

# CONFIDENCE INTERVALS FOR $N$ IN THE EXPONENTIAL ORDER STATISTICS PROBLEM

Mark Finkelstein, Howard G. Tucker, and Jerry Alan Veeh

University of California, Irvine  
Irvine, California 92697-3875

and

University of California, Irvine  
Irvine, California 92697-3875

and

Auburn University  
Auburn, Alabama 36849-5310

*Key Words:* exponential order statistics; Jelinski-Moranda model; software reliability.

## ABSTRACT

A simplified proof of the basic properties of the estimators in the Exponential Order Statistics (Jelinski-Moranda) Model is given. The method of constructing confidence intervals from hypothesis tests is applied to find conservative confidence intervals for the unknown parameters in the model.

## 1. INTRODUCTION

Testing of a software product has revealed 10 bugs. How many more bugs are there in the software?

A team of explosives experts has removed 112 landmines from a known minefield. How many landmines did the team fail to find?

These questions fit into the general framework of what is called the order statistics problem. In product reliability terms the problem is formulated more precisely as follows. A product contains  $N$  defects, where  $N \geq 0$  is unknown. Associated with the  $i$ th defect is a time  $D_i$  at which this defect will be observed once the product is in operation. The product is placed into operation at time 0 and the operation of the product will be observed until some fixed time  $T > 0$ . Defects which are observed are assumed to

be immediately reported. The number  $R$  of defects which are reported up to time  $T$  as well as the times  $T_1 \leq T_2 \leq \dots \leq T_R$  at which defects are reported are observed. The random variables  $T_1, \dots, T_R$  are thus the first  $R$  order statistics of the random variables  $D_1, \dots, D_N$ . In this study the random variables  $D_1, \dots, D_N$  are assumed to be independent and identically distributed random variables each having the exponential distribution with rate parameter  $\lambda > 0$  which is unknown; under this assumption the problem is known as the exponential order statistics problem. The objective is to estimate  $N$  and  $\lambda$  and find confidence intervals for each.

The exponential order statistics problem was first put forward as a model for software reliability questions by Jelinski and Moranda (1972) and is often referred to as the Jelinski-Moranda model. The contributions of this paper are twofold. First, a simplified presentation of the derivation and properties of the maximum likelihood estimator of  $N$  is given. This derivation is based only on elementary techniques and does not depend on the notion of total positivity. Second, the method of constructing confidence intervals from hypothesis tests is used to find conservative confidence intervals for both  $N$  and  $\lambda$ . The advantage of this over asymptotic techniques is that the confidence level of the derived confidence intervals can be proved to exceed the nominal level. No appeal to asymptotics is required.

## 2. MAXIMUM LIKELIHOOD ESTIMATION

This section and the next one are devoted to deriving some basic properties of the maximum likelihood estimators of  $N$  and  $\lambda$ . The arguments given here are straightforward and self-contained; the presentation does not depend on the notion of total positivity, as for example, in Joe and Reid (1985), who obtained similar results.

**Theorem 1.** *In the notation above, denote  $S = \sum_{j=1}^R T_j$ . On the event*

$[R \geq 1, S < T(R + 1)/2]$  the maximum likelihood estimators of  $N$  and  $\lambda$  exist and are unique, except on a sub-event of probability 0. On the event  $[S \geq T(R + 1)/2]$  there is no maximum likelihood estimator of  $N$  or  $\lambda$ .

The proof of the theorem also provides an effective method for computing the maximum likelihood estimators.

To begin, the joint density of the observations for  $N > 0$  is given by

$$f_{R, T_1, \dots, T_R}(r, t_1, \dots, t_r) = r! \binom{N}{r} \lambda^r e^{-\lambda((N-r)T + \sum_{j=1}^r t_j)}$$

for  $1 \leq r \leq N$  and  $0 < t_1 < \dots < t_r < T$ . If  $N = 0$  the joint density is degenerate. Denote by  $s = \sum_{j=1}^r t_j$  and  $S = \sum_{j=1}^R T_j$ . The Factorization Theorem shows that  $R$  and  $S$  are jointly sufficient for  $N$  and  $\lambda$ .

The parameter space here is  $\{(N, \lambda) : N \in \{0, 1, 2, \dots\}, \lambda > 0\}$ , which is discrete in  $N$  and continuous in  $\lambda$ . In order to find the joint maximum likelihood estimates of  $N$  and  $\lambda$  the location of the maximum of the joint density over this parameter space must be found. As will be seen, in certain situations there will be no maximum likelihood estimates, while in others there will be multiple maximum likelihood estimates. This necessitates a detailed discussion of several cases.

If  $r = 0$  the unique maximum likelihood estimate of  $N$  is 0 and any positive real number is a maximum likelihood estimate of  $\lambda$ .

Suppose  $r \geq 1$ . Calculus shows that for each fixed value of  $N \geq 1$  the joint density is maximized as a function of  $\lambda$  at the value  $r/((N - r)T + s)$ . After substitution of this value for  $\lambda$  in the joint density, a function of the single variable  $N$  is obtained. The values of  $N$ , if any, which maximize this function are then the maximum likelihood estimates of  $N$ . The corresponding values of  $r/((N - r)T + s)$  are the maximum likelihood estimates of  $\lambda$ . Denote

$$L(N, r, s) = \sum_{j=0}^{r-1} \log(N - j) - r \log((N - r)T + s) - r + r \log r$$

as the value of the log likelihood after this substitution. The objective is to find the value(s) of  $N$  which maximize this function.

Viewing the discrete parameter  $N$  as a continuous variable which takes values  $N \geq r$  provides some insight into the value of the maximum likelihood estimator. Simple computation shows that  $\frac{\partial}{\partial N} L(N, r, s) = 0$  if and only if

$$\sum_{j=0}^{r-1} \frac{1}{N-j} = \frac{rT}{(N-r)T + s}.$$

This equation will be referred to as the **continuous likelihood equation**.

The first claim is that there is at most one solution of the continuous likelihood equation. To prove this claim denote by  $\mathcal{S}(N, r)$  the unique real number that satisfies

$$\sum_{j=0}^{r-1} \frac{1}{N-j} = \frac{rT}{(N-r)T + \mathcal{S}(N, r)}.$$

Implicit differentiation of this expression with respect to  $N$  shows that

$$\sum_{j=0}^{r-1} \frac{-1}{(N-j)^2} = \frac{-rT(T + \frac{\partial}{\partial N} \mathcal{S}(N, r))}{((N-r)T + \mathcal{S}(N, r))^2},$$

which, upon using the defining equation for  $\mathcal{S}(N, r)$ , becomes

$$\sum_{j=0}^{r-1} \frac{1}{(N-j)^2} = \frac{1}{rT} \left( \sum_{j=0}^{r-1} \frac{1}{(N-j)} \right)^2 \left( T + \frac{\partial}{\partial N} \mathcal{S}(N, r) \right).$$

Further algebraic simplification gives

$$\frac{1}{T} \frac{\partial}{\partial N} \mathcal{S}(N, r) = \frac{\frac{1}{r} \sum_{j=0}^{r-1} \frac{1}{(N-j)^2} - \left( \frac{1}{r} \sum_{j=0}^{r-1} \frac{1}{N-j} \right)^2}{\left( \frac{1}{r} \sum_{j=0}^{r-1} \frac{1}{N-j} \right)^2}.$$

The numerator of the right hand side is the variance of a discrete random variable which takes the values  $1/N, 1/(N-1), \dots, 1/(N-r+1)$  with equal probability. Hence the right hand side is positive and  $\frac{\partial}{\partial N} \mathcal{S}(N, r) > 0$

for  $N > r$ . Now suppose that  $s$  and  $r$  are fixed. If  $N_1$  and  $N_2$  are two solutions of the continuous likelihood equation then  $\mathcal{S}(N_1, r) = \mathcal{S}(N_2, r)$ . Hence  $N_1 = N_2$  since  $\mathcal{S}$  is strictly increasing. This proves the claim.

Denote by  $\tilde{N}(r, s)$  the value of  $N$  (if any) that solves the continuous likelihood equation. Note that  $\tilde{N}(r, s)$  need not be an integer.

The function  $\mathcal{S}(N, r)$  introduced in proof of the first claim is very useful in understanding the solution of the continuous likelihood equation. Since  $\mathcal{S}$  is increasing,  $\mathcal{S}(\infty, r) = \lim_{N \rightarrow \infty} \mathcal{S}(N, r)$  must exist. As will be seen,  $\mathcal{S}(\infty, r)$  identifies the region in which there is no maximum likelihood estimate of  $N$ .

The second claim is that  $\mathcal{S}(\infty, r) = T(r+1)/2$ . To prove this claim first write the defining equation for  $\mathcal{S}(N, r)$  as  $\frac{1}{r} \sum_{j=0}^{r-1} \frac{1}{N-j} = \frac{1}{(N-r)+\mathcal{S}(N, r)/T}$ . Since an average is at least as large as the smallest term being averaged, this equation yields the inequality  $\frac{1}{N} \leq \frac{1}{(N-r)+\mathcal{S}(N, r)/T}$ . This in turn shows that  $\mathcal{S}(N, r) \leq rT$ . To refine this bound, rewrite each of the terms in the continuous likelihood equation as  $\frac{1}{N-j} = \frac{1}{N} \frac{1}{1-\frac{j}{N}} = \sum_{k=0}^{\infty} j^k / N^{k+1}$  and similarly  $\frac{r}{N-r+\mathcal{S}(N, r)/T} = \sum_{k=0}^{\infty} r(r - \mathcal{S}(N, r)/T)^k / N^{k+1}$ . (The preliminary bound  $\mathcal{S}(N, r) \leq rT$  together with the fact that  $r < N$  (eventually) shows that these series expressions are convergent.) After substitution and simplification this leads to the equation  $\frac{r(r-1)}{2N^2} + \sum_{j=0}^{r-1} \sum_{k=2}^{\infty} j^k / N^{k+1} = \frac{r(r-\mathcal{S}(N, r)/T)}{N^2} + \sum_{k=2}^{\infty} r(r - \mathcal{S}(N, r)/T)^k / N^{k+1}$ . Multiply both sides by  $N^2$  and then let  $N \rightarrow \infty$  to conclude that

$$\frac{r(r-1)}{2} = r(r - \mathcal{S}(\infty, r)/T)$$

which upon simplification finishes the proof of the claim.

The third claim is that if  $s \geq \mathcal{S}(\infty, r)$  then there is no maximum likelihood estimate of  $N$ . Compute the derivative of the log likelihood with respect to  $N$  to obtain  $\frac{\partial}{\partial N} L(N, r, s) = \sum_{j=0}^{r-1} \frac{1}{N-j} - \frac{rT}{(N-r)T+s}$ . This expression is an increasing function of  $s$  which is 0 when  $s = \mathcal{S}(N, r)$ . Thus when

$s \geq \mathcal{S}(\infty, r) > \mathcal{S}(N, r)$ , the derivative is positive for all  $N$  and the likelihood does not attain its maximum value.

The fourth claim is that if  $s \leq \mathcal{S}(r, r)$  then the unique maximum likelihood estimate of  $N$  is  $r$ . To prove this claim notice that if  $s < \mathcal{S}(N, r)$  then the derivative of the log likelihood with respect to  $N$  is negative. Since  $\mathcal{S}(N, r)$  is increasing in  $N$ , if  $s \leq \mathcal{S}(r, r)$  then  $s \leq \mathcal{S}(N, r)$  for all  $N$ . Thus the derivative of the log likelihood is negative for all  $N \geq r$ . This means that the log likelihood is maximized at  $N = r$ , proving the claim.

Notice that since an average of numbers is bounded above by the largest of those numbers,  $\mathcal{S}(r, r) \geq T$ . Thus by claim 4 if  $s \leq T$  the unique maximum likelihood estimate of  $N$  is  $r$ .

The remaining cases in which the maximum likelihood estimate of  $N$  must be found are those in which  $\mathcal{S}(r, r) < s < \mathcal{S}(\infty, r)$ . For  $s$  in this interval  $\tilde{N}(r, s)$  exists and is unique because  $\tilde{N}(r, s)$  (as a function of  $s$ ) is the inverse of the function  $\mathcal{S}(N, r)$  (as a function of the continuous variable  $N$ ). This observation shows that for  $s$  in this interval either  $[\tilde{N}(r, s)]$  or  $[\tilde{N}(r, s)] + 1$  or both are maximum likelihood estimates for  $N$ . Here  $[x]$  is the greatest integer function.

The preceding discussion shows that the maximum likelihood estimator can be computed using what might be called the ratio method. To describe the method, return now to the situation in which the parameter  $N$  is discrete. Denote by  $\rho(N, r, s)$  the ratio of the value of the likelihood at  $N + 1$  to the value of the likelihood at  $N$ . Simple algebra shows

$$\rho(N, r, s) = \frac{N + 1}{N - r + 1} \left( \frac{(N - r)T + s}{(N - r + 1)T + s} \right)^r.$$

The analysis above shows that, as a function of  $N$ ,  $\rho(N, r, s) \geq 1$  until the ratio falls below 1 at either  $N = [\tilde{N}(r, s)]$  or  $N = [\tilde{N}(r, s)] + 1$  and then remains below 1 for all larger values of  $N$ . Hence if  $\rho([\tilde{N}(s, r)], r, s) = 1$  then both  $[\tilde{N}(s, r)]$  and  $[\tilde{N}(s, r)] + 1$  are maximum likelihood estimators of  $N$ .

If  $\rho([\tilde{N}(s, r)], r, s) > 1$  then  $[\tilde{N}(s, r)] + 1$  is the unique maximum likelihood estimator of  $N$ . Finally, if  $\rho([\tilde{N}(s, r)], r, s) < 1$  then  $[\tilde{N}(s, r)]$  is the unique maximum likelihood estimator of  $N$ .

### 3. UNIQUENESS OF THE MLE

The computation of the last section shows that the maximum likelihood estimator of  $N$  need not be unique. In this section the non-uniqueness will be shown to occur only on an event of probability 0.

Denote by  $B = \{(r, s) : r \geq 1, 0 < s < rT, \rho([\tilde{N}(r, s)], r, s) = 1\}$ . Because for each fixed  $r \geq 1$  the conditional distribution of  $S$  given  $R = r$  is absolutely continuous, the almost sure uniqueness of the maximum likelihood estimator of  $N$  will be proved if the set  $B$  is shown to be countable. To this end, define a mapping from  $B$  into the integers by  $(r, s) \mapsto 2^r 3^{[\tilde{N}(r, s)]}$ . The claim is that this mapping is injective. Suppose  $(r, s)$  and  $(r', s')$  are in  $B$  and that  $2^r 3^{[\tilde{N}(r, s)]} = 2^{r'} 3^{[\tilde{N}(r', s')]}.$  Then  $r = r'$  and  $[\tilde{N}(r, s)] = [\tilde{N}(r', s')]$  by unique factorization. Simple analysis of the formula for  $\rho$  shows that  $\rho$  is strictly increasing in  $s$  for each fixed  $N$  and  $r$ . Since by definition of  $B$ ,  $\rho([\tilde{N}(r, s)], r, s) = 1 = \rho([\tilde{N}(r', s')], r', s')$  the conclusion that  $s = s'$  follows. Thus the mapping is injective and  $B$  is countable.

### 4. ASYMPTOTIC BEHAVIOR OF THE MLE

The behavior of the maximum likelihood estimators of  $N$  and  $\lambda$  will now be studied as the length of the observation period becomes infinite. In this discussion the maximum likelihood estimator of  $N$  is defined to be infinity in those cases in which  $S \geq T(R + 1)/2$ . The corresponding value of the maximum likelihood estimator of  $\lambda$  is taken to be 0. As was shown above, the situation in which there are two maximum likelihood estimates can be neglected.

**Theorem 2.** *The maximum likelihood estimator of  $N$  converges almost*

surely to  $N$  as  $T \rightarrow \infty$ . The maximum likelihood estimator of  $\lambda$  converges almost surely to  $N/\sum_{j=1}^N D_j$  as  $T \rightarrow \infty$ .

**proof.** As seen above, the maximum likelihood estimators are functions of the statistics  $R = R_T$  and  $S = S_T$ . In terms of the original unobservable random variables  $D_1, \dots, D_N$

$$R_T = \sum_{j=1}^N 1_{[D_j \leq T]}$$

and

$$S_T = \sum_{j=1}^N D_j 1_{[D_j \leq T]}.$$

Since the random variables  $D_1, \dots, D_N$  are each finite with probability 1 each of the indicators in these expressions converges upward to the constant function 1 as  $T \rightarrow \infty$ , with probability 1. Thus  $R_\infty = \lim_{T \rightarrow \infty} R_T = N$  and  $S_\infty = \lim_{T \rightarrow \infty} S_T = \sum_{j=1}^N D_j$  both exist almost surely.

In these terms the comment following the fourth claim above shows that if  $S_T \leq T$  then the unique maximum likelihood estimate of  $N$  is  $R_T$ . Let  $\Omega_0$  be the event of probability 1 one which  $R_\infty$  and  $S_\infty$  both exist. Given  $\omega \in \Omega_0$  there is a  $T_0 = T_0(\omega)$  so that for any  $T > T_0$ ,  $R_T = R_\infty$  and  $S_T = S_\infty$ . If  $T > \max\{T_0, S_\infty\}$  then  $S_T < T$  and the unique maximum likelihood estimator of  $N$  is  $R_T = R_\infty = N$ . This proves the strong consistency of the maximum likelihood estimator of  $N$ .

The same argument shows that the maximum likelihood estimator of  $\lambda$  converges almost surely to  $R_\infty/S_\infty = N/\sum_{j=1}^N D_j$ . Thus the maximum likelihood estimator of  $\lambda$  is not strongly consistent. ■

## 5. CONFIDENCE INTERVALS FOR $N$ -AN OUTLINE

Confidence intervals for  $N$  and  $\lambda$  are found by controlling the  $p$ -values of certain hypothesis tests. This method is described in Bickel and Doksum



(1977), as well as in Kendall and Stuart (1987). The method was successfully applied in Finkelstein, Tucker, and Veeh (1998).

As motivation for the method, suppose the null hypothesis  $H_0 : N = N_0$  is to be tested against the alternative  $H_1 : N > N_0$  at level of significance  $\alpha$ . Suppose the test statistic for conducting this test is an observable random variable  $X$  and the null hypothesis is rejected if  $X$  is too large. Let  $x$  be the (fixed) observed value of  $X$ . One concludes that  $X$  was too large if  $\sup_{\lambda > 0} P_{N_0, \lambda}[X \geq x] \leq \alpha$ .

This hypothesis testing technique suggests the following method of finding a one sided confidence interval for  $N$ . Let  $N_0$  be the largest value of  $i$  so that  $\sup_{\lambda > 0} P_{i, \lambda}[X \geq x] \leq \alpha$ . Then the interval  $[N_0, \infty)$  is an intuitively reasonable choice of a confidence interval for  $N$ . Note that  $N_0$  depends on the observed value  $x$  of  $X$  and therefore is a random variable.

The presence of the nuisance parameter  $\lambda$  makes direct application of this method unwieldy. Instead a two stage procedure is used which embodies some of the same ideas. Here  $R$  has a binomial distribution based on  $N$  independent trials each having success probability  $1 - e^{-\lambda T}$ . If  $\lambda$  were known, the above technique could be applied easily to obtain a confidence interval for  $N$ . Since  $\lambda$  is unknown, a confidence interval will first be found for  $\lambda$  and the endpoints of this confidence interval will be used to set limits on the success probability for  $R$ . The above technique will then be used to get a confidence interval for  $N$ . This procedure is similar to Barnard's (1984) solution of the Behrens-Fisher problem.

## 6. CONFIDENCE INTERVALS FOR BINOMIAL $N$

For the first step in the two stage procedure outlined above, the method of constructing confidence intervals from hypothesis tests will be applied to find confidence intervals for the unknown number  $N$  of independent Bernoulli trials when the success probability  $p$ ,  $0 < p < 1$ , on each trial is known. Here

$X$  will denote the observed number of successes observed in the  $N$  trials. Also  $P_i[E]$  will denote the probability of the event  $E$  computed when the number of Bernoulli trials is assumed to be  $i$ .

First, a one sided confidence interval for  $N$  will be found. To be precise, a statistic  $\mathcal{N}^L(X, p)$  will be defined with the property that  $P_N[\mathcal{N}^L(X, p) \leq N] \geq 1 - \alpha$  for all  $N \geq 0$ .

Define a function  $\mathcal{N}^L$  with domain  $\{0, 1, 2, \dots\}$  by setting  $\mathcal{N}^L(0, p) = 0$ . For  $j \geq 1$  let  $\mathcal{N}^L(j, p)$  be the largest integer  $i$  satisfying  $P_i[X \geq j] \leq \alpha$ . This inequality has a unique maximal solution since  $P_i[X \geq j]$  is a non-decreasing function of  $i$  which has the value 0 at  $i = 0$  (by convention) and limit 1 as  $i \rightarrow \infty$ . Notice that  $0 = \mathcal{N}^L(0, p) \leq \mathcal{N}^L(1, p) \leq \mathcal{N}^L(2, p) \dots$ . Observe also that  $\mathcal{N}^L(j, p)$  is a decreasing function of  $p$  for each fixed  $j$ .

**Theorem 3.** *In the notation above,  $P_N[\mathcal{N}^L(X, p) \leq N] \geq 1 - \alpha$  for each  $N \geq 0$ .*

**proof.** The theorem will be proved by showing that  $P_N[N < \mathcal{N}^L(X, p)] \leq \alpha$  for each  $N \geq 0$ . The proof proceeds by considering various cases. First suppose  $0 = \mathcal{N}^L(0, p) \leq N < \mathcal{N}^L(1, p)$ . For such  $N$

$$[N < \mathcal{N}^L(X, p)] = [X \geq 1].$$

Thus

$$P_N[N < \mathcal{N}^L(X, p)] = P_N[X \geq 1] \leq P_{\mathcal{N}^L(1, p)}[X \geq 1] \leq \alpha$$

where the next to last inequality follows from the fact that  $P_N[X \geq 1]$  is an increasing function of  $N$  with  $P_{\mathcal{N}^L(1, p)}[X \geq 1] \leq \alpha$ . Next consider the case  $\mathcal{N}^L(1, p) \leq N < \mathcal{N}^L(2, p)$ . For  $N$  in this interval  $[N < \mathcal{N}^L(X, p)] = [X \geq 2]$ , so  $P_N[N < \mathcal{N}^L(X, p)] = P_N[X \geq 2] \leq \alpha$  since  $P_N[X \geq 2]$  is an increasing function of  $N$  with  $P_{\mathcal{N}^L(2, p)}[X \geq 2] \leq \alpha$ . A similar argument holds whenever  $\mathcal{N}^L(j - 1, p) \leq N < \mathcal{N}^L(j, p)$  for some  $j$ . Observe that  $\mathcal{N}^L(j, p) \geq j - 1$ , so

that  $\lim_{j \rightarrow \infty} \mathcal{N}^L(j, p) = \infty$ . Thus each  $N$  will fall into some such interval. This concludes the proof of the theorem. ■

The problem of finding a two sided confidence interval will now be examined. Let  $\mathcal{N}^L$  be defined as above. Define a function  $\mathcal{N}^U$  with domain  $\{0, 1, 2, \dots\}$  as follows. Let  $\mathcal{N}^U(j, p)$  be the smallest integer  $i$  satisfying  $P_i[X \leq j] \leq \alpha$ . This inequality has a smallest solution since  $P_i[X \leq j]$  is a non-increasing function of  $i$  which has the value 1 at  $i = j$  and limit 0 as  $i \rightarrow \infty$ . Notice that  $\mathcal{N}^U(0, p) \leq \mathcal{N}^U(1, p) \leq \mathcal{N}^U(2, p) \dots$ . Note too that  $\mathcal{N}^U(j, p)$  is a decreasing function of  $p$  for each fixed  $j$ . The following theorem has a proof which is similar to the proof of Theorem 3.

**Theorem 4.** *In the notation above,  $P_N[N \leq \mathcal{N}^U(X, p)] \geq 1 - \alpha$  for each  $N \geq 0$ .*

In order to obtain two sided confidence intervals it is important to establish an order relation between  $\mathcal{N}^L(j, p)$  and  $\mathcal{N}^U(j, p)$ . It is clear that  $\mathcal{N}^L(0, p) = 0 \leq \mathcal{N}^U(0, p)$ . For  $j \geq 1$  note that  $P_i[X \leq j - 1] = 1 - P_i[X \geq j]$  for all  $i$ . Replacing  $i$  by  $\mathcal{N}^L(j, p)$  gives  $P_{\mathcal{N}^L(j, p)}[X \leq j - 1] \geq 1 - \alpha > \alpha$  for  $\alpha < 1/2$ . This shows that  $\mathcal{N}^L(j, p) < \mathcal{N}^U(j - 1, p)$  and hence that  $\mathcal{N}^L(j, p) < \mathcal{N}^U(j, p)$ .

**Theorem 5.** *In the notation above,  $P_N[\mathcal{N}^L(X, p) \leq N \leq \mathcal{N}^U(X, p)] \geq 1 - 2\alpha$  for each  $N \geq 0$ .*

**proof.** The theorem is proved by noting that  $1 - P_N[\mathcal{N}^L(X, p) \leq N \leq \mathcal{N}^U(X, p)] = P_N[N < \mathcal{N}^L(X, p)] + P_N[\mathcal{N}^U(X, p) < N] \leq \alpha + \alpha = 2\alpha$  by the preceding two theorems. ■

In the present setting suppose that a confidence interval  $\Lambda^L(R, S) \leq \lambda \leq \Lambda^U(R, S)$  has been found. The resulting confidence interval for  $N$  will then be  $\mathcal{N}^L(R, 1 - e^{-\Lambda^U(R, S)T}) \leq N \leq \mathcal{N}^U(R, 1 - e^{-\Lambda^L(R, S)T})$ . This procedure will be shown to have the correct confidence level below. First though, a

confidence interval for  $\lambda$  must be found. A technical result concerning the behavior of tail probabilities is needed first.

## 7. MONOTONICITY OF TAIL PROBABILITIES

Direct computation shows that  $P_{N,\lambda}[S \geq s|R = r]$  depends on  $N$  only through the requirement that  $r \leq N$ . The objective is to show that this probability is strictly decreasing in  $\lambda$  for each fixed  $r$  and  $s$ .

**Theorem 6.** *If  $r \leq N$ ,  $P_{N,\lambda}[S \geq s|R = r]$  is strictly decreasing in  $\lambda$  for each fixed  $0 < s < T$ .*

The conditional distribution of  $S$  given  $R = r$  is as the sum of  $r$  independent random variables each with density  $\lambda e^{-\lambda x}/(1 - e^{-\lambda T})$  on the interval  $[0, T]$ . The following lemma can be easily proved.

**Lemma.** *If  $X_\lambda$  and  $Y_\lambda$  are independent processes and if both  $X_\lambda$  and  $Y_\lambda$  are stochastically decreasing in  $\lambda$  then  $X_\lambda + Y_\lambda$  is stochastically decreasing in  $\lambda$ .*

**Proof of Theorem 6.** By virtue of the lemma, it suffices to show the conditional distribution of  $S$  given  $R = 1$  is stochastically decreasing in  $\lambda$ . Direct computation gives  $P_{N,\lambda}[S \leq s|R = 1] = (1 - e^{-\lambda s})/(1 - e^{-\lambda T})$  for  $0 < s < T$ . Write  $x = e^{-\lambda T}$ . The objective is to show that for each fixed  $a$  with  $0 < a < 1$  the function  $H(x) = (1 - x^a)/(1 - x)$  is decreasing in  $x$  on the interval  $0 < x \leq 1$ . Simple computation gives  $H'(x) = (1 - x^a - ax^{a-1} + ax^a)/(1 - x)^2$ , so the objective is to show that  $g(x) = 1 - x^a - ax^{a-1} + ax^a$  is non-positive on the interval  $0 < x \leq 1$ . Note that  $g(1) = 0$  and  $g(0+) = -\infty$ . Direct computation shows that  $g'(x) = 0$  if and only if  $x = 1$ . Now if  $g$  were ever non-negative on the interval  $0 < x < 1$  then by continuity there would be a  $c$  with  $g(c) = 0$ . Since  $g(1) = 0$  too, the Mean Value Theorem shows that there is a  $d$  with  $c < d < 1$  and  $g'(d) = 0$ . But this is impossible. Thus  $g$  is always negative and  $H$  is decreasing, as claimed. ■

## 7. CONFIDENCE INTERVALS FOR $\lambda$

In order to find a  $100(1 - \alpha)\%$  one sided confidence interval for  $\lambda$  of the form  $(0, \Lambda^U(R, S)]$  the previous method will again be used. If the null hypothesis  $H_0 : \lambda = \lambda_0$  is tested against the alternative  $H_1 : \lambda < \lambda_0$  the results of the previous section suggest that the null hypothesis should be rejected if the observed value  $s$  of  $S$  is too large. The  $p$ -value of this test would then be  $P_{N, \lambda_0}[S \geq s | R = r]$ . The desired value of  $\Lambda^U(r, s)$  would be the smallest value of  $\lambda_0$  for which the  $p$ -value did not exceed  $\alpha$ . With this motivation in mind, define  $\Lambda^U(0, s) = \infty$  for all  $s \geq 0$  and for  $r \geq 1$  set

$$\Lambda^U(r, s) = \inf\{\lambda > 0 : P_{N, \lambda}[S \geq s | R = r] \leq \alpha, N \geq r\}.$$

**Theorem 7.** *For any  $N \geq 1$  and any  $\lambda > 0$ ,  $P_{N, \lambda}[\lambda \leq \Lambda^U(R, S)] \geq 1 - \alpha$ .*

**proof.** By the Theorem of Total Probability, the theorem will be proved if the stronger assertion

$$P_{N, \lambda}[\lambda \leq \Lambda^U(R, S) | R = r] \geq 1 - \alpha$$

is established for all  $r \geq 0$ . By definition of  $\Lambda^U$  this inequality holds in the case  $r = 0$ . For  $r \geq 1$  fixed, observe that  $\Lambda^U(r, s)$  is a continuous decreasing function of  $s$  which is 0 for large enough  $s$  and has limit  $\infty$  as  $s \rightarrow 0^+$ . Hence for each  $\lambda > 0$  there is a unique value of  $s$ , say  $s_\lambda$ , with the property  $\Lambda^U(r, s_\lambda) = \lambda$ . Moreover,  $\Lambda^U(r, s) \geq \lambda$  if and only if  $s \leq s_\lambda$ . Now for  $N \geq r$

$$\begin{aligned} P_{N, \lambda}[\lambda \leq \Lambda^U(R, S) | R = r] &= P_\lambda[S \leq s_\lambda | R = r] \\ &= 1 - P_\lambda[S \geq s_\lambda | R = r] \\ &= 1 - P_{\Lambda^U(r, s_\lambda)}[S \geq s_\lambda | R = r] \\ &= 1 - \alpha \end{aligned}$$

where the fact that, by continuity,  $P_{\Lambda^U(r, s)}[S \geq s | R = r] = \alpha$  has been used. ■

In a similar way, if the null hypothesis  $H_0 : \lambda = \lambda_0$  is tested against the alternative  $H_1 : \lambda > \lambda_0$  the null hypothesis should be rejected if the observed value  $s$  of  $S$  is too small. The  $p$ -value of this test would then be  $P_{N,\lambda_0}[S \leq s | R = r]$ . The desired value of  $\Lambda^L(r, s)$  would be the largest value of  $\lambda_0$  for which the  $p$ -value did not exceed  $\alpha$ . Define  $\Lambda^L(0, s) = 0$  for all  $s \geq 0$ , and for  $r \geq 1$  set

$$\Lambda^L(r, s) = \sup\{\lambda > 0 : P_{N,\lambda}[S \leq s | R = r] \leq \alpha, N \geq r\}$$

with the convention that the supremum of the empty set is infinity. The following theorem has a proof that is similar to that of Theorem 7.

**Theorem 8.** *For any  $N \geq 1$  and any  $\lambda > 0$ ,  $P_{N,\lambda}[\Lambda^L(R, S) \leq \lambda] \geq 1 - \alpha$ .*

The following result is a simple consequence of the previous two theorems.

**Theorem 9.** *For any  $N \geq 1$  and any  $\lambda > 0$ ,  $P_{N,\lambda}[\Lambda^L(R, S) \leq \lambda \leq \Lambda^U(R, S)] \geq 1 - 2\alpha$ .*

## 8. CONFIDENCE INTERVALS FOR $N$

The preceding results can finally be combined in order to get confidence intervals for  $N$ .

**Theorem 10.** *If  $\Lambda^U(R, S)$  and  $\mathcal{N}^L(R, p)$  denote the endpoints of the conservative  $100(1 - \alpha)\%$  one sided confidence intervals for  $\lambda$  and  $N$  found above then for any  $N \geq 1$  and any  $\lambda > 0$ ,  $P_{N,\lambda}[\mathcal{N}^L(R, 1 - e^{-\Lambda^U(R, S)T}) \leq N] \geq 1 - 2\alpha$ .*

**proof.** The proof is obtained by looking at how the proposed confidence

interval can fail.

$$\begin{aligned}
& P_{N,\lambda}[N < \mathcal{N}^L(R, 1 - e^{-\Lambda^U(R,S)T})] \\
&= P_{N,\lambda}[N < \mathcal{N}^L(R, 1 - e^{-\Lambda^U(R,S)T}), \lambda > \Lambda^U(R, S)] \\
&\quad + P_{N,\lambda}[N < \mathcal{N}^L(R, 1 - e^{-\Lambda^U(R,S)T}), \lambda \leq \Lambda^U(R, S)] \\
&\leq P_{N,\lambda}[\lambda > \Lambda^U(R, S)] + P_{N,\lambda}[N < \mathcal{N}^L(R, 1 - e^{-\lambda T})] \\
&\leq 2\alpha
\end{aligned}$$

where the fact that  $\mathcal{N}^L(r, p)$  is decreasing in  $p$  for each fixed  $r$  has been used. ■

The following two theorems can be obtained by similar methods.

**Theorem 11.** *If  $\Lambda^L(R, S)$  and  $\mathcal{N}^U(R, p)$  denote the endpoints of the conservative  $100(1-\alpha)\%$  one sided confidence intervals for  $\lambda$  and  $N$  found above, then for any  $N \geq 1$  and any  $\lambda > 0$ ,  $P_{N,\lambda}[R \leq N \leq \mathcal{N}^U(R, 1 - e^{-\Lambda^L(R,S)T})] \geq 1 - 2\alpha$ .*

**Theorem 12.** *If  $\Lambda^U(R, S)$ ,  $\Lambda^L(R, S)$ ,  $\mathcal{N}^U(R, p)$ , and  $\mathcal{N}^L(R, p)$  denote the endpoints of the conservative  $100(1-\alpha)\%$  one sided confidence intervals for  $\lambda$  and  $N$  respectively, then for any  $N \geq 1$  and any  $\lambda > 0$ ,  $P_{N,\lambda}[\mathcal{N}^L(R, 1 - e^{-\Lambda^U(R,S)T}) \leq N \leq \mathcal{N}^U(R, 1 - e^{-\Lambda^L(R,S)T})] \geq 1 - 4\alpha$ .*

## 9. APPLICATIONS

The method given above was applied to the DACS software reliability datasets (Musa (1980)). The conservative 95% confidence intervals obtained by the methods here are compared with intervals obtained by Osborne and Severini (1998) using asymptotic (in  $N$ ) methods based on the integrated

likelihood.

Data Set	$r$	$s$	MLE	O& S Left	O& S Right	FTV Left	FTV Right
1	136	36.9	141	137	150	136	161
14c	36	14.4	47	38	122	36	$\infty$
17	38	9.78	38	38	44	38	53
2	54	14.2	55	54	62	54	71
27	41	7.70	41	41	43	41	47
3	38	7.59	38	38	40	38	45
4	53	7.16	53	53	53	53	56
40	101	21.8	102	101	105	101	111
5	831	372	1758	1409	2451	1327	2929
6	73	25.8	85	76	114	73	148
ssl a	112	51.4	265	156	961	140	$\infty$
ssl b	375	183	2535	860	11189	814	$\infty$
ssl c	277	114	414	349	552	329	676
ss2	192	97.5	$\infty$	441	13279	450	$\infty$
ss3	278	111	387	337	489	319	575
ss4	196	88.8	435	290	1065	264	26515

The comparison is quite favorable in those cases in which  $N$  is evidently small. The extra width of the intervals obtained by our methods is compensated for in these cases by the provable level of confidence of our intervals.

## 10. IMPLEMENTATION DETAILS

The computation of the confidence limits for  $\lambda$  is numerically very challenging. The tail probability  $P_\lambda[S \geq s | R = r]$  depends on  $\lambda$  in a complicated way and the exact form of the tail probability is difficult to determine. Solving the necessary equations to find  $\Lambda^L$  and  $\Lambda^U$  exhibited numerical instability in the solutions. The only practical method of computing this tail probability is to use an Edgeworth approximation. The first 2 terms of the Edgeworth series were found to provide an adequate approximation, in limited testing.

In finding confidence intervals for  $N$  the following simpler procedure was used. Suppose  $M$  is proposed as the lower endpoint of a confidence interval



for  $N$ . Find the smallest success probability  $p$  that would allow  $M$  to be a lower endpoint, that is, find the smallest  $p$  so that  $M = \mathcal{N}^L(R, p)$ . This  $p$  satisfies  $P[b(M, p) \geq r] = \alpha$ . There is a unique  $p$  for each  $M \geq r$ , and as  $M$  increases the corresponding  $p$  decreases. Moreover, this  $p$  can be easily and accurately found numerically using the relationship between the binomial distribution and the incomplete beta function. Having found  $p$  compute the corresponding value  $\lambda = -\log(1 - p)/T$  and then find  $P_\lambda[S \geq s | R = r]$  for this particular value of  $\lambda$ . If this tail probability is less than  $\alpha$  then  $\lambda \geq \Lambda^U(r, s)$  so  $M$  is an acceptable candidate for the lower endpoint for  $N$ . As  $M$  is increased, the computed value of  $\lambda$  decreases and hence the computed tail probability increases. As  $M$  is increased the computed tail probability will either eventually exceed  $\alpha$  or always remain below  $\alpha$ . In the first case, the largest value of  $M$  for which the tail probability is still below  $\alpha$  is the lower endpoint of the confidence interval for  $N$ . The second case occurs if  $P_{\lambda=0}[S \geq s | R = r] < \alpha$ , in which case  $\infty$  is the lower endpoint of the confidence interval for  $N$ .

A similar procedure is used to find the upper endpoint of a confidence interval for  $N$ . Suppose  $M$  is proposed as the upper endpoint of a confidence interval for  $N$ . Find the largest success probability  $p$  that would allow  $M$  to be an upper endpoint, that is, find the largest  $p$  so that  $M = \mathcal{N}^L(R, p)$ . This  $p$  satisfies  $P[b(M, p) \leq r] = \alpha$ . There is a unique  $p$  for each  $M > r$ , and as  $M$  increases the corresponding  $p$  decreases. Having found  $p$  compute the corresponding value  $\lambda = -\log(1 - p)/T$  and then find  $P_\lambda[S \leq s | R = r]$  for this particular value of  $\lambda$ . If this tail probability exceeds  $\alpha$  then  $\lambda \geq \Lambda^L(r, s)$  so  $M$  is an acceptable candidate for the upper endpoint for  $N$ . As  $M$  is increased, the computed value of  $\lambda$  decreases and hence the computed tail probability decreases. As  $M$  is increased the computed tail probability will either eventually fall below  $\alpha$  or always exceed  $\alpha$ . In the first case, the first

value of  $M$  for which the tail probability falls below  $\alpha$  is the upper endpoint of the confidence interval for  $N$ . The second case occurs if  $P_{\lambda=0}[S \leq s | R = r] > \alpha$ , in which case  $\infty$  is the upper endpoint of the confidence interval for  $N$ .

## BIBLIOGRAPHY

- Barnard, G. A. (1984). “Comparing the Means of Two Independent Samples.” *Applied Statistics*, 33 no. 3, 266–271.
- Bickel, P. J. and K. A. Doksum (1977). *Mathematical Statistics*. San Francisco: Holden-Day.
- Finkelstein, M., H. G. Tucker, and J. Veeh (1998). “Confidence intervals for the number of unseen types.” *Statistics and Probability Letters*, 37, 423–430.
- Jelinski, Z. and P. Moranda (1972). “Software Reliability Research.” In *Statistical Computer Performance Evaluation*, Walter Freiberger, editor, pages 465–484. New York: Academic Press.
- Joe, H. and N. Reid (1985). “Estimating the Number of Faults in a System.” *Journal of the American Statistical Association*, 80 no. 389, 222–226.
- Kendall, M. G. and A. Stuart (1991). *Advanced Theory of Statistics*, volume 2 fifth edition. London: Edward Arnold.
- Musa, J. D. (1980). *Software Reliability Data*. Data and Analysis Center for Software. Also at [www.dacs.dtic.mil/databases/sled/swrel.shtml](http://www.dacs.dtic.mil/databases/sled/swrel.shtml).
- Osborne, J. A. and T. A. Severini (preprint 1998). “Inference for Exponential Order Statistic Models based on an Integrated Likelihood Function.”