Tests and Confidence Sets for Comparing Two Mean Residual Life Functions

Roger L. Berger; Dennis D. Boos; Frank M. Guess

Biometrics, Vol. 44, No. 1 (Mar., 1988), 103-115.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28198803%2944%3A1%3C103%3ATACSFC%3E2.0.CO%3B2-%23

Biometrics is currently published by International Biometric Society.

BIOMETRICS 44, 103–115
March 1988

Tests and Confidence Sets for Comparing Two Mean Residual Life Functions

Roger L. Berger and Dennis D. Boos

Statistics Department, Box 8203, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A.

and

Frank M. Guess

Statistics Department, University of South Carolina, Columbia, South Carolina 29208, U.S.A.

SUMMARY

The mean residual life function of a population gives an intuitive and interesting perspective on the aging process. Here we present new nonparametric methods for comparing mean residual life functions based on two independent samples. These methods have the flexibility to handle crossings of the functions and result in a new type of confidence set. We also discuss similar methods for comparison of median residual life functions.

1. Introduction

The mean residual life of an animal at age t is the average remaining life among those population members who have survived until time t. If lifelengths of the population are described by a random variable X with distribution function $F(x)$, then the mean residual life function is defined by

$$e_F(t) = \mathrm{E}(X - t \mid X > t) = \int_t^\infty \bar{F}(u)\, du/\bar{F}(t),$$

where $\bar{F}(t) = 1 - F(t)$ is the survival function.

The mean residual life function gives a different picture of survival or aging than is seen through the more commonly studied survival function $\bar{F}(t)$ or hazard (or failure) rate function $f(t)/\bar{F}(t)$. A researcher should find all three measures interesting and complementary. To illustrate a different perspective, consider a male cancer patient undergoing chemotherapy. He would be very interested in knowing how long he should expect to live given that his chemotherapy began $t = 6$ months ago. His expected remaining life is $e_F(6)$. The proportion of patients like him who survive the first 6 months is $\bar{F}(6)$ and his instantaneous probability of dying tomorrow is $f(6)/\bar{F}(6)$.

The variable of interest X need not be a "lifetime" and $t = 0$ need not correspond to "birth." As in the above example, $t = 0$ will often denote the beginning of a treatment, and $e_F(t)$ is the expected remaining life of a subject who has survived the treatment for time t. Indeed, X need not measure time. If X is the medical cost of a patient, then $e_F(t)$ is the

Key words: Intersection–union principle; Mean residual life function; Median residual life function; Nonparametric confidence sets; Nonparametric test.

expected additional cost given that an amount t has already been charged. A variety of other applications is provided in the recent review paper of Guess and Proschan (1988). Further discussion of the mean residual life function in biological contexts may be found in Deevey (1947), Chiang (1960), Bryson and Siddiqui (1969), and Chen, Hollander, and Langberg (1983).

In this paper we introduce a new nonparametric test and confidence procedures for comparing the mean residual life functions from two populations or treatment groups. For a specified age interval $[T_1, T_2]$, the testing problem is

$$H_0: \quad e_F(t) \leq e_G(t) \qquad \text{for some } t \in [T_1, T_2]$$

versus (1.1)

$$H_a: \quad e_F(t) > e_G(t) \qquad \text{for all } t \in [T_1, T_2],$$

where F and G are the two population distribution functions. If, for example, H_0 were rejected when comparing two human populations over the age interval [60 years, 70 years], then an annuity company might make lower monthly payments to people from the first population (distribution F) because the expected number of monthly payments is greater for these people. This would be true for any person starting payments between the ages of 60 and 70.

An important feature of the hypotheses (1.1) is that the test may be inverted to obtain confidence statements of the form "$e_F(t) > e_G(t)$ for all $t \in \hat{I}$," where $\hat{I}$ is an interval of values determined by the data. This is a new approach to the problem of constructing confidence statements for comparing two functions. In the above example, our procedures would allow the calculation of an interval, with center fixed at $t = 65$ years, for which the statement could be made that $e_F(t) > e_G(t)$ for all t in the interval.

Figure 1 shows the empirical mean residual life functions calculated from samples from two populations of guinea pigs. The two populations correspond to injections with different concentrations of tubercle bacilli. For small values of t, it is reasonable to hypothesize that the mean residual life at time t is larger for group 1 (guinea pigs receiving the smaller concentration). On the other hand, for a large value of t, a guinea pig from group 2 that has survived to this time t might be a particularly hardy animal and might have a longer expected remaining life than a survivor to this time from group 1. The empirical graphs in Figure 1 also suggest that the group 1 mean residual life function may be larger than the group 2 mean residual life function for small values of t. Using our procedures, we can test, for example,

$$H_0: \quad e_1(t) \leq e_2(t) \qquad \text{for some } t \in [0, 80]$$

versus

$$H_a: \quad e_1(t) > e_2(t) \qquad \text{for all } t \in [0, 80].$$

Alternatively, we could construct an interval of the form $\hat{I} = [0, \hat{\theta})$, where $\hat{\theta}$ is calculated from the data, and make the confidence statement that $e_1(t) > e_2(t)$ for all $t \in [0, \hat{\theta})$. Further analysis of this data is given in Sections 3 and 4.

Our test and confidence statements are quite different from those suggested by other authors. The tests proposed by Cheng (1985) to compare failure rate functions and by Joe and Proschan (1984) to compare percentile residual life functions, as well as many other tests proposed to compare functions, test the null hypothesis that two functions are equal for all t, versus the alternative that one function dominates the other for all t. These models do not account for the realistic possibility that the functions cross. A test designed only to test the null hypothesis of equality may have a large probability of rejecting this null hypothesis for two populations whose functions cross. Rejection of the null hypothesis by

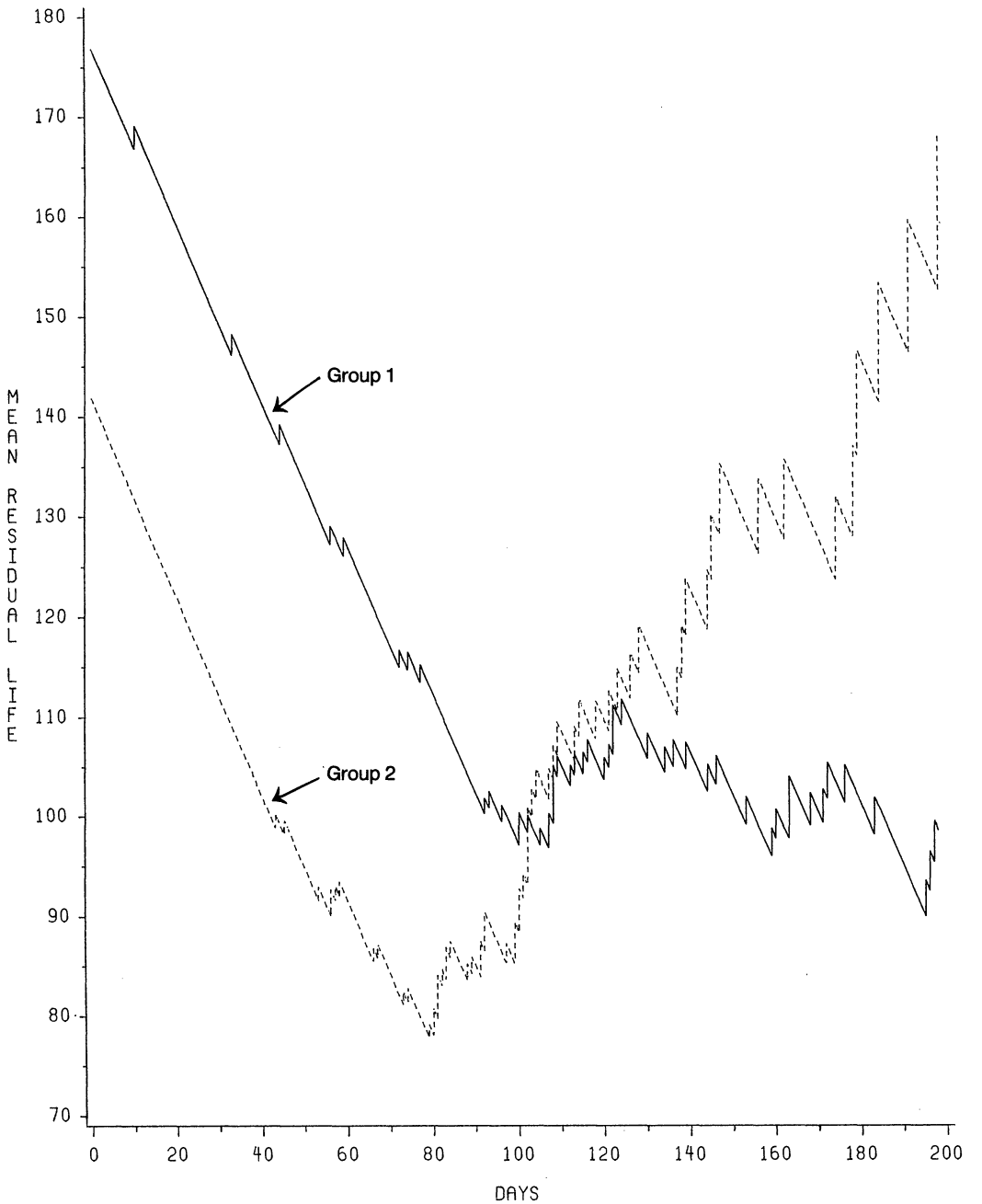

Figure 1. Mean residual life of guinea pigs injected with different amounts of tubercle bacilli. From Bjerkedal (1960).

such a test can be interpreted as evidence that one function dominates the other only if the possibility of crossing functions can be eliminated a priori. The distinctive feature of our formulation is that our H_0 and H_a include all possibilities, including the possibility that the functions $e_F(t)$ and $e_G(t)$ cross. In any given problem, either H_0 or H_a is true. It is our consideration of H_0 and H_a, including crossing functions, that results in the confidence statements we have mentioned. Such confidence statements do not arise from tests of an

equality null hypothesis. Because these confidence statements specify a range of values for which one function dominates the other, they are easily interpretable and meaningful in many contexts.

In Section 2, the test and confidence statements are described. These procedures are applied to several examples in Section 3. Some possible modifications are presented in Section 4 to cover cases in which censoring occurs or robustness considerations are necessary. The theory justifying our procedure is presented in Section 5 followed by an Appendix with theorem proofs.

2. Hypothesis Test and Confidence Procedures

The goal is to test (1.1) based on independent random samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from F and G and to give confidence procedures of the type mentioned in the Introduction. Our approach is to use the intersection–union principle discussed by Gleser (1973) and Berger and Sinclair (1984). We define an asymptotic level α test for each simple problem

$$\begin{aligned} H_{0t}&: \ e_F(t) \leqslant e_G(t), \\ H_{at}&: \ e_F(t) > e_G(t), \end{aligned} \tag{2.1}$$

and then reject H_0 of (1.1) if and only if H_{0t} is rejected at level α for each $t \in [T_1, T_2]$. In Section 5 we show that the proposed test is asymptotically a size α test of (1.1).

The obvious test statistic for (2.1) is

$$Z_{mn}(t) = \left[\hat{e}_F(t) - \hat{e}_G(t)\right] \Big/ \left[\frac{S_m^2(t)}{m(t)} + \frac{S_n^2(t)}{n(t)}\right]^{1/2},$$

where $\hat{e}_F(t) + t$ is the average of the X_i's greater than t, $m(t)$ is the number of X_i's greater than t, and $S_m^2(t)$ is the sample variance of the X_i's greater than t with denominator $m(t) - 1$. The analogous quantities for the Y_i's are $\hat{e}_G(t)$, $S_n^2(t)$, and $n(t)$. Define $S_m^2(t) = 0$ when $m(t) \leqslant 1$, $S_n^2(t) = 0$ when $n(t) \leqslant 1$, and $Z_{mn}(t) = 0$ when $\min\{m(t), n(t)\} \leqslant 1$. Chiang (1960, pp. 226–227) discusses a discrete version of $Z_{mn}(t)$ for life-table analysis and Elandt-Johnson and Johnson (1980, §8.3) discuss related statistics.

Let z_α be the $(1 - \alpha)$th percentile of a standard normal distribution. Then the test that rejects H_{0t} if and only if $Z_{mn}(t) > z_\alpha$ is asymptotically a size α test of (2.1). Thus, the asymptotically size α test we propose for (1.1) rejects H_0 if and only if

$$Z_{mn}(t) > z_\alpha \quad \text{for every } t \in [T_1, T_2] \tag{2.2}$$

or, equivalently, if and only if

$$\inf_{t \in [T_1, T_2]} Z_{mn}(t) > z_\alpha.$$

The statistic $Z_{mn}(t)$, as a function of t, is a step function that has jumps only at t values that are X_i or Y_j observations. Thus, to perform the test it is necessary only to compute $Z_{mn}(t)$ for $t = T_1$ and those t values between T_1 and T_2 that are observed lifelengths in at least one of the samples, and compare these values of $Z_{mn}(t)$ with z_α. If $m(t)$ or $n(t)$ is small, we recommend replacing z_α by the appropriate t distribution percentile obtained by using Welch's unequal variances approximation (e.g., see Best and Rayner, 1987). In this case $Z_{mn}(t)$ would be compared to a different critical value for each value of t at which $Z_{mn}(t)$ is computed.

The test in (2.2) can be inverted to yield three different types of confidence statements. Each produces a confidence statement of the form "$e_F(t) > e_G(t)$ for all $t \in \hat{I}$" where

$\hat{I} \subset [0, \infty)$ is a random interval computed from the data. We show in Section 5 that the asymptotic confidence level for each procedure is $1 - \alpha$.

The three procedures differ in the form of the interval $\hat{I}$, but each requires that a value $T \geqslant 0$ be specified in advance of sampling. If $Z_{mn}(T) \leqslant z_\alpha$, then no confidence statement is made. If $Z_{mn}(T) > z_\alpha$, then the three procedures assert that $e_F(t) > e_G(t)$ for all $t \in \hat{I}$, where the $\hat{I}$ for each procedure is defined as follows:

$$\text{Procedure 1:} \quad \hat{I} = [T, \hat{\theta}_1) \text{ where } \hat{\theta}_1 = \inf\{t \geqslant T\colon Z_{mn}(t) \leqslant z_\alpha\}.$$

$$\text{Procedure 2:} \quad \hat{I} = (\hat{\theta}_2, T] \text{ where } \hat{\theta}_2 = \sup\{t \leqslant T\colon Z_{mn}(t) \leqslant z_\alpha\}.$$

$$\text{Procedure 3:} \quad \hat{I} = (\max(0, T - \hat{\delta}), T + \hat{\delta}) \text{ where}$$

$$\hat{\delta} = \sup\{d \geqslant 0\colon \inf_{t \in [\max(0, T-d), T+d]} Z_{mn}(t) > z_\alpha\}.$$

As mentioned earlier, we replace z_α by t distribution percentiles whenever $m(t)$ and $n(t)$ are small.

In the first type of interval, T might be chosen to be 0. Then $\hat{\theta}_1$ is the smallest value of t for which $Z_{mn}(t) \leqslant z_\alpha$. Figure 2 illustrates the use of all three types of intervals with simulated data from F lognormal such that $\log X$ is N($\log 2$, $\frac{1}{4}$) and G is exponential with mean 1.8. The sample sizes were $m = n = 200$. The statistic $Z_{mn}(t) \leqslant z_{.10} = 1.282$ for the first time at $\hat{\theta}_1 = .479$. Thus, the first confidence procedure asserts with confidence 90% that $e_F(t) > e_G(t)$ for all $t \in [0, .479)$.

To illustrate the second type of interval, suppose $T = 2$ were chosen as the upper limit of our interest. Reverse the roles of F and G in Figure 2. The largest $t \leqslant T = 2$ for which $Z_{mn}(t) \leqslant 1.282$ is $\hat{\theta}_2 = .999$. Actually $Z_{mn}(t) \geqslant -1.282$ in Figure 2. So the second confidence procedure asserts with confidence 90% that $e_G(t) > e_F(t)$ for all $t \in (.999, 2.000]$.

To illustrate the third type of interval, choose $T = 2$ as the center of the interval of interest. The value of t closest to $T = 2$ at which $Z_{mn}(t) \geqslant -1.282$ is $t = 2.425$. Thus, $\hat{\delta} = .425$ and the third confidence procedure asserts, with confidence 90%, that $e_G(t) > e_F(t)$ for all $t \in (1.575, 2.425)$.

As mentioned before, the function $Z_{mn}(t)$ changes values only at observed lifelengths. Thus, the endpoints, $\hat{\theta}_1$ and $\hat{\theta}_2$, for procedures 1 and 2 and one of the endpoints for procedure 3 will be observed lifelengths.

3. Examples

We illustrate the use of the hypothesis test and confidence procedures on data from two different experiments. The first experiment studied the lifelengths of guinea pigs after injection with different concentrations of tubercle bacilli. This experiment was introduced in Section 1.

Guinea pigs are known to have a high susceptibility to human tuberculosis, which is one reason for choosing this species. Bjerkedal (1960) studies the acquisition of resistance in these animals and provides the data that we use here. His study labeled M is for animals in a single cage under the same regimen. The regimen number is the common logarithm of the number of bacillary units in .5 ml of the challenge solution. For example, regimen 4.3 corresponds to 2.2×10^4 bacillary units per .5 ml, $\log_{10}(2.2 \times 10^4) = 4.342$. Here we compare regimen 4.3 to that of regimen 5.5. It is reasonable a priori to hypothesize H_a of (1.1) with $T_1 = 0$, $e_F(t) = e_{4.3}(t)$, and $e_G(t) = e_{5.5}(t)$. Some natural T_2 values to consider are $T_2 = 30$, 60, or 90 days. Figure 1 shows the estimates $\hat{e}_{4.3}(t)$ and $\hat{e}_{5.5}(t)$ labeled Group 1 and Group 2, respectively. The corresponding $Z_{mn}(t)$ function is

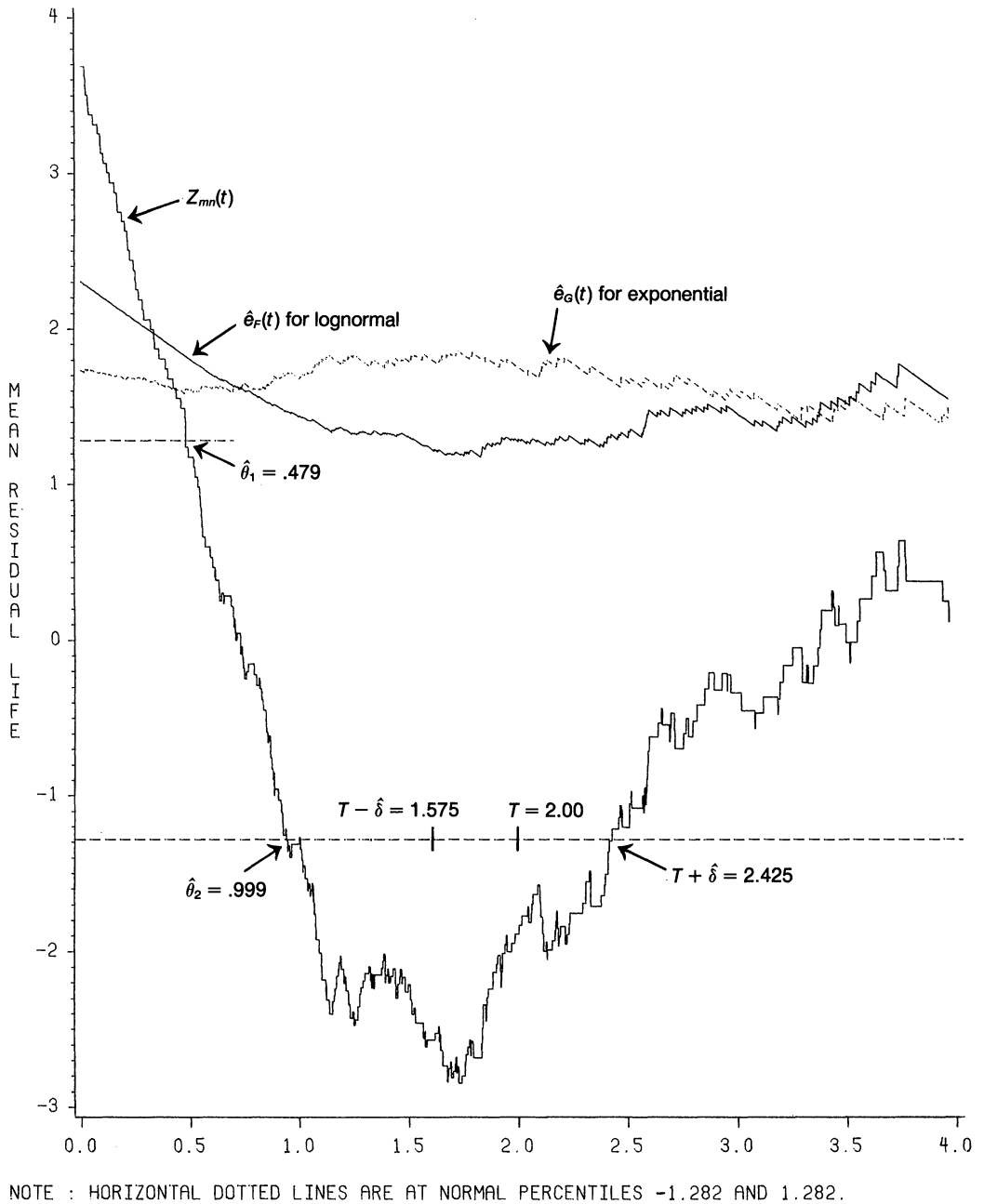

Figure 2. Mean residual life functions from simulated lognormal and exponential distributions.

around 2.0 for the first 60 days and then slowly decreases to 1.365 at $t = 81$ and to 1.277 at $t = 82$. Thus, at the approximate level $\alpha = .10$ we reject H_0 in favor of H_a for $T_2 = 30$ and $T_2 = 60$ since $Z_{mn}(t) > 1.282$ for $0 \leq t \leq 60$. For $T_2 = 90$ we do not reject H_0 since $Z_{mn}(82) < 1.282$.

The confidence procedures provide more insight into the testing results. If we specify $T = 0$ in the first confidence procedure, then we can assert with 90% confidence that $e_{4.3}(t) > e_{5.5}(t)$ for all $t \in [0, 82)$. Note that $t = 82$ is the first time that $Z_{mn}(t)$ crosses

$z_{.10} = 1.282$, the 90th percentile of the standard normal distribution. If our main interest were in the region near $T = 60$, then the second confidence procedure would give the interval (0, 60] and the third would give (38, 82).

Note in Figure 1 that the empirical mean residual life functions cross each other. One explanation is that the shock of tubercle bacilli is more severe under regimen 5.5 initially, and thus $e_{4.3}(t) > e_{5.5}(t)$ for the first interval. The later reversal suggests that a more vigorous subgroup has been *screened* after surviving the initial disease state under regimen 5.5. Thus, $e_{4.3}(t) < e_{5.5}(t)$ is reasonable for the later period. This example illustrates the flexibility of the techniques under natural crossings.

In the second example the empirical mean residual life functions do not cross. The context involves the influence of different diets on the aging process in rats where research indicates that diet restriction promotes longevity. Yu et al. (1982) study the effects of a restricted diet on rats versus an *ad libitum* diet, i.e., free eating (cf. also Witten, 1985). Table 1 contains the data. The numbers in each group are slightly different than reported by Yu et al. (1982) because the treatments began after an initial weaning period during which several rats died.

Table 1
Rat lifelength data in days

Restricted diet								
105	605	811	931	1011	1073	1133	1190	1244
193	630	833	940	1012	1076	1136	1203	1258
211	716	868	957	1014	1085	1138	1206	1268
236	718	871	958	1017	1090	1144	1209	1294
302	727	875	961	1032	1094	1149	1218	1316
363	731	893	962	1039	1099	1160	1220	1327
389	749	897	974	1045	1107	1166	1221	1328
390	769	901	979	1046	1119	1170	1228	1369
391	770	906	982	1047	1120	1173	1230	1393
403	789	907	1001	1057	1128	1181	1231	1435
530	804	919	1008	1063	1129	1183	1233	
604	810	923	1010	1070	1131	1188	1239	
***Ad libitum* diet**								
89	536	630	668	695	717	739	770	801
104	545	635	670	697	720	741	773	806
387	547	639	675	698	721	743	777	807
465	548	648	677	702	730	746	779	815
479	582	652	678	704	731	749	780	836
494	606	653	678	710	732	751	788	838
496	609	654	681	711	733	753	791	850
514	619	660	684	712	735	764	794	859
532	620	665	688	715	736	765	796	894
533	621	667	694	716	738	768	799	963

Let e_F correspond to the mean residual life function under the restricted diet and e_G under *ad libitum*. If is of interest to test (1.1) for $T_1 = 0$ and $T_2 = 730$ days or, equivalently, 2 years. The test overwhelmingly rejects H_0 in favor of H_a at level $\alpha = .01$. In fact, $Z_{mn}(t) \geq 9.2$ for all $0 \leq t \leq 730$. Although a subjective graphical analysis would lead to similar conclusions, a distinct advantage of our procedures is the objective quantitative assessment of the data.

For a confidence statement let $T = 0$. Then we can assert with 99% confidence that $e_F(t) > e_G(t)$ for $t \in [0, 894)$ days.

4. Censoring and Robustness

Whenever the data are censored in the right tail or one expects the right tail of the distribution to be heavy, it makes sense to consider other residual life measures such as the median residual life or trimmed mean residual life. The same basic approach as in Section 2 can be carried out, but modifications in the test statistic $Z_{mn}(t)$ are of course required. We will illustrate with the guinea pig example.

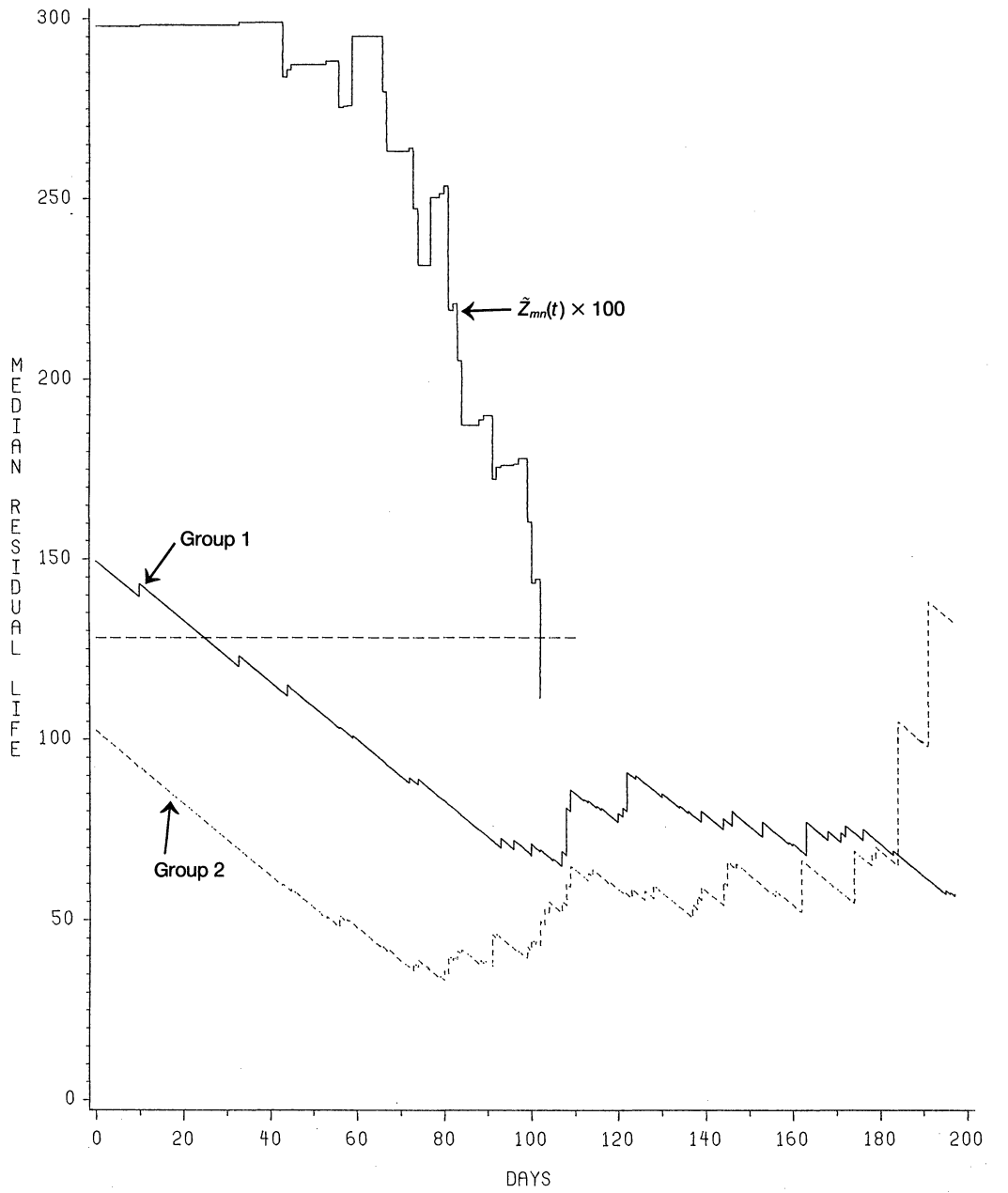

Figure 3. Median residual life of guinea pigs injected with tubercle bacilli.

The median residual life of a random variable X is just

$$\text{med}(t) = \text{median}(X - t \mid X > t),$$

i.e., replace "mean" by "median" in the definition of mean residual life. The sample median residual life is $\hat{\text{med}}(t) = \text{median}\{X_{(k)} - t, \ldots, X_{(m)} - t\}$, where $X_{(k)}$ is the first ordered X value greater than t. Figure 3 shows $\hat{\text{med}}(t)$ for the two groups of guinea pigs discussed in Section 3. When any type of censoring in the right tail occurs, the median residual life can still be calculated for all t until a censored observation appears in the first half of the remaining observations. For example, Figure 3 would not change at all if the experiment had been stopped at the end of 1 year, i.e., censoring of all values greater than $t = 365$. If the experiment had been ended at day $t = 250$, $\hat{\text{med}}_1(t)$ would be truncated at $t = 171$.

The modification of $Z_{mn}(t)$ for median residual life that we suggest is taken from Fligner and Rust (1982). Let

$$\tilde{Z}_{mn}(t) = \frac{n^{1/2}[T_{mn}(t) - \frac{1}{2}]}{\hat{\sigma}(t)},$$

where $T_{mn}(t)$ is a two-sample median test statistic and $\hat{\sigma}(t)$ is an estimate of its standard deviation. Both $T_{mn}(t)$ and $\hat{\sigma}(t)$ are based only on those observations greater than t. We chose this median test because it is an approximate size α test of H_0: $\text{med}(F) \leqslant \text{med}(G)$, not of H_0: $F = G$. In Figure 3 we display $\tilde{Z}_{mn}(t)$ ($\times$ 100 for scaling) up through $t = 102$, but not for $t > 102$ because it interferes with the $\hat{\text{med}}(t)$ graphs. Since $\tilde{Z}_{mn}(t)$ first dips below 1.282 at $t = 102$, we are 90% confident that $\text{med}_1(t) > \text{med}_2(t)$ on [0, 102). The calculation of $\hat{\sigma}(t)$ requires a few more observations than does $\hat{\text{med}}(t)$, but still no changes in $\tilde{Z}_{mn}(t)$ through $t = 147$ would have been caused by terminating the experiment at $t = 250$.

Finally, if random censoring occurs throughout the data rather than only in the right tail, then further adjustments to our estimators and test statistics are required.

5. Size, Power, and Confidence Levels

In this section we state theorems giving conditions under which the test we propose has the prescribed asymptotic size α and under which the related confidence procedures have asymptotic confidence level $1 - \alpha$. We also give conditions under which the test is consistent for all pairs of distributions (F, G) in H_a, but for no others. As mentioned in Section 1, this feature is distinctive for our test. It means that rejection of H_0 provides valid statistical evidence that H_a is true and that the mean residual life functions do not cross in the region of interest.

In all these theorems, $X_1, \ldots, X_m$ and $Y_1 \ldots, Y_n$ are independent samples with distribution functions $F(x) = \Pr(X_1 \leqslant x)$ and $G(y) = \Pr(Y_1 \leqslant y)$, respectively. We assume that $EX_1^2 < \infty$ and $EY_1^2 < \infty$. Let $\sigma_F^2(t) = \text{var}(X_1 \mid X_1 > t)$ and $\sigma_G^2(t) = \text{var}(Y_1 \mid Y_1 > t)$ with the convention that $\sigma_F(t) = e_F(t) = 0$ if $\bar{F}(t) = 0$ and $\sigma_G(t) = e_G(t) = 0$ if $\bar{G}(t) = 0$. Also we break $Z_{mn}(t)$ into two parts,

$$\begin{aligned} Z_{mn}(t) &= \frac{m^{1/2}\{[\hat{e}_F(t) - e_F(t)] - [\hat{e}_G(t) - e_G(t)]\}}{D_{mn}(t)} + \frac{m^{1/2}[e_F(t) - e_G(t)]}{D_{mn}(t)} \\ &\equiv A_{mn}(t) + B_{mn}(t), \end{aligned}$$

where $D_{mn}^2(t) = [S_m^2(t)/m(t) + S_n^2(t)/n(t)]m$.

The first theorem deals with the behavior of the statistic $Z_{mn}(t)$ at a single value of t.

Theorem 1. Let $t \geq 0$ be a value such that $\min(\bar{F}(t), \bar{G}(t)) > 0$. Then, as $\min(m, n) \to \infty$,

(i) $S_m^2(t) \to \sigma_F^2(t)$ and $S_n^2(t) \to \sigma_G^2(t)$ with probability 1,
(ii) $A_{mn}(t) \to$ N(0, 1) in distribution, where N(0, 1) is a standard normal random variable,

and thus

$$\text{(iii)} \quad \Pr(Z_{mn}(t) > z_\alpha) \to \begin{cases} 0 & \text{if } e_F(t) < e_G(t) \\ \alpha & \text{if } e_F(t) = e_G(t) \\ 1 & \text{if } e_F(t) > e_G(t) \end{cases}.$$

Part (iii) tells us that the test of H_{0t} versus H_{at} which rejects H_{0t} if and only if $Z_{mn}(t) > z_\alpha$ is asymptotically level α and consistent. But this result about a single value of t also implies that our overall test of H_0 versus H_a is asymptotically level α. To see this note that for any (F, G) in H_0, there is a $t_0 \in [T_1, T_2]$ with $e_F(t_0) \leq e_G(t_0)$. By (iii) in Theorem 1,

$$\lim_{\min(m,n)\to\infty} \Pr\left(\inf_{t\in[T_1,T_2]} Z_{mn}(t) > z_\alpha\right) \leq \lim_{\min(m,n)\to\infty} \Pr(Z_{mn}(t_0) > z_\alpha) \leq \alpha. \tag{5.1}$$

Since $\{\inf_{t\in[T_1,T_2]} Z_{mn}(t) > z_\alpha\}$ is the rejection region for the proposed test, (5.1) says that the test is asymptotically level α. Theorem 1 can also be used to show that the confidence procedures have asymptotic confidence level $1 - \alpha$. We need the additional assumption that $F(x)$ is continuous to obtain this result for procedures 2 and 3.

Theorem 2. Under the model,

$$\lim_{\min(n,m)\to\infty} \Pr(\text{Procedure 1 makes an incorrect statement}) \leq \alpha.$$

If in addition $F(x)$ is a continuous function, then

$$\lim_{\min(m,n)\to\infty} \Pr(\text{Procedure } i \text{ makes an incorrect statement}) \leq \alpha \quad \text{for } i = 2, 3.$$

It is also easy to show using Theorem 1 that lim Pr(no statement is made) = lim $\Pr(Z_{mn}(T) \leq z_\alpha) = 0$ if $e_F(T) > e_G(T)$. However, Pr(no statement is made) can be large when m and n are moderate and $e_F(T)$ exceeds $e_G(T)$ by only a small amount.

If the inequality in (5.1) were strict for every (F, G) in H_0, then the test would be too conservative and would rarely or never reject H_0. But under some additional assumptions, it can be shown that the test is asymptotically size α. That is, there are (F, G) in H_0 for which lim Pr(reject H_0) = α. Thus, the test is not too conservative. Also, we would want the test to be consistent. That is, for any (F, G) in H_a, a reasonable test should satisfy lim Pr(reject H_0) = 1. Conditions under which these properties hold are given in Theorem 3.

Theorem 3. Suppose $F(x)$ and $G(y)$ are continuous distribution functions. Suppose $\min(m, n) \to \infty$ such that $m/(m + n) \to \lambda$, $0 \leq \lambda \leq 1$. In addition assume that

$$\mathrm{E}\,|X_1|^r < \infty, \ \mathrm{E}\,|Y_1|^r < \infty \quad \text{for some } r > 2$$

and

$$\bar{F}(T_2) > 0, \ \bar{G}(T_2) > 0.$$

If (F, G) is in H_a, that is, if

$$e_F(t) > e_G(t) \quad \text{for all } t \in [T_1, T_2], \tag{5.2}$$

then

(i) $\Pr(\inf_{t\in[T_1,T_2]} Z_{mn}(t) > z_\alpha) \to 1$.

If (F, G) is an element of H_0 such that

$$e_F(t_0) = e_G(t_0) \quad \text{for some } t_0 \in [T_1, T_2]$$

and

$$e_F(t) > e_G(t) \quad \text{for all } t \in \{[T_1, T_2] - t_0\},$$

then

(ii) $\Pr(\inf_{t\in[T_1,T_2]} Z_{mn}(t) > z_\alpha) \to \alpha$.

ACKNOWLEDGEMENTS

We gratefully acknowledge a helpful discussion with Dr F. Proschan. We also thank Dr B. P. Yu for his data in Table 1, which Dr M. Witten provided, and Mr J. Cho for the computations used in Section 3.

RÉSUMÉ

La fonction d'espérance de vie moyenne d'une population donne des informations intuitives et intéressantes sur la processus de vieillissement. Nous présentons ici de nouvelles méthodes non-paramétriques pour comparer, à partir de deux échantillons indépendants, des fonctions d'espérance de vie moyenne. Ces méthodes sont assez souples pour permettre de tenir compte des relations entre fonctions et conduisent à un nouveau type d'intervalle de confiance. Nous discutons aussi de méthodes similaires pour comparer des espérances de vie moyenne.

REFERENCES

Berger, R. L. and Sinclair, D. F. (1984). Testing hypotheses concerning unions of linear subspaces. *Journal of the American Statistical Association* **79,** 158–163.

Best, D. J. and Rayner, J. C. W. (1987). Welch's approximate solution for the Behrens–Fisher problem. *Technometrics* **29,** 205–210.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

Bjerkedal, T. (1960). Acquisition of resistance in guinea pigs infected with different doses of virulent tubercle bacilli. *American Journal of Hygiene* **72,** 130–148.

Bryson, M. C. and Siddiqui, M. M. (1969). Some criteria for aging. *Journal of the American Statistical Association* **64,** 1472–1483.

Chen, Y. Y., Hollander, M., and Langberg, N. A. (1983). Tests for monotone mean residual life, using randomly censored data. *Biometrics* **39,** 119–127.

Cheng, K. F. (1985). Tests for the equality of failure rates. *Biometrika* **72,** 211–215.

Chiang, C. L. (1960). A stochastic study of the life table and its applications: II. Sample variance of the observed expectation of life and other biometric functions. *Human Biology* **32,** 221–238.

Deevey, E. S. (1947). Life tables for natural populations of animals. *Quarterly Review of Biology* **22,** 283–314.

Elandt-Johnson, R. C. and Johnson, N. L. (1980). *Survival Models and Data Analysis.* New York: Wiley.

Fligner, M. A. and Rust, S. W. (1982). A modification of Mood's median test for the generalized Behrens–Fisher problem. *Biometrika* **69,** 221–226.

Gleser, L. J. (1973). On a theory of intersection–union tests (Abstract). *Institute of Mathematical Statistics Bulletin* **2,** 233.

Guess, F. and Proschan, F. (1988). Mean residual life: Theory and applications. To appear in *Handbook of Statistics, Volume 7. Quality Control and Reliability*, P. R. Krishnaiah and C. R. Rao (eds), 215–224. Amsterdam: North-Holland.

Joe, H. and Proschan, F. (1984). Comparison of two life distributions on the basis of their percentile residual life functions. *Canadian Journal of Statistics* **12,** 91–97.

Witten, M. (1985). Reliability theoretic methods and aging: Critical elements, hierarchies, and longevity—Interpreting survival curves. In *Molecular Biology of Aging*, A. D. Woodland, A. D. Plackett, and A. Hollaender (eds), 345–360. New York: Plenum Press.

Yu, B. P., Masoro, E. J., Murata, I., Bertrand, H. A., and Lynd, F. T. (1982). Lifespan study of SPF Fisher 344 male rats fed *ad libitum* or restricted diets: Longevity, growth, lean body mass and disease. *Journal of Gerontology* **37**, 130–141.

Received February 1986; *revised June and August* 1987.

Appendix

Proof of Theorem 1. (i) Let $X_{(1)} \leq \cdots \leq X_{(m)}$ be the ordered X values. Then $S_m^2(t)$ can be written as a continuous function of $m^{-1}\sum_{i=1}^m I(X_i > t)X_i^2$, $m^{-1}\sum_{i=1}^m I(X_i > t)X_i$, $\bar{F}_m(t) = m^{-1}\sum_{i=1}^m I(X_i > t)$, and $I(X_{(m)} > t)$, each of which converges with probability 1 to the appropriate quantity. A similar result holds for $S_n^2(t)$. (ii) The central limit theorem applied to the sums $\sum_{i=1}^m I(X_i > t)[X_i - t - e_F(t)]$ and $\sum_{i=1}^n I(Y_i > t)[Y_i - t - e_G(t)]$ along with Slutsky's theorem and (i) yield the result. Note that if $n/(m+n) \not\to \lambda$, $0 < \lambda < 1$, then a subsequence argument may be used. (iii) follows from (ii) since $B_{mn}(t)$ converges in probability to $-\infty$ if $e_F(t) < e_G(t)$ and $B_{mn}(t)$ converges in probability to $+\infty$ if $e_F(t) > e_G(t)$.

Proof of Theorem 2. Let F and G be fixed distributions. All limits are as $\min(m, n) \to \infty$. For each of the three procedures if $e_F(T) \leq e_G(T)$, then by Theorem 1

$$\lim \Pr(\text{Procedure } i \text{ makes an incorrect statement}) = \lim \Pr(Z_{mn}(T) > z_\alpha) \leq \alpha.$$

Now assume $e_F(T) > e_G(T)$. For Procedure 1 define $\theta_1 = \inf\{t \geq T\colon e_F(t) \leq e_G(t)\}$. If $\theta_1 = \infty$ then Pr(Procedure 1 makes an incorrect statement) = 0. For the case $\theta_1 < \infty$ note that mean residual life functions are right-continuous. Thus, $e_F(t) - e_G(t)$ is also right-continuous and $e_F(\theta_1) \leq e_G(\theta_1)$. Procedure 1 will make an incorrect statement only if $\hat{\theta}_1 > \theta_1$. Thus, by Theorem 1,

$$\lim \Pr(\text{Procedure 1 makes an incorrect statement}) \leq \lim \Pr(\hat{\theta}_1 > \theta_1) \leq \lim \Pr(Z_{mn}(\theta_1) > z_\alpha) \leq \alpha.$$

For procedures 2 and 3, define

$$\theta_2 = \sup\{t \leq T\colon e_F(t) \leq e_G(t)\}$$

and

$$\delta = \sup\left\{d \geq 0\colon \inf_{t \in [\max(0, T-d), T+d]} [e_F(t) - e_G(t)] > 0\right\}.$$

The proof of the validity of these procedures follows as for procedure 1 except that the continuity of F is used to ensure that $e_F(\theta_2) \leq e_G(\theta_2)$ or $e_F(T-\delta) \leq e_G(T-\delta)$ if

$$\lim_{t \uparrow T-\delta} [e_F(t) - e_G(t)] \leq 0.$$

The continuity of F ensures the continuity of e_F. Also e_G, like any mean residual life function, is upper semicontinuous so we have

$$\lim_{t \uparrow t^*} [e_F(t) - e_G(t)] \geq e_F(t^*) - e_G(t^*).$$

Proof of Theorem 3. First we give two lemmas whose proofs are given in unpublished work by Berger, Boos, and Guess (Institute of Statistics Mimeo Series #1673, North Carolina State University, 1985). Let $D[T_1, T_2]$ be the space of functions on $[T_1, T_2]$ that are right-continuous and have left-hand limits (see Billingsley, 1968, Chap. 3). We often write $\inf_t$ and $\sup_t$ as shorthand notation for infimum and supremum over $t \in [T_1, T_2]$.

Lemma A1. Under the assumptions of Theorem 3,

$$A_{mn}(t) \to A(t) \text{ weakly in } D[T_1, T_2]$$

with respect to the Skorohod topology, where $A(t)$ is a mean zero Gaussian process with continuous sample paths.

Lemma A2. Under the conditions of Theorem 3 with $\lambda < 1$, we have

$$\sup_{t\in[T_1,T_2]} [D_{mn}(t) - D(t)| \quad \text{converges almost surely to 0}$$

where $D^2(t) = \sigma_F^2(t)/\bar{F}(t) + [\lambda/(1-\lambda)]\sigma_G^2(t)/\bar{G}(t)$.

Proof of Theorem 3(i). Without loss of generality we assume $\lambda < 1$, otherwise we redefine $D_{mn}(t)$ with $n^{-1/2}$ factored out. Then

$$\begin{aligned}\Pr(\inf_t Z_{mn}(t) \leqslant z_\alpha) &= \Pr(\inf_t\{A_{mn}(t) + B_{mn}(t)\} \leqslant z_\alpha)\\ &\leqslant \Pr(\inf_t A_{mn}(t) + \inf_t B_{mn}(t) \leqslant z_\alpha)\\ &\leqslant \Pr(\inf_t A_{mn}(t) < z_\alpha - n^{1/4}) + \Pr(\inf_t B_{mn}(t) \leqslant n^{1/4}).\end{aligned}$$

Now, $\inf_t(\cdot)$ is a continuous functional in $D[T_1, T_2]$ with respect to the Skorohod topology and thus Lemma A1 gives

$$\inf_t A_{mn}(t) \text{ converges in distribution to } \inf_t A(t).$$

We can show that $\inf_t A(t)$ is finite-valued with probability 1 and thus $\Pr(\inf_t A_{mn}(t) < z_\alpha - n^{1/4}) \to 0$. Using Lemma A2 and (5.2), we can show that $m^{-1/2}[\inf_t B_{mn}(t)]$ stays strictly above 0 and thus $\Pr(\inf_t B_{mn}(t) \leqslant n^{1/4}) \to 0$ and the result 3(i) follows.

Proof of Theorem 3(ii). From the proofs of Lemmas A1 and A2 we can assert that $A_{mn}(t)$ and $D_{mn}(t)$ converge jointly to $A(t)$ and $D(t)$ in $D[T_1, T_2] \times D[T_1, T_2]$ and use Skorohod's theorem to get representations with the same exact distribution and such that almost surely

$$\sup_t |A_{mn}(t) - A(t)| + \sup_t |D_{mn}(t) - D(t)| \to 0. \tag{1}$$

Let Ω_0 be the subset of the underlying probability space on which the above convergence holds. We will show that for each $\omega \in \Omega_0$

$$\inf_t Z_{mn}(t, \omega) = \inf_t \left\{A_{mn}(t, \omega) + m^{1/2}\frac{[e_F(t) - e_G(t)]}{D_{mn}(t, \omega)}\right\} \to A(t_0, \omega), \tag{2}$$

where $A(t, \omega)$ is the sample path of $A(t)$ corresponding to ω. If (2) holds, then certainly $\inf_t Z_{mn}(t)$ converges in distribution to $A(t_0)$ and $A(t_0)$ is a standard normal random variable so that

$$\Pr\left(\inf_t Z_{mn}(t) > z_\alpha\right) \to \Pr(A(t_0) > z_\alpha) = \alpha.$$

To show (2), note first that since $B_{mn}(t_0) = 0$,

$$\inf_t Z_{mn}(t, \omega) \leqslant Z_{mn}(t_0, \omega) = A_{mn}(t_0, \omega) \tag{3}$$

and thus

$$\overline{\lim_{m,n\to\infty}} \inf_t Z_{mn}(t, \omega) \leqslant \overline{\lim_{m,n\to\infty}} A_{mn}(t_0, \omega) = A(t_0, \omega) \tag{4}$$

by (1). Since $Z_{mn}(t)$ is a step function with a finite number of jumps, $\inf_t Z_{mn}(t, \omega)$ is attained for some $t_{mn} \in [T_1, T_2]$. We can show $t_{mn} \to t_0$ by a contradiction argument. We then have

$$\begin{aligned}\inf_t Z_{mn}(t, \omega) &= A_{mn}(t_{mn}, \omega) + \frac{m^{1/2}[e_F(t_{mn}) - e_G(t_{mn})]}{D_{mn}(t_{mn}, \omega)}\\ &\geqslant A_{mn}(t_{mn}, \omega)\end{aligned}$$

since the second term is nonnegative. Thus,

$$\underline{\lim_{m,n\to\infty}} \inf_t Z_{mn}(t, \omega) \geqslant \underline{\lim_{m,n\to\infty}} A_{mn}(t_{mn}, \omega) = A(t_0, \omega) \tag{5}$$

using (1) and the convergence $t_{mn} \to t_0$. Putting (4) and (5) together yields (2).