1991

M.C.J. van Pul

Software reliability and the bootstrap

Department of Operations Research, Statistics, and System Theory Report BS-R9123 September

CWI, nationaal instituut voor onderzoek op het gebied van wiskunde en informatica

CWI is the research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a non-profit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands organization for scientific research (NWO).

Software Reliability and the Bootstrap

Mark C. van Pul

CWI
P.O. Box 4079, 1009 AB Amsterdam
The Netherlands

In this paper we will show that in a general framework of counting processes, the parametric bootstrap is asymptotically consistent. Furthermore we will consider some applications in the context of software reliability.

1980 Mathematics Subject Classification: 62N05, 62M99.
Key Words and Phrases: counting process, intensity function, maximum likelihood estimation, asymptotic behaviour, (asymptotic) consistency, asymptotic normality and efficiency, S-LAN, S-regularity, Skohorod representation, parametric bootstrap, software reliability model of Jelinski-Moranda, martingale theory.

1. Introduction

A lot of important software reliability models can be formulated in terms of counting processes, counting the number of failure occurrences. In an earlier paper (Van Pul (1990)) we studied some asymptotic properties of the maximum likelihood estimation method for parametric counting process models. For a general class of intensity functions, it was proved that the maximum likelihood estimators for the model parameters are asymptotically normally distributed. A novel aspect of our approach is the fact that—in order to treat asymptotic theory—instead of increasing the time variable or the number of data as is usually the case, we will (conceptually) increase one of the model parameters itself. To illustrate the problem and to motivate our concepts, we will present here one of the oldest and most elementary software reliability models, namely that of Jelinski-Moranda (1972), as an example.

Example 1: the Jelinski-Moranda model. A computer program has been executed during a specified exposure period and the interfailure times are observed. The repairing of a fault takes place immediately after it produces a failure and no new faults are introduced with probability one.

Let N the unknown number of faults initially present in the software. Let the exposure period be $[0,\tau]$ and let $n(t)$, $t \in [0,\tau]$, denote the number of faults detected up to time t. Define $T_0 := 0$ and let T_i, $i=1,2,...,n(\tau)$, the failure time of the i-th occurring failure, while $t_i := T_i - T_{i-1}$, $i=1,2,...,n(\tau)$, denotes the interfailure time, that is the time between the i-th and the $(i-1)$-th occurring failure. Finally we define $t_{n(\tau)+1} := \tau - T_{n(\tau)}$.

In the Jelinski-Moranda model, introduced in 1972 and a few years later generalized by Musa (1975), the failure rate of the program is at any time proportional to the number of remaining faults and each fault still present makes the same contribution to the failure rate. So if $(i-1)$ faults have already been detected, the failure rate for the i-th occurring failure, λ_i, becomes:

$$\lambda_i = \phi_0 \left[N_0 - (i-1) \right], \tag{1.1}$$

where ϕ_0 is the true failure rate per fault (the occurrence rate) and N_0 is the true number of faults initially present in the software. In terms of counting processes we can write:

Report BS-R9123
ISSN 0924-0659
CWI
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

$$\lambda^{JM}(t) = \phi_0 \left[N_0 - n(t-) \right], \quad t \in [0,\tau], \tag{1.2}$$

where $\lambda(t)$, $t \in [0,\tau]$ denotes the failure rate at time t. The interfailure times $t_i, i = 1,...,n(\tau)$, are independent and exponentially distributed with parameter λ_i given by (1.1).

By using the information obtained from the test experiment one can estimate the the parameters of the underlying model. Mostly maximum likelihood estimation is used for this purpose. Now let us look at the way we will treat asymptotic behaviour. It does not make sense to let τ, the stopping time, grow to infinity. In the long run the estimate of the total number of faults will trivially be equal to the true number of faults. It makes more sense to (conceptually) increase the number of faults in the program. The idea is that then asymptotics should be relevant to the practical situation in which N_0 is large and $n(\tau)/N_0$ not close to zero or one. *(End of example 1)*

In an other paper (Van Pul (1991)) we studied—in the context of software reliability models—how fast the asymptotic behaviour of the maximum likelihood estimators appeared in practice. It turned out that convergence is rather slow and that empirical distributions of the MLE's for small N_0 tend to be very skew. This leads of course to systematic errors in the coverage percentages of confidence intervals based on the classical normal test statistic. In Van Pul (1990, 1991) also confidence intervals are constructed, based on another test statistic, namely the Wilks Likelihood Ratio Test (WLRT) statistic, which can be proved to be asymptotically χ^2-distributed (see Rao (1973)). The coverage percentages of confidence intervals based on the WLRT statistic were much better. Upperbounds based on the WLRT statistic for small N_0 (that is, when the asymptotic normal behaviour has not occurred yet) are often infinite and therefore in some sense less informative. On the other hand, the fact that the data can not exclude the possibility of constant (or even increasing) failure intensity, could be important to know in practical situations. Another idea is to apply a parametric bootstrap method on a standardized and studentized version of the approximate normal test statistic or to estimate the second order term in the Edgeworth expansion of the distribution of the MLE's (Hall (1988)).

In the next section of this paper we will introduce the general framework of counting processes and intensities as described in Van Pul (1990) and discuss how parametric bootstrap methods could be applied. In section 3 we will show that in this general setting the parametric bootstrap 'works', that is, is asymptotically consistent. The proof which makes use of the Skohorod-Dudley-Wichura almost-sure-representation theorem (see for instance Pollard (1989)), will be based on stronger versions of the concepts of LAN and regularity. In the fourth and last section of this paper we will give some applications of the developed theory in the context of software reliability. The bootstrap method is applied on simulated data according to the model of Jelinski-Moranda, as described above.

2. The general framework

Let a counting process $n(t)$ be given. Only during a specific time interval $[0,\tau]$, are jumps of the counting process $n(t)$ observed. In this paper we will assume that the intensity function associated with the counting process exists and is a member of some specified parametric family, that is:

$$\lambda(t) := \lambda(t; N, \psi), \quad t \in [0,\tau],\ N \in \mathbb{N},\ \psi \in \Psi,\ \Psi \subset \mathbb{R}^{p-1} \tag{2.1}$$

for an integer p. Let N_0 and ψ_0 be the true parameter values. Typically the parameter N_0 represents the scale or size of the problem (sometimes $N_0 = n(\infty)$), while ψ_0 is a nuisance vector parameter. We will be interested in estimation of N_0 and ψ_0 as $N_0 \to \infty$. We assume that the model is also meaningful for non-integer N. For instance the intensity function (1.2) of the Jelinski-Moranda model can be generalized to

$$\lambda^{JM}(t) = \phi \left[N - n(t-) \right] I\{n(t) < N\}, \quad t \in [0,\tau], \tag{2.2}$$

where $I\{.\}$ denotes the indicator function. As we are particularly interested in the parameter estimation when N_0 is large, we will introduce a series of counting processes $n_\nu(t)$, $t\in[0,\tau], \nu=1,2,...$ and let N conceptually increase. Let $N = N_\nu \to \infty$ for $\nu \to \infty$. By the reparametrization

$$N_\nu = \nu\gamma_\nu \tag{2.3}$$

with a dummy variable γ_ν, we can denote the associated intensity functions by

$$\lambda_\nu(t;\gamma,\psi) := \lambda(t;\nu\gamma,\psi), \quad t\in[0,\tau], \quad \gamma\in\mathbb{R}^+, \quad \psi\in\Psi, \quad \nu=1,2,.... \tag{2.4}$$

Now we consider the estimation of γ and ψ as $\nu\to\infty$. If the real-life situation has $\nu=N_0$, then $\gamma=\gamma_0=1$ and $\psi=\psi_0$. It is rather unorthodox to increase a model parameter itself, in this case N. This complication is solved by estimating γ. We will assume that the maximum likelihood estimators $(\hat{\gamma}_\nu,\hat{\psi}_\nu)$ for (γ_0,ψ_0) exist. Typically, $(\hat{\gamma}_\nu,\hat{\psi}_\nu)$ is a root of the likelihood equations

$$\frac{\partial}{\partial(\gamma,\psi)} \log L_\nu(\gamma,\psi;\tau) = 0, \quad \nu=1,2,..., \tag{2.5}$$

where the likelihood function at time t $L_\nu(\gamma,\psi;t)$ is given by (see Aalen (1978)):

$$L_\nu(\gamma,\psi;t) := \exp\left[\int_0^t \log \lambda_\nu(s;\gamma,\psi)\, dn_\nu(s) - \int_0^t \lambda_\nu(s;\gamma,\psi)\, ds\right]. \tag{2.6}$$

We define for $\nu=1,2,...$ the stochastic process $x_\nu(t)$ by:

$$x_\nu(t) := \nu^{-1} n_\nu(t), \quad t\in[0,\tau]. \tag{2.7}$$

In most practical situations, this sequence of stochastic processes converges uniformly on $[0,\tau]$ in probability to a deterministic function $x_0(t)$ as $\nu\to\infty$ (Kurtz (1983)). We assume that the counting processes n_ν are generated by associated intensity functions $\lambda_\nu(t)$, satisfying:

$$\lambda_\nu(t) = \nu\beta(t;\theta;x_\nu(t-)), \tag{2.8}$$

for an arbitrary non-negative and non-anticipating function $\beta:[0,\tau]\times\Theta\times K \to \mathbb{R}^+$ where the model-parameter $\theta=(\gamma,\psi)$ consists of the parameter of most interest γ and a nuisance parameter vector ψ. Under classical smoothness and boundedness conditions on the function β (see for instance Borgan (1984) or Van Pul (1990)), we have the following result:

Theorem 1:

(i) *Consistency of ML-estimators: With a probability tending to 1, the likelihood equations (2.5) have exactly one consistent solution $\hat{\theta}_\nu$. Moreover this solution provides a local maximum of (2.6).*

(ii) *Asymptotic normality of the ML-estimators: Let $\hat{\theta}_\nu$ be the consistent solution of the maximum likelihood equations (2.5), then*

$$\sqrt{\nu}(\hat{\theta}_\nu - \theta_0) \to \mathcal{N}(0,\Sigma^{-1}), \quad \nu\to\infty, \tag{2.9}$$

where the matrix $\Sigma=\{\sigma_{ij}(\theta_0)\}$ with for $i,j\in\{1,2,..,p\}$, $\theta\in\Theta_0$: is given by

$$\sigma_{ij}(\theta) = \int_0^\tau \frac{\frac{\partial}{\partial\theta_i}\beta(s,\theta,x_0)\frac{\partial}{\partial\theta_j}\beta(s,\theta,x_0)}{\beta(s,\theta,x_0)}\, ds. \tag{2.10}$$

(iii) *Local asymptotic normality of the model: There exist a sequence $U_\nu, \nu=1,2,...$ such that for all $h\in\mathbb{R}^p$:*

$$\log\frac{dP_{\theta_\nu}}{dP_{\theta_0}} - h^T U_\nu + \frac{1}{2}h^T\Sigma h \xrightarrow{P_{\theta_0}} 0, \quad \nu\to\infty, \tag{2.11}$$

where $\theta_\nu = \theta_0 + \nu^{-\frac{1}{2}} h$ and $U_\nu \xrightarrow{D} \mathcal{N}(0,\Sigma)$, Σ given by (2.10).

(iv) *Asymptotic efficiency of the ML-estimators:* $\hat{\theta}_\nu$ *is asymptotically efficient in the sense that the limit distribution for any other regular estimator* $\tilde{\theta}_\nu$ *for* θ_0 *satisfies:*

$$\sqrt{\nu}(\tilde{\theta}_\nu - \theta_0) \overset{D(\theta_0)}{\rightarrow} Z + Y, \tag{2.12}$$

where $Z \sim_d \mathcal{N}(0,\Sigma^{-1})$, *Z and Y independent. (For a definition of the regularity of an estimator we refer to Van der Vaart (1988).)*

This theorem and its proof can be found more thoroughly in Van Pul (1990).

Simulations of the Jelinski-Moranda model showed (see Van Pul (1990,1991) that asymptotic convergence to the normal distribution is appearing very slowly and that for values of ν not extremely big the empirical distribution functions of the components of $\hat{\theta}_\nu$ can be significantly skew. Hence, confidence intervals based on approximate normal test statistics turned out to be disappointing. One solution, which is already suggested in Van Pul (1990), is to make use of the Wilks Likelihood Ratio Test Statistic, which can be proved to be asymptotically χ^2 distributed. Another way to deal with deviations from normality is to make use of bootstrap methods. Suppose we want to construct confidence intervals for a one-dimensional real parameter θ. The concept of parametric bootstrapping in the context of software reliability consists of simulating a so called bootstrap counting process according to the failure intensity $\lambda(t,\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimator for θ. Repeating this simulation experiment, say M times, we get bootstrap estimators $\hat{\theta}_i^*, i=1,...,M$. Defining

$$G_\nu := \text{Law}_{\theta_0}\left[\sqrt{\nu}(\hat{\theta}_\nu - \theta_0)\right] \tag{2.13}$$

and

$$G_\nu^* := \text{Law}_{\hat{\theta}_\nu}\left[\sqrt{\nu}(\hat{\theta}_\nu^* - \hat{\theta}_\nu)\right], \tag{2.14}$$

we will say that the parametric bootstrap 'works' (or is asymptotically consistent) if and only if

$$\sup_{x \in \mathbb{R}} | G_\nu^*(x) - G_\nu(x) | \overset{P_{\theta_0}}{\rightarrow} 0. \tag{2.15}$$

See also Bickel & Freedman (1981) and Singh (1981). This result will be derived in the next section. Note that, as G_ν converges to a continuous distribution function, a consequence of (2.15) is that confidence intervals for θ based on G_ν^* will have asymptotically the right coverage probabilities, i.e.:

$$\mathbb{P}\left[\hat{\theta}_\nu - \nu^{-1/2}Z_\nu(1-\alpha/2) \leqslant \theta_0 \leqslant \hat{\theta}_\nu - \nu^{-1/2}Z_\nu(\alpha/2)\right] \rightarrow 1-\alpha, \quad \nu \rightarrow \infty, \tag{2.16}$$

where $Z_\nu(\alpha) := G_\nu^{*\,-1}(\alpha)$. In practice one often uses studentized versions of (2.13) and (2.14),

$$G_\nu^{ST} := \text{Law}_{\theta_0}\left[\frac{\sqrt{\nu}(\hat{\theta}_\nu - \theta_0)}{\hat{\sigma}_\nu}\right] \tag{2.17}$$

and

$$G_\nu^{ST*} := \text{Law}_{\hat{\theta}_\nu}\left[\frac{\sqrt{\nu}(\hat{\theta}_\nu^* - \hat{\theta}_\nu)}{\hat{\sigma}_\nu^*}\right], \tag{2.18}$$

expecting the second order terms of the Edgeworth expansions to be the same too (see for instance Helmers (1991)). In this paper we will determine $\hat{\sigma}_\nu$ and $\hat{\sigma}_\nu^*$ simply by substituting respectively $\hat{\theta}_\nu$ and $\hat{\theta}_\nu^*$ for θ_0 in the expected information matrix Σ, given by (2.10).

An alternative way to estimate σ consistently, is to make use of the observed information matrix

$$I_\nu(\theta,\tau) \;:=\; \frac{\partial^2}{\partial\theta^2} L_\nu(\theta,\tau). \tag{2.19}$$

3. Asymptotic consistency of the parametric bootstrap

Supposing we are in the software reliability situation, sketched in section 2, we can prove (2.15). The following two lemma's will be very useful:

Lemma 1: *Under the conditions of theorem 1, we have SLAN (strong local asymptotic normality), that is: there exist a sequence $U_\nu, \nu=1,2,...$ such that for all $h \in \mathbb{R}^p$:*

$$\log\frac{dP_{\theta_\nu}}{dP_{\theta_0}} \;-\; h^T U_\nu \;+\; \frac{1}{2} h^T \Sigma h \;\overset{P_{\theta_0}}{\rightarrow}\; 0, \quad \nu\rightarrow\infty, \tag{3.1}$$

where $U_\nu \rightarrow_d \mathcal{N}(0,\Sigma)$, Σ given by (2.6), but now with

$$\theta_\nu = \theta_0 + \nu^{-1/2} h + o(\nu^{-1/2}). \tag{3.2}$$

Lemma 2: *Under the conditions of theorem 1, asymptotic normality and SLAN imply S-regularity (strong regularity), that is:*

$$\sqrt{\nu}\left[\hat{\theta}_\nu \;-\; \theta_\nu\right] \;\overset{D(\theta_\nu)}{\rightarrow}\; N(0,\Sigma^{-1}) \tag{3.3}$$

for all sequences θ_ν of the form (3.2).

The proofs of lemma's 1 and 2 are slight modifications of the proofs of Theorem 1 (iii)&(iv), given in Van Pul (1990), and are therefore omitted here. We are now able to formulate the main result of this paper:

Theorem 2: The parametric bootstrap is asymptotically consistent.
Proof of theorem 2: The asymptotic normality of the MLE yields:

$$\lim_{\nu\rightarrow\infty} \sup_{x\in\mathbb{R}} \mid G_\nu(x) \;-\; G(x) \mid \;=\; 0, \tag{3.4}$$

where $G(x) := N(0,\Sigma^{-1})$. So to prove (2.15) it is sufficient to show that for all $\epsilon>0$:

$$\lim_{\nu\rightarrow\infty} \mathbb{P}\left[\; \sup_{x\in\mathbb{R}} \mid G_\nu^*(x) \;-\; G(x) \mid \;>\; \epsilon \;\right] \;=\; 0. \tag{3.5}$$

Defining $Z_\nu := \sqrt{\nu}(\hat{\theta}_\nu - \theta_0)$ the asymptotic normality assures that:

$$Z_\nu \;\overset{D}{\rightarrow}\; Z \tag{3.6}$$

where $\text{Law}_{\theta_0}(Z) \;=\; G \;=\; N(0,\Sigma^{-1})$. The almost-sure-representation theorem (see for instance Pollard [4]) states that there exist $\tilde{Z}_\nu \;=_d\; Z_\nu$ and a $\tilde{Z} \;=_d\; Z$ such that:

$$\tilde{Z}_\nu \;\overset{a.s.}{\rightarrow}\; \tilde{Z}. \tag{3.7}$$

As (for fixed ν) θ_0 and $\sqrt{\nu}$ are constants, Z_ν is only a function of $\hat{\theta}_\nu$. So, we can write

$$\tilde{Z}_\nu \;:=\; \sqrt{\nu}(\tilde{\theta}_\nu - \theta_0) \;\overset{a.s.}{\rightarrow}\; \tilde{Z}, \tag{3.8}$$

for some $\tilde{\theta}_\nu \in \mathbb{R}^p$, that is

$$\tilde{\theta}_\nu := \theta_0 + \nu^{-1/2}\tilde{Z} + o(\nu^{-1/2}), \quad \text{a.s.} \tag{3.9}$$

Now the S-regularity of $\hat{\theta}_\nu$ gives under $P(\tilde{\theta}_\nu)$:

$$\sqrt{\nu}\left[\hat{\theta}_\nu - \tilde{\theta}_\nu\right] \overset{D}{\to} G, \quad \text{a.s.} \tag{3.10}$$

or in other words

$$\tilde{G}_\nu^* := \text{Law}_{\tilde{\theta}_\nu}\left[\sqrt{\nu}(\hat{\theta}_\nu^* - \tilde{\theta}_\nu)\right] \overset{a.s.}{\to} G. \tag{3.11}$$

But this is to say

$$\sup_{x \in \mathbb{R}} \mid \tilde{G}_\nu^*(x) - G(x) \mid \overset{a.s}{\to} 0, \tag{3.12}$$

which implies

$$\sup_{x \in \mathbb{R}} \mid \tilde{G}_\nu^*(x) - G(x) \mid \overset{P}{\to} 0. \tag{3.13}$$

Because $\tilde{G}_\nu$ is a function of $\tilde{\theta}_\nu$ (or equivalently $\tilde{Z}_\nu$) only, and because $\tilde{Z}_\nu =_d Z_\nu$, we have $\tilde{G}_\nu^* =_d G_\nu^*$ and can conclude from (3.13) that

$$\sup_{x \in \mathbb{R}} \mid G_\nu^*(x) - G(x) \mid \overset{P_{\theta_0}}{\to} 0. \tag{3.14}$$

So we have derived (3.5) and theorem 2 is proved completely. □

Remark 1: The result of theorem 2 also holds for studentized versions of the parametric bootstrap.

4. Some applications

Recently we made the beginning with the study of the behaviour of the ML-estimators in practice, computed from data simulated according to the Jelinski-Moranda model. We compared the estimated coverage probabilities of confidence intervals for $\hat{N}$ based on the approximate normal and on the Wilks likelihood ratio test statistic. From the simulation results, presented by Van Pul (1991), we can draw several conclusions. Firstly, they confirm the asymptotic theory as derived in Van Pul (1990) for large N_0 ($\sim$5000). For medium large N_0 ($\sim$500) the simulation results show that the asymptotic theory does not occur yet, but also shows that the Wilks likelihood ratio test statistic provides confidence intervals which are much better. Finally, for small N_0 ($\sim$50) we found that confidence intervals based on the WLRT statistic still have reasonable estimated coverage probabilities; the upperbounds of those confidence intervals can be extremely large and sometimes even be infinite, however. There is a clear interpretation of the model and its likelihood when $N_0 = \infty$; we are then namely in the special case of a Poisson process with constant intensity.

In this section we will therefore concentrate on the case that N_0 is small (say 50) and compare estimators obtained from parametric bootstrapping methods with the estimators (based on the approximate normal and on the WLRT test statistic) constructed earlier. For a full description of the simulation experiments according to the Jelinski-Moranda model and for remarks on the construction of confidence intervals based on the approximate normal and WLRT statistic, we refer to Van Pul (1991). To construct bootstrapped confidence intervals we used a set consisting of 1000 ML-estimator-pairs $(\hat{N}_i, \hat{\phi}_i)$, $i = 1...1000$, simulated according to the Jelinski-Moranda model with $N_0 = 50$ and $\phi_0 = 1$. For each estimator-pair $(\hat{N}_i, \hat{\phi}_i)$ we now constructed 1000 bootstrapped estimator-pairs $(\hat{N}_{ij}^*, \hat{\phi}_{ij}^*)$, $j = 1...1000$ and used formulas analogous to (2.16) to construct one- and two-sided confidence intervals for N_0.

In table 1 we compared the estimated coverage probabilities of upperbounds (for various levels of confidence α) constructed with use of the approximate normal test statistic (AN), the Wilks likelihood ratio test statistic (WLR), (non-studentized) bootstrapping (B) and studentized bootstrapping (SB). An asterisk (*) denotes that upperbounds were sometimes infinite.

Analogously in table 2 hitting and miss percentages of two-sided confidence intervals are given. In each entry (x_1:x_2:x_3) in table 2, x_1 corresponds with the miss-under percentage (that is the percentage of the confidence intervals with upperbound smaller than N_0), x_2 denotes the hitting percentage and x_3 represents the miss-over percentage. Again an asterisk (*) denotes that upperbounds were sometimes infinite.

$1-\alpha$	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
AN	0.35	0.39	0.44	0.47	0.51	0.55	0.59	0.63	0.69	0.75
W	0.37	0.39	0.44	0.49	0.55	0.62*	0.68*	0.75*	0.81*	0.89*
B	0.44	0.46	0.48	0.51	0.53	0.57	0.58	0.63	0.67	0.71
SB	0.47	0.51	0.55	0.62	0.66	0.72	0.76	0.80	0.86	0.90

Table 1: Hitting percentages of one-sided confidence intervals.

$1-\alpha$	0.50	0.55	0.60	0.65	0.70
AN	0.45:0.55:0.00	0.43:0.57:0.00	0.41:0.59:0.00	0.38:0.62:0.00	0.37:0.63:0.00
W	0.39:0.52:0.09*	0.35:0.58:0.07*	0.32:0.63:0.05*	0.28:0.68:0.04*	0.25:0.72:0.03*
B	0.43:0.31:0.26	0.42:0.34:0.24	0.41:0.37:0.22	0.40:0.40:0.20	0.38:0.45:0.17
SB	0.28:0.45:0.27	0.27:0.47:0.26	0:25:0.52:0.23	0.22:0.56:0.22	0.21:0.58:0.21

$1-\alpha$	0.75	0.80	0.85	0.90	0.95
AN	0.34:0.66:0.00	0.31:0.69:0.00	0.28:0.72:0.00	0.25:0.75:0.00	0.22:0.78:0.00
W	0.22:0.76:0.02*	0.19:0.80:0.01*	0.14:0.85:0.01*	0.11:0.89:0.00*	0.07:0.93:0.00*
B	0.36:0.49:0.15	0.34:0.53:0.13	0.33:0.56:0.11	0.30:0.61:0.09	0.26:0.68:0.06
SB	0.18:0.62:0.20	0.16:0.66:0.18	0.12:0.71:0.17	0.10:0.75:0.15	0.07:0.81:0.12

Table 2: Hitting and miss percentages of two-sided confidence intervals.

From table 1 and table 2 we may conclude the following:

1 Non-studentized bootstrap results are poor.

2 Studentizing has proved to be very fruitful.

3 Hitting percentages of one-sided confidence intervals based on a studentized bootstrap method, are comparable to those based on the WLRT statistic (and much better than those based on the approximate normal test statistic).

4 Hitting percentages of two-sided confidence intervals based on a studentized bootstrap method, are comparable to those based on the approximate normal test statistic (and significantly worse than those based on the WLRT statistic).

5 (Studentized) bootstrap upperbounds are, in contrary to the WLRT ones, always finite.

6 (Studentized) bootstrap results have the advantage (in comparison with the approximate normal and WLRT results) that the miss percentages are more symmetric distributed.

Except perhaps for 4, all observations confirm the developed theory. It is not clearly understood, however, why the studentized bootstrap results for two-sided confidence intervals are less promising than expected.

Acknowledgement

This research was carried out under a grant of the Netherlands Foundation for Applied Technology (STW).

References

Aalen,O.O. (1978), Non-parametric inference for a family of counting processes. *Annals of Statistics* **6** 701-726.

Bickel,P.J. & Freedman D. (1981), Some asymptotic theory for the bootstrap. *Annals of Statistics* **9** 1196-1217.

Borgan,Ø. (1984), Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics* **1** 1-16.

Hall,P. (1988), Theoretical comparison of bootstrap confidence intervals. *Annals of Statistics* Volume **16,** No. 3, 927-953.

Helmers,R. (1991), On the edgeworth expansion and the bootstrap approximation for a studentized U-statistic. *Annals of Statistics* Volume **19,** No. 1, 470-484.

Jelinski,Z. & Moranda, P. (1972), Software Reliability Research. *Statistical Computer Performance Evaluation* 465-484.

Kurtz,T.G. (1983), Gaussian approximations for Markov chains and counting processes. *Bulletin of the International Statistical Institute* **50** 361-375.

Moek,G. (1983), *Software Reliability Models on Trial: Selection, Improved Estimation, and Practical Results.* Report MP 83059 U, National Aerospace Laboratory NLR, Amsterdam.

Moek,G. (1984), Comparison of some software reliability models for simulated and real failure data. *International Journal of Modelling & Simulation,* Volume 4, No. 1, 29-41.

Musa,J.D. (1975), A theory of software reliability and its application. *IEEE Transactions on Software Engineering,* **3,** 312-327.

Pollard,D. (1989), *Convergence of stochastic Processes.* Springer Series in Statistics, New York.

van Pul, M.C. (1990), *Asymptotic Properties of Statistical Models in Software Reliability.* Report BS-R9011, Centre for Mathematics and Computer Science, Amsterdam. (To appear in the SJS (1992).)

van Pul, M.C. (1991), *Simulations on the Jelinski-Moranda Model of Software Reliability.* Report BS-R9122, Centre for Mathematics and Computer Science, Amsterdam.

Rao,C.R. (1973), *Linear Statistical Interference and its Applications.* Wiley, New York.

Singh,K. (1981), On asymptotic accuracy of Efron's bootstrap. *Annals of Statistics* **9** 1187-1195.

van der Vaart, A.W. (1987), *Statistical Estimation in Large Parameter Spaces.* Tract 44, Centre for Mathematics and Computer Science, Amsterdam.