STAT331

1-Sample Problem: Confidence Intervals for Quantiles and Confidence Bands

In this unit we use the weak convergence results from Unit 15 to obtain approximate confidence intervals for $S(t)$ (for fixed t), approximate confidence intervals for quantiles of $S(\cdot)$, say $t_p = S^{-1}(p)$ for fixed p, and approximate confidence bands for $S(\cdot)$.

Confidence Intervals for $S(t)$ (fixed t) : From (15.5) we have that

$$\sqrt{n}\,(\widehat{S}(\cdot) - S(\cdot)) \xrightarrow{w} -S(\cdot)\cdot X(\cdot), \tag{16.1}$$

where $X(\cdot)$ is a zero mean Gaussian process with independent increments and variance function $\sigma^2(t)$ which can be estimated by (15.3). For fixed t, it follows that the sequence of random variables

$$\sqrt{n}\,\left(\widehat{S}(t) - S(t)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, S^2(t)\,\sigma^2(t)).$$

Thus,

$$\widehat{S}(t) \overset{apx}{\sim} \mathcal{N}\left(S(t), S^2(t)\sigma^2(t)/n\right),$$

so that an approximate 95% confidence interval for $S(t)$ is given by

$$\hat{S}(t) \pm 1.96\sqrt{\frac{\hat{\sigma}^2(t)\hat{S}^2(t)}{n}}.$$

Alternatively, one could replace the term within the square root by Greenwood's formula (see Unit 5), or by (15.6).

The above formula can lead to a confidence interval whose upper bound exceeds 1, or whose lower bound is less than 0. An alternative approach that would avoid this is to use the δ-method to find a CI for $\log(-\log(S(t)))$, and then transform this interval back to get an interval for $S(t)$.

Confidence Regions for Quantiles of $S(\cdot)$: Now suppose we want to find an approximate confidence interval (CI) for a quantile of $S(\cdot)$, say $t_p = S^{-1}(p)$ for some fixed p. For example, we may want an approximate confidence interval for the median survival time $t_{.5} = S^{-1}(.5)$. We briefly summarize methods for this below. For more details, see Emerson (1982) and Brookmeyer & Crowley (1982).

To allow for discrete and "partially flat" distributions, define the inverse function and quantile t_p (for $0 < p < 1$) as:

$$t_p \stackrel{def}{=} S^{-1}(p) \stackrel{def}{=} \inf\{t \ : S(t) \leq p\}.$$

For continuous, strictly decreasing $S(\cdot)$, $S^{-1}(\cdot)$ is just the usual inverse function.

One approach for getting an approximate CI for t_p, analogous to the approach used to get an approximate CI for S(t), is to use the

functional delta method to show that

$$\sqrt{n}\left(\widehat{S}^{-1}(\cdot) - S^{-1}(\cdot)\right)$$

converges weakly to a zero mean Gaussian process. The conditions for the delta method are satisfied: because of Theorem II.8.2 and Proposition II.8.5 of Andersen et al., if we restrict the domain to $[p_0, p_1]$ such that F is continuously differentiable on an open interval containing $[p_0, p_1]$, with positive derivative. However, the variance of the asymptotic distribution depends on the density, $f(t) = -S'(t)$, which (being a density function) is hard to estimate precisely in practice.

Thus, as an alternative, we consider an approximate CI for t_p by inverting a hypothesis test. For a fixed t, consider testing the hypothesis

$$H_t : S(t) = p,$$

for some fixed p. This can be tested with the statistic

$$Z_t = \frac{\widehat{S}(t) - p}{\sqrt{\widehat{\mathrm{Var}}\left(\widehat{S}(t)\right)}}$$

where the term in the square root is, for example, Greenwood's formula. Since $Z_t \xrightarrow{\mathcal{L}} N(0,1)$ under H_t, an approximate 2-sided test would reject H_t at the .05 level if $\mid Z_t \mid > 1.96$. Now consider the set

$$\mathcal{T} = \{t \ : \ |Z_t| \leq 1.96\},$$

i.e., set of all t for which H_t is not rejected at the 0.05 level. Notice that $t_p \in \mathcal{T}$ if and only if $\mid Z_{t_p} \mid \leq 1.96$. Thus, $\mathcal{T}$ is an approximate 95% confidence region for t_p:

$$P\left(\mathcal{T} \text{ contains } t_p\right) = P\left(\left|Z_{t_p}\right| \leq 1.96\right) \to 0.95$$

if t_p is the true pth quantile.

Several things are worth noting about CIs obtained in this way:

- Since $\widehat{S}(\cdot)$ only jumps at observed failure times, say $\tau_1 < \tau_2 < \cdots$, then $\tau_i \in \mathcal{T}$ implies that $[\tau_i, \tau_{i+1}) \subseteq \mathcal{T}$.
- In some instances, Z_t is not a monotonically decreasing function of t. As a result, $\mathcal{T}$ need not be an <u>interval</u>, although this is not common. In cases that $\mathcal{T}$ is not an interval, one option would be to use the interval formed by its smallest and largest elements.
- If the largest observed failure time is in $\mathcal{T}$, then $\mathcal{T}$ is of the form $[a, \infty)$.
  If p is not reached by $\widehat{S}(\cdot)$, then $\mathcal{T}$ is also of this form.

Figure 16-1, taken from Brookmeyer & Crowley, illustrates this approach for p=1/2; that is, getting an approximate CI for the median. The data are survival times for 53 patients with colorectal cancer being treated with the drug 5-FU. Sixteen of the 53 survival times are right censored. The estimated median survival time is 61 weeks, and the corresponding approximate 95% CI is [38,73).

Figure 16-1

From Brookmeyer & Crowley, Biocs, 1982, p. 20

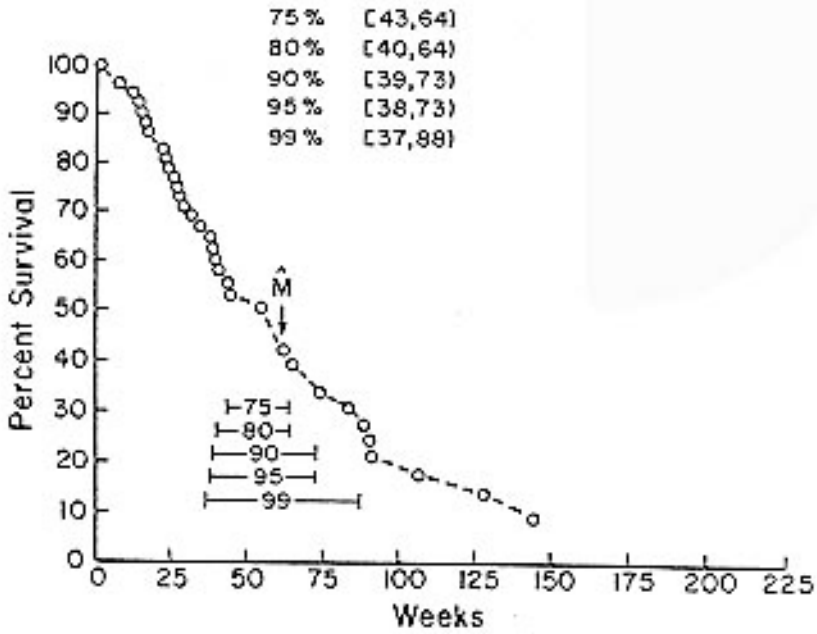

Figure 2. Confidence intervals for Treatment 1.

Confidence Bands for $S(\cdot)$: Now consider how we might obtain a level of 1- α <u>confidence band</u> for $S(t)$ for $0 \leq t \leq \tau$, for some τ. Throughout we suppose that τ is in the interior of the support of $F(\cdot)$ and $G(\cdot)$, so that $1-F(\tau) > 0$ and $1-G(\tau) > 0$. The goal is to find random functions, say $L(\cdot)$ and $U(\cdot)$ defined over $[0, \tau]$, such that

$$P\,(\ L(t) \leq S(t) \leq U(t) \text{ for all } t \in [0, \tau]\ ) \geq 1 - \alpha \qquad (16.2).$$

Note that:

- One way to think of a confidence band for $S(\cdot)$ is as a set of CIs, say $[L(t), U(t)]$, for $t \in [0, \tau]$ for which all the CIs simultaneously contain their theoretical counterparts with probability at least 1- α.
- With an ordinary CI, we can choose how to 'spread' our α error in the sense that the parameter of interest will fall above the interval with probability α_1 and below the interval with probability $\alpha_2 = \alpha - \alpha_1$. In most applications we take $\alpha_1 = \alpha_2 = \alpha/2$. With confidence bands we have this choice but also have the choice of how to spread our α over the interval $[0, \tau]$. In some applications we might want our bands to be of equal width but in other applications we might want them to be narrower for certain values of t, (e.g., larger t) than for other values of t because we are more interested in knowing about $S(t)$ at certain values.

<u>General Approach:</u> A $(1 - \alpha)100\%$ confidence bands for a function, say $S(\cdot)$, can be formed by finding a process $R(\cdot)$ and a nonnegative deterministic function $c(\cdot)$ such that

$$P\left(\left|R(t)(\hat{S}(t)-S(t))\right| \leq c(t) \text{ for all } t \in [0,\tau]\right) \quad \geq \quad 1-\alpha. \qquad (16.3)$$

Then, by taking

$$L(t) = \hat{S}(t) - c(t)/ \mid R(t) \mid$$

and

$$U(t) = \hat{S}(t) + c(t)/ \mid R(t) \mid$$

we would have that

$$1-\alpha \leq P\left(\left|R(t)(\hat{S}(t)-S(t))\right| \leq c(t), \quad \text{for all } t \in [0,\tau]\right)$$

$$= P\left(L(t) \leq S(t) \leq U(t) \text{ for all } t \in [0,\tau]\right),$$

so that L(·) and U(·) selected in this way represent a level 1-α confidence band for S(·). The width of these bands at time t is $2c(t)/ \mid R(t) \mid$. If $c(\cdot)$ were constant, say $c(t) = c$, we can write (16.3) as

$$P\left(\sup_{0\leq t\leq\tau} \mid R(t)(\hat{S}(t)-S(t)) \mid \;\leq\; c\right) \quad \geq \quad 1-\alpha. \quad (16.4)$$

If we take this approach, the problem of constructing a confidence band reduces to finding the constant c so that (16.4) holds. This is generally done by finding the limiting distribution of the process $R(\cdot)[\hat{S}(\cdot) - S(\cdot)]$ as $n \to \infty$ and then applying the Continuous Mapping Theorem to find the limiting distribution of the random variable $Z_n = \sup_{0\leq t\leq\tau} \mid R(t)(\hat{S}(t) - S(t)) \mid$.

Confidence Bands Based on Critical Values of a Wiener Process: Note that we can write (16.1) equivalently as

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

where $W(\cdot)$ is a Wiener process. It follows (e.g. from Slutzky's Lemma) that

$$\frac{1}{S(\cdot)}\sqrt{n}\left(\widehat{S}(\cdot) - S(\cdot)\right) \xrightarrow{w} W\left(\sigma^2(\cdot)\right)$$

and, from Slutsky's Lemma, that

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

Thus,

$$\frac{1}{\sigma(\tau)}\sqrt{n}\left(\frac{\widehat{S}(\cdot) - S(\cdot)}{\widehat{S}(\cdot)}\right) \xrightarrow{w} W\left(\frac{\sigma^2(\cdot)}{\sigma^2(\tau)}\right).$$

Also, $\sigma^2(0) = 0$ and $\sigma^2(t)$ is nondecreasing and continuous.

We applied Slutzky's Lemma twice to division, although it was only proved for sums and products. However, this follows from multiplication:

Lemma 16.1 If $\hat{S}(\cdot) - S(\cdot)$ converges in probability to the zero-process on $[0, \tau]$ and S is bounded away from 0 on $[0, \tau]$, also $\frac{1}{\hat{S}(\cdot)} - \frac{1}{S(\cdot)}$ converges in probability to the zero-process on $[0, \tau]$.

Proof: Notice first that

$$\left|\frac{1}{\hat{S}(t)} - \frac{1}{S(t)}\right| = \frac{\left|S(t) - \hat{S}(t)\right|}{\hat{S}(t)S(t)} \leq \frac{\left|S(t) - \hat{S}(t)\right|}{\hat{S}(\tau)S(\tau)}.$$

Now let $\epsilon > 0$ be given. We want to show that the probability that the above expression exceeds ϵ goes to 0. We know that for any $\delta > 0$, the probability that $\sup_{t\in[0,\tau]} \left|S(t) - \hat{S}(t)\right| > \delta$ converges to 0. If $\sup_{t\in[0,\tau]} \left|S(t) - \hat{S}(t)\right| \leq \delta$, from above,

$$\left|\frac{1}{\hat{S}(t)} - \frac{1}{S(t)}\right| \leq \frac{\delta}{(S(\tau) - \delta)S(\tau)}.$$

We want to show that the probability that this is greater than ϵ converges to 0. So, choose δ small: e.g. in such a way that $S(\tau) - \delta > S(\tau)/2$ and such that $\delta < \epsilon S(\tau)S(\tau)/2$. Then, again because of the above, if $\sup_{t\in[0,\tau]} \left|S(t) - \hat{S}(t)\right| \leq \delta$, then

$$\begin{aligned}\left|\frac{1}{\hat{S}(t)} - \frac{1}{S(t)}\right| &\leq \frac{\delta}{S(\tau)(S(\tau) - \delta)} \\ &\leq \frac{\epsilon S(\tau)S(\tau)/2}{S(\tau)S(\tau)/2} = \epsilon.\end{aligned}$$

Therefore,

$$P\left(\sup_{t\in[0,\tau]} \left|\frac{1}{\hat{S}(t)} - \frac{1}{S(t)}\right| > \epsilon\right) \leq P\left(\sup_{t\in[0,\tau]} \left|S(t) - \hat{S}(t)\right| > \delta\right) \to 0$$

by assumption. This concludes the proof. □

Notice that the proof works to prove Slutzky's Lemma for division for convergence in probability of elements of $D[0, \tau]$, not just for the survival function, as long as the limiting function is bounded away from 0.

Going on where we left before Lemma 16.1:

Thus, by the Continuous Mapping Theorem and Slutsky's Lemma,

$$\sup_{0\le t\le\tau}\left\{\frac{\sqrt{n}}{\hat{\sigma}(\tau)}\left|\frac{\hat{S}(t)-S(t)}{\hat{S}(t)}\right|\right\} \xrightarrow{\mathcal{L}} \sup_{0\le t\le\tau} \mid W(\sigma^2(t)/\sigma^2(\tau)) \mid$$

$$= \sup_{0\le u\le 1} |W(u)| \,. \qquad (16.5)$$

This is a useful result because such critical values for a Wiener process can be expressed analytically (cf, Hall & Wellner, 1980). Specifically,

$$P\left(\sup_{0\le u\le 1}|W(u)|\le c\right) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)}\, e^{\frac{-\pi^2(2k+1)^2}{8c^2}}.$$

Tabulated values are given in M. Schumacher (1984). For example, c=2.241 when α=0.05, so that from (16.5),

$$0.95 \;=\; P[\sup_{0\le u\le 1} \mid W(u) \mid \;\le\; 2.241\,] \cong P[\sup_{0\le t\le\tau} \frac{\sqrt{n}}{\hat{\sigma}(\tau)} \mid \frac{\hat{S}(t)-S(t)}{\hat{S}(t)} \mid \;\le\; 2.241].$$

Thus, an approximate 95% confidence band for $S(\cdot)$ is given by

$$\hat{S}(t) \pm \frac{2.241\;\hat{\sigma}(\tau)\cdot\hat{S}(t)}{\sqrt{n}} \qquad \text{for} \quad 0\le t\le\tau. \qquad (16.6)$$

Fleming & Harrington(1991) refer to these confidence bands as "Gill bands". These bands correspond to the general approach in (16.3) with $R(t) = \sqrt{n}/(\hat{\sigma}(\tau)\hat{S}(t))$. The width of this band at time t is $2(2.241)\cdot\hat{\sigma}(\tau)\cdot\hat{S}(t)/\sqrt{n}$. Thus, since $\hat{S}(t)$ is nonincreasing in t, the band gets narrower as t goes from 0 to τ.

The top graph on Figure 16-2 is taken from Fleming & Harrington (1991, page 241). It provides the Kaplan-Meier estimator and approximate 95% confidence bands based on (16.6). Also included are the pointwise confidence intervals for S(t) at each t. The estimates are based on 418 subjects. Note that the confidence bands are wider than the pointwise confidence intervals, as the former simultaneously contain the entire function with approximate probability .95.

Figure 16-2

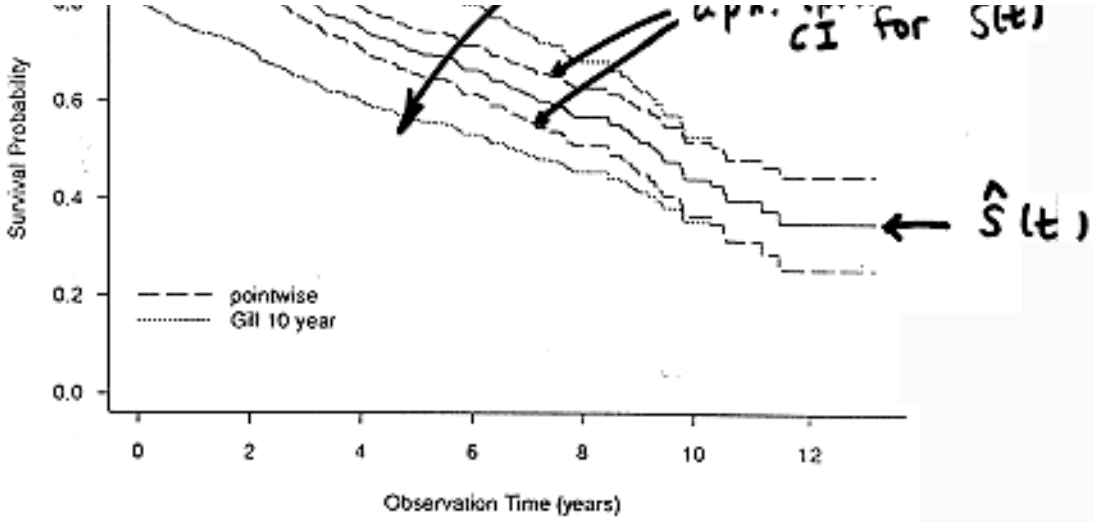

Figure 6.3.2a Pointwise confidence intervals and confidence bands (using the Gill ten-year bands) for survival probability using all 418 patients in the PBC data set.

252

LARGE SAMPLE RESULTS

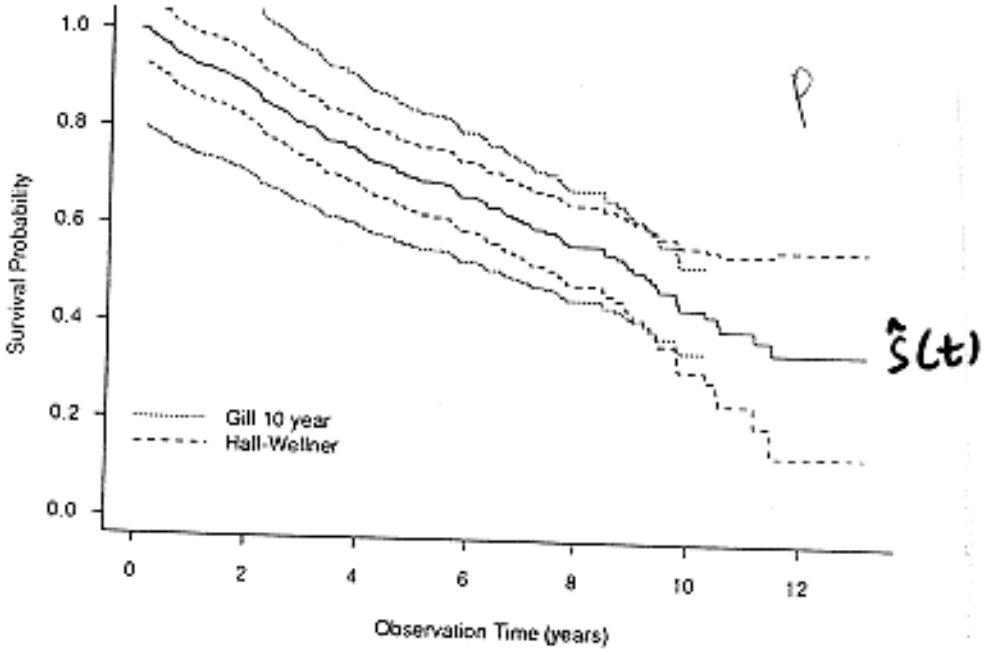

Figure 6.3.5b Gill 10-year and Hall-Wellner confidence bands in the PBC data

Confidence Bands Based on Critical Values of a Brownian Bridge:

Differently-shaped bands can be formed by considering other multipliers of $\sqrt{n}\left(\widehat{S}(\cdot) - S(\cdot)\right)$. Hall & Wellner (1980, p. 133) do this by making use of the fact that if $W(\cdot)$ is a Wiener process on $[0, \infty)$, then

$$W_0(t) \stackrel{def}{=} (1-t) \cdot W\left(\frac{t}{1-t}\right) \qquad 0 < t < 1$$

is a Brownian Bridge (Exercise 2). Define

$$K(t) = \frac{\sigma^2(t)}{1+\sigma^2(t)}$$

so that

$$\sigma^2(t) = \frac{K(t)}{1-K(t)}.$$

Then, as before,

$$\sqrt{n}\left(\frac{\widehat{S}(\cdot) - S(\cdot)}{\widehat{S}(\cdot)}\right) \stackrel{w}{\longrightarrow} W(\sigma^2(\cdot))$$

so that, e.g. because of Slutzky's Lemma,

$$(1-K(\cdot))\sqrt{n}\left(\frac{\widehat{S}(\cdot) - S(\cdot)}{\widehat{S}(\cdot)}\right) \stackrel{w}{\longrightarrow} (1-K(\cdot))W\left(\frac{K(\cdot)}{1-K(\cdot)}\right)$$

That is,

$$(1-K(\cdot))\sqrt{n}\left(\frac{\widehat{S}(\cdot) - S(\cdot)}{\widehat{S}(\cdot)}\right) \stackrel{w}{\longrightarrow} W_0(K(\cdot)),$$

and hence (Slutzky's Lemma)

$$\left(1-\widehat{K}(\cdot)\right)\ \sqrt{n}\ \left(\frac{\widehat{S}(\cdot)-S(\cdot)}{\widehat{S}(\cdot)}\right) \xrightarrow{w} W_0(K(\cdot)),$$

where

$$\widehat{K}(\cdot)=\frac{\widehat{\sigma}^2(\cdot)}{1+\widehat{\sigma}^2(\cdot)}.$$

Thus, by the Continuous Mapping Theorem,

$$\sup_{0\le t\le\tau}\left|(1-\hat{K}(t))\sqrt{n}\left(\frac{\hat{S}(t)-S(t)}{\hat{S}(t)}\right)\right| \xrightarrow{\mathcal{L}} \sup_{0\le t\le\tau}|W_0(K(t))|$$

or

$$\sup_{0\le t\le\tau}\left|(1-\hat{K}(t))\sqrt{n}\left(\frac{\hat{S}(t)-S(t)}{\hat{S}(t)}\right)\right| \xrightarrow{\mathcal{L}} \sup_{0\le u\le u^*}|W_0(u)|$$

where $u^*=K(\tau)$. Here we have used the fact that $K(t)\nearrow$ in t, $K(0)=0$ and $K(\tau)\le 1$.

Percentiles of the supremum of a Brownian Bridge have been tabulated. For example, Schumacher (1984) gives the values of c_{α,u^*} satisfying

$$P\left(\sup_{0\le u\le u^*}|W_0(u)|\le c_{\alpha,u^*}\right)=1-\alpha$$

for various choices of α and u^*. We can use these to form the approximate $1-\alpha$ confidence bands for $(S(t) \ : \ 0 \le t \le \tau)$ defined by

$$\widehat{S}(t) \pm \frac{\widehat{S}(t) \cdot c_{\alpha,u^*}}{\left(1-\widehat{K}(t)\right)\sqrt{n}} \qquad 0 \le t \le \tau.$$

Note 1: In most situations, τ is such that $\widehat{K}(\tau) \ge .7$. Here, $c_{\alpha,u^*} \approx c_{\alpha,1}$ (recall that the Brownian Bridge returns to the origin at time 1) and so we can use $c_{\alpha,1}$ in the bands. This is illustrated in the table below for $\alpha = 0.05$, which was taken from Schumacher (1984).

u^*	c_{α,u^*}
.6	1.3211
.7	1.3471
.8	1.3564
.9	1.3581
1.0	1.3582

Note 2: This approach corresponds to (16.3) with $R(t) = (1-\hat{K}(t))\sqrt{n}/\hat{S}(t)$. The width of these bands at time t is

$$\frac{2 \cdot c_{\alpha,u^*}}{\sqrt{n}} \ \frac{\hat{S}(t)}{1-\hat{K}(t)} = \frac{2c_{\alpha,u^*}}{\sqrt{n}} \ \cdot \ \left(1+\hat{\sigma}^2(t)\ \right) \ \hat{S}(t).$$

The factor involving $\hat{\sigma(t)}$ is increasing in t, while the factor involving $\hat{S}(t)$ is decreasing in t. As a result, these bands are more nearly constant in width than the bands defined in (16.6). This can be seen on the bottom plot in Figure 16-2, where the Hall-Wellner bands are plotted along with the so called "Gill" bands defined in (16.6).

Exercises

1. Suppose that $X(\cdot)$ is a zero-mean Gaussian process with independent increments, X(0)=0, and variance function Var(X(t))=$\sigma^2(t)$. Suppose also that $\sigma(0)$=0 and that $\sigma(t)$ is increasing in t. Define the new process $W(\cdot)$ by $W(\sigma^2(t)) = X(t)$. Prove that $W(\cdot)$ is a Wiener process.

2. Suppose that $W(\cdot)$ is a Wiener process.
   Define $X(t) = \begin{cases} (1-t)W\left(\frac{t}{1-t}\right), & 0 \le t < 1 \\ 0, & t = 1 \end{cases}$.
   Prove that $X(\cdot)$ is a Brownian Bridge.

3. Consider the 1-sample problem and let $\hat{\Lambda}(\cdot)$ denote the Nelson-Aalen estimator of the cumulative hazard function $\Lambda(\cdot)$ based on n i.i.d. observations. Suppose that as $n \to \infty$,

$$\sqrt{n}(\hat{\Lambda}(\cdot) - \Lambda(\cdot)) \xrightarrow{w} X(\cdot),$$

   in $[0, \tau]$, where $X(\cdot)$ is a zero-mean Gaussian process with independent increments and variance function at time t given by $Var(X(t)) = \sigma^2(t)$. We note that $\sigma^2(t)$ is increasing in t for $0 \le t \le \tau$, and assume that we have a consistent estimator denoted $\hat{\sigma}^2(t)$.

   Using the fact that for a Wiener process $W(\cdot)$,

$$P(sup_{0<u<1} \mid W(u) \mid < 2.241) \;=\; 0.95,$$

   construct an approximate 95% confidence band for $\Lambda(t)$ for $0 < t < \tau$. Please briefly justify your answer.

4. Show that under the assumptions in this unit, $\hat{K}(\cdot) \xrightarrow{P} K(\cdot)$ as processes on $[0, \tau]$ on page 14. How is this used on page 14?

References

Emerson, J. (1982) *Biometrics*, p. 17.

Brookmeyer R & Crowley J (1982), *Biometrics*, p. 29.

Fleming TR and Harrington DP (1991). Counting Processes and Survival Analysis. Wiley, New York.

Hall J & Wellner J (1980), *Biometrika*

Schumacher M (1984). *Intl Stat Review* 52:263-281.