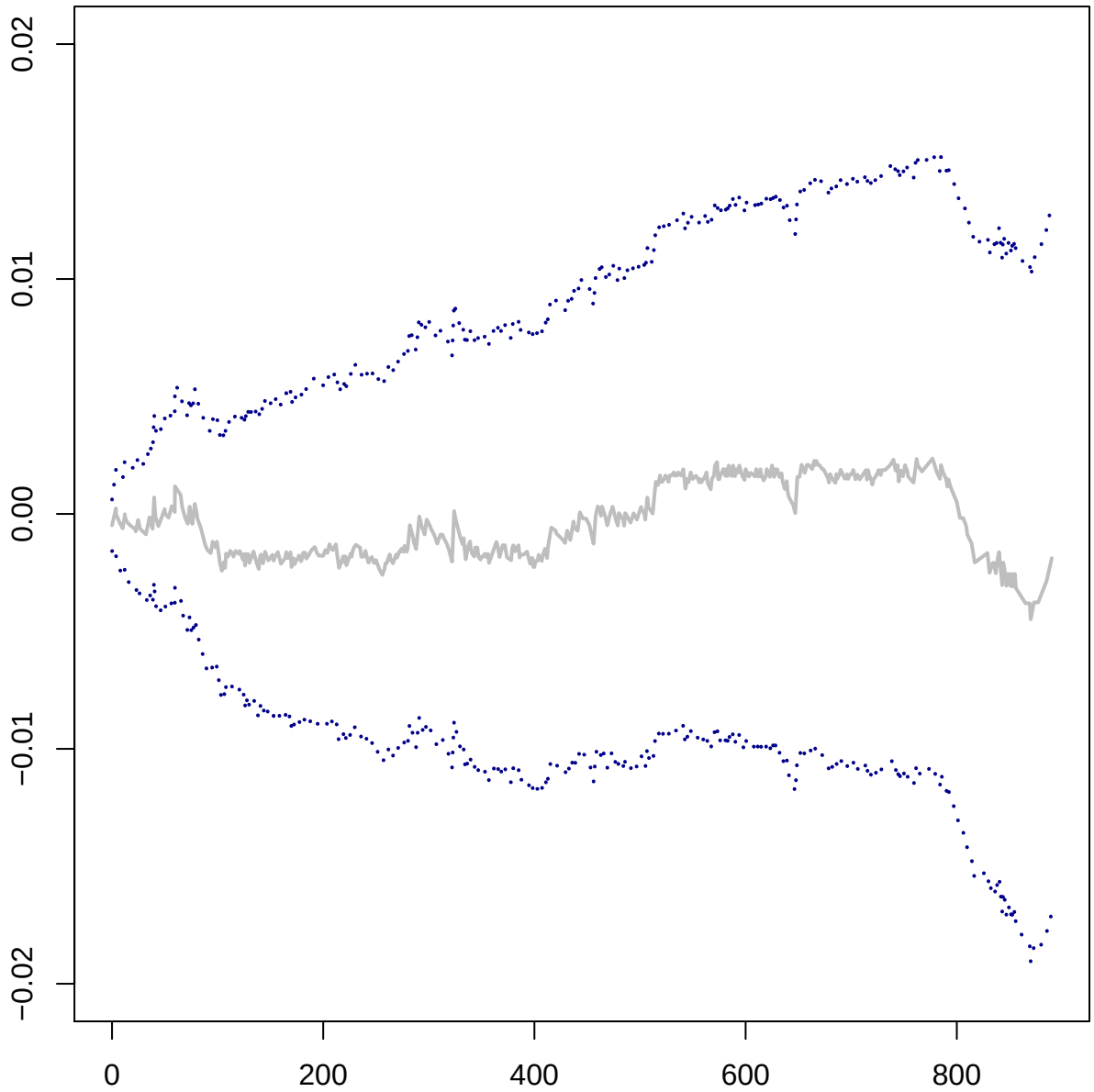

Figure 2: the cumulative hazard difference between the placebo arms of the SAVOR clinical trial. The estimated cumulative hazard difference is in grey and the fitted AMA band in blue.

References

Georg M. Goerg (2016). LambertW: An R package for Lambert W x F Random Variables. R package version 0.6.4.

Georg M. Goerg (2011): Lambert W random variables - a new family of generalized skewed distributions with applications to risk estimation. Annals of Applied Statistics 3(5). 2197-2230.

Georg M. Goerg (2014): The Lambert Way to Gaussianize heavy-tailed data with the inverse of Tukey's h transformation as a special case. The Scientific World Journal.

Hall, Wendy J., and Jon A. Wellner. Confidence bands for a survival curve from censored data. Biometrika 67.1 (1980): 133-143.

Kendall, W.S., Marin, J.M. and Robert, C.P. (2007). Confidence bands for Brownian motion and applications to Monte Carlo simulation. Statistics and Computing, 17(1), pp.1-10.

Klein, Moeschberger with modifications by Jun Yan (2012). KMsurv: Data sets from Klein and Moeschberger (1997), Survival Analysis. R package version 0.1-5}. https://CRAN.R-project.org/package=KMsurv.

Myles Hollander , Ian W. McKeague & Ian W. McKeague. Likelihood Ratio-Based Confidence Bands for Survival Functions. Journal of the American Statistical Association 92:437 (1997): 215-226

Nair, Vijayan N. Confidence bands for survival functions with censored data: a comparative study. Technometrics 26.3 (1984): 265-275.

Parzen, M. I., L. J. Wei, and Z. Ying. A resampling method based on pivotal estimating functions. Biometrika 81.2 (1994): 341-350.

Parzen, M. I., L. J. Wei, and Z. Ying. Simultaneous confidence intervals for the difference of two survival functions. Scandinavian Journal of Statistics 24.3 (1997): 309-314.

Ralf Strobl (2009). km.ci: Confidence intervals for the Kaplan-Meier estimator. R package version 0.5-2. https://CRAN.R-project.org/package=km.ci

Scirica, Benjamin M, et al. Saxagliptin and Cardiovascular Outcomes in Patients with Type 2 Diabetes Mellitus. New England Journal of Medicine 369 (2013): 1317-1326.

Terry M. Therneau and Patricia M. Grambsch (2000). Modeling Survival Data: Extending the Cox Model. Springer, New York. ISBN 0-387-98784-3.

Therneau T (2015). A Package for Survival Analysis in S. version 2.38, <URL: https://CRAN.R-project.org/package=survival>.

Thomas, David R., and Gary L. Grunkemeier. Confidence interval estimation of survival probabilities for censored data. Journal of the American Statistical Association} 70.352 (1975): 865-871.

Tian, Lu, et al. The highest confidence density region and its usage for joint inferences about constrained parameters. Biometrics 67.2 (2011): 604-610.

Appendix

The adjustment process for κ would depend on t_L, t_U, γ, and $\sigma^2(t)$. Kendall et al. only performed a Monte Carlo simulation for the case $t_L = 0, t_U = 1, \gamma = 0.95, \sigma^2(t) = t$ (the standard Brownian motion on the unit interval) in deriving $\kappa = 0.105$. We provide a functional approximation that behaves very well.

First, we may perform a transformation to standardize the entire process. Without loss of generality, consider minimizing the interval around $W(\frac{\sigma^2(t)}{\sigma^2(t_U)})$ for $t \in [t_L, t_U]$, since $\frac{1}{\sqrt{\sigma^2(t_U)}} W(\sigma^2(t))$ is equal in law to $W(\frac{\sigma^2(t)}{\sigma^2(t_U)})$.

Then, the AMA solution is $u^*_\kappa(t) = \psi(\kappa \frac{\sigma^2(t)}{\sigma^2(t_U)}) \cdot$

$\sqrt{\frac{\sigma^2(t)}{\sigma^2(t_U)}}$, and hence we would need to adjust κ such that $|W(s)| \leq \psi(\kappa s)\sqrt{s}$ for all $s \in [\frac{\sigma^2(t_L)}{\sigma^2(t_U)}, 1]$ occurs with probability γ, which is now only a function of γ and $L = \frac{\sigma^2(t_L)}{\sigma^2(t_U)}$. The authors have done so through simulation, and derived the empirical relationship

$$\gamma = 1 + (a + bL)\kappa + a\kappa^2 \implies \kappa = -\frac{(a + bL) + \sqrt{(a + bL)^2 - 4a(1 - \gamma)}}{2a} \tag{1}$$

where $a = -0.4272, b = 0.2848$ are empirically calculated. The resulting γ from the simulation ranges from 0.871 to 0.999, encompassing accurate interpolation for the clinically-relevant coverage levels of 0.90 to 0.99. The simulation was performed in R and the file and resulting coverage table are shared with the online version of this paper. This coverage table is displayed as a heatmap of coverage values in Figure A1.

Let $X_i \sim F$ and $C_i \sim G$ denote the random variables representing the failure times and censoring times, respectively. We only observe $T_i = \min(X_i, C_i)$ and $\Delta_i = \mathbb{I}\{X_i \leq C_i\}$. Given observed values $(t_i, \delta_i)_{i=1}^n$, let $d_i = \sum_{j:t_j=t_i} t_j\delta_j$ be the number of failures at t_i, and $n_i = |\{j : t_j \geq t_i\}|$ be the number of individuals at risk at time t_i, and $\widehat{H}(t)$ be either the Nelson-Aalen or Kaplan-Meier estimator of the true cumulative hazard $H(t)$. Restricting our interval of interest to $[t_L, t_U]$, standard asymptotic results show that $\sqrt{n}\frac{\widehat{H}(t)-H(t)}{\sigma(t_U)}$ weakly converges to a $W\left(\frac{\sigma^2(t)}{\sigma^2(t_U)}\right)$, where

$$\sigma^2(t) = \int_0^t \frac{\lambda(s)}{(1 - F(s))(1 - G(s))}ds$$

i.e. the variance of the cumulative hazard, which is clearly strictly increasing. Constructing the AMA bands with $W\left(\frac{\sigma^2(t)}{\sigma^2(t_U)}\right)$ yields

$$\left|\sqrt{n}\frac{\widehat{H}(t) - H(t)}{\sigma(t_U)}\right| \leq \psi\left(\kappa\frac{\sigma^2(t)}{\sigma^2(t_U)}\right)\sqrt{\frac{\sigma^2(t)}{\sigma^2(t_U)}}$$

where κ is calculated from Eq 1 to yield a coverage level of γ. Since we do not observe the true variance $\sigma^2(t)$, we replace it with some estimator, most commonly that computed from Greenwood's formula:

$$\widehat{\sigma^2}(t) = n\sum_{t_j \leq t}\frac{d_j}{n_j(n_j - d_j)}$$

Pivoting, an approximate $100\gamma\%$ CI for $H(t)$ is $\widehat{H}(t) \pm c_{CH}(t)$, where

$$c_{CH}(t) = \psi\left(\kappa\frac{\widehat{\sigma^2}(t)}{\widehat{\sigma^2}(t_U)}\right)\sqrt{\frac{\widehat{\sigma^2}(t)}{n}}$$

Next, consider constructing AMA bands for the difference of two cumulative hazard functions $H_1(t) - H_2(t)$. Again, asymptotic results dictate

$$[\widehat{H}_1(t) - \widehat{H}_2(t)] - [H_1(t) - H_2(t)] = [\widehat{H}_1(t) - H_1(t)] - [\widehat{H}_2(t) - H_2(t)]$$

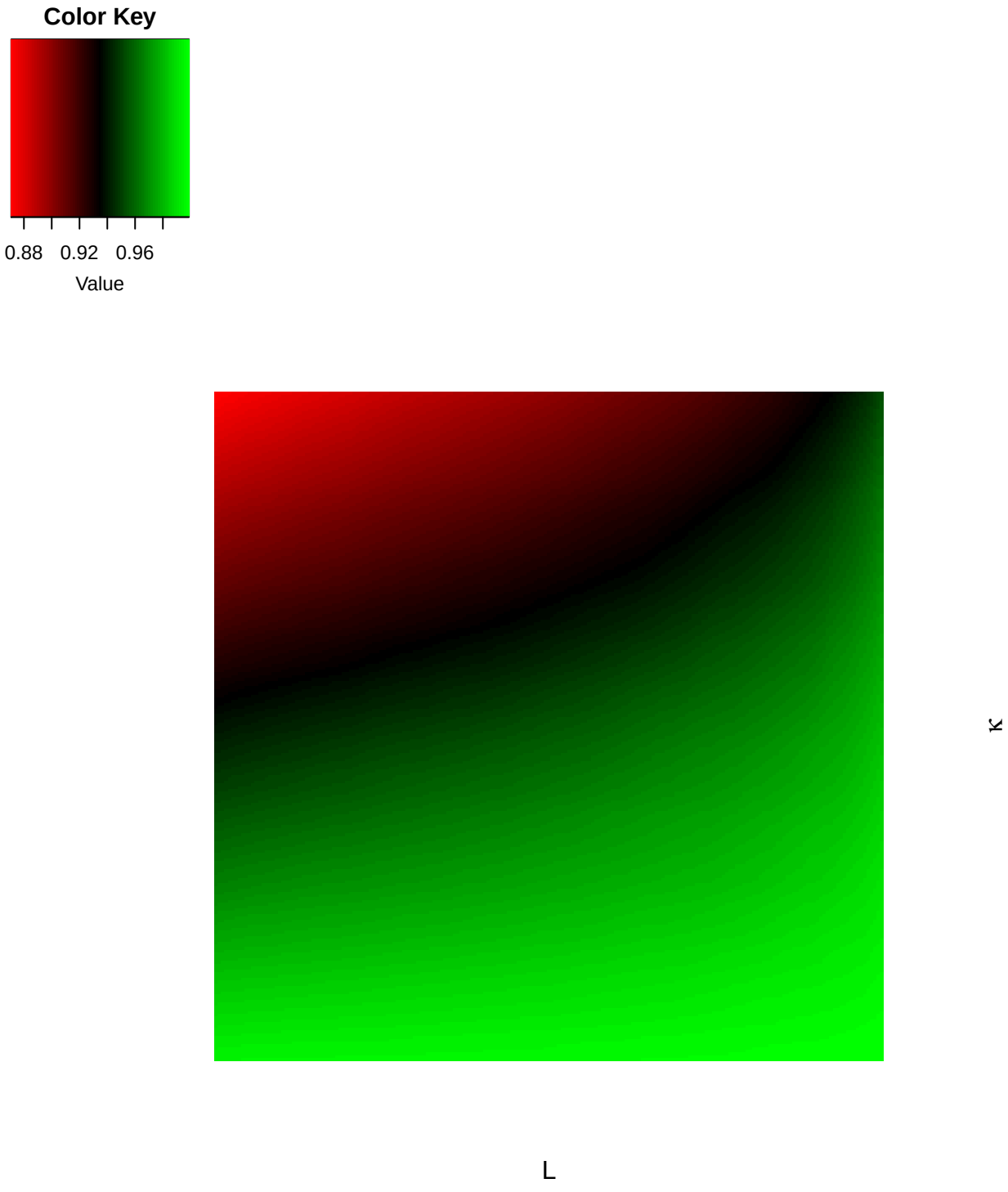

Figure A1: heatmap of simulated coverage table for kappa and gamma values.

weakly converges to $W\left(\frac{\sigma_1^2(t)}{n_1}+\frac{\sigma_2^2(t)}{n_2}\right)$. Applying the same strategy from before, we derive the $100\gamma\%$ CB's for $H_1(t)-H_2(t)$ as $[\widehat{H}_1(t)-\widehat{H}_2(t)]\pm c_{2CH}(t)$, where

$$c_{2CH}(t)=\psi\left(\kappa\frac{\widehat{\sigma^2}_{\text{agg}}(t)}{\widehat{\sigma^2}_{\text{agg}}(t_U)}\right)\sqrt{\frac{\widehat{\sigma^2}_{\text{agg}}(t)}{n_1+n_2}}$$

and $\widehat{\sigma^2}_{\text{agg}}(t)=(n_1+n_2)\left(\frac{\widehat{\sigma_1^2}(t)}{n_1}+\frac{\widehat{\sigma_2^2}(t)}{n_2}\right)$.

Let $\widehat{S}(t)$ be the Nelson-Aalen or Kaplan-Meier estimator for the true survival function $S(t)$. Standard results show that

$$\sqrt{n}\frac{\widehat{S}(t)-S(t)}{S(t)}\xrightarrow{w}W\left(\sigma^2(t)\right)$$

If we were to apply the result from before, we would ultimately be minimizing the confidence bands surrounding $\frac{\widehat{S}(t)-S(t)}{S(t)}$, not $\widehat{S}(t)-S(t)$. We fix this problem by constructing AMA bands weighted by $S(t)$; that is, we target $\widetilde{u}(t)=S(t)u(t)$ so that $\sqrt{n}|\widehat{S}(t)-S(t)|\leq\widetilde{u}(t)$ with the same AMA procedure. The bands for $S(t)$ are $\widehat{S}(t)\left(1\pm c_S(t)\right)$, where

$$c_S(t)=\psi\left(\kappa\widehat{S}(t)\frac{\widehat{\sigma^2}(t)}{\widehat{\sigma^2}(t_U)}\right)\sqrt{\frac{\widehat{\sigma^2}(t)}{n}}$$

and κ is calculated according to Eq 2 to obtain a coverage level of γ:

$$\kappa=-\frac{\tilde{b}+\sqrt{\tilde{b}^2-4\tilde{a}\tilde{c}}}{2\tilde{a}} \tag{2}$$

where

$$\begin{aligned}
\tilde{a}&=a\overline{S}_{K-1}^2\\
\tilde{b}&=\frac{b}{\widehat{\sigma^2}(t_U)}\left\{\sum_{i=1}^{K-2}\overline{S}_i(\widehat{\sigma^2}(\xi_i)-\widehat{\sigma^2}(\xi_{i+1}))\right\}+\left(a+b\frac{\widehat{\sigma^2}(\xi_{K-1})}{\widehat{\sigma^2}(t_U)}\right)\overline{S}_{K-1}\\
\tilde{c}&=1-\gamma
\end{aligned}$$

and $\overline{S}_i=(\widehat{S}(\xi_i)+\widehat{S}(\xi_{i+1}))/2$, where ξ_i is the ith observed (non-censored) time point.