NRC Research and/or Technical Assistance Rept

PDR

EGG-M-11682

10/5/82

# APPROXIMATE TOLERANCE INTERVALS, BASED ON MAXIMUM LIKELIHOOD ESTIMATES

Corwin L. Atwood
EG&G Idaho, Inc.

ABSTRACT

Let $X_1, \ldots, X_n$ be independent random variables with distributions depending on a possibly multidimensional $\theta$. Let Y be an unobserved continuously distributed random variable whose distribution depends on $\theta$. A tolerance interval for Y is desired, satisfying $P[Y \epsilon I(X_1, \ldots, X_n)] = \beta$. A naive interval would estimate $\theta$ from the X's, and construct the interval assuming that the estimate is exactly correct. This paper assumes standard regularity conditions, and uses Taylor approximations to construct correction terms of order $1/n$. The resulting interval is longer than the naive interval, because it takes into account the uncertainty in the estimate of $\theta$. Two examples, one simple and one complex, illustrate the method.

## APPROXIMATE TOLERANCE INTERVALS, BASED ON MAXIMUM LIKELIHOOD ESTIMATES

### 1. INTRODUCTION AND TERMINOLOGY

Based on observations $X_1,\ldots,X_n$, inference is to be made about an unobserved continuously distributed random variable Y. The inference consists of an interval $I = I(X_1,\ldots,X_n)$, such that I contains Y according to some probability statement. There are two examples that will be used repeatedly for illustrations.

Normal Example. Here $X_1,\ldots,X_n$ and Y are independent normal$(\mu,\sigma^2)$ random variables, with $\mu$ and $\sigma^2$ unknown. Based on the observable values $X_1,\ldots,X_n$, we want to predict a future value Y, by an interval $I = I(X_1,\ldots, X_n)$ that contains Y according to some probability statement.

Gamma-Poisson Example. Here $\lambda$ has a gamma$(\alpha,\beta)$ distribution, with $\alpha$ and $\beta$ unknown. For $i = 1,\ldots,n$, values $\lambda_i$ are independently generated from this distribution. Corresponding to each $\lambda_i$, a Poisson$(\lambda_i t_i)$ random variable $X_i$ is observed, with $t_i$ known and $\lambda_i$ unknown. Based on the observed values $X_1,\ldots,X_n$ we wish to cover most of the distribution of $\lambda$ by an interval I. Equivalently, if Y is some future randomly generated $\lambda$, we wish to construct $I = I(X_1,\ldots,X_n)$ such that I contains Y according to some probability statement.

A commonly used probability statement to precisely relate I and Y is the following. The interval I is a tolerance interval with content $\beta$ and confidence coefficient $\alpha$ if

$$P\,\{P\,[Y\epsilon I(X_1,\ldots,X_n)\ \big|\ X_1,\ldots,X_n] \geq \beta\} \geq a \quad . \tag{1}$$

In words, this says that, with probability at least $\alpha$, the interval covers at least $\beta$ of the distribution of Y. This definition is mentioned here only because it is so commonly used. It is usually difficult, even in problems as simple as our normal example, to find an interval I satisfying (1). Solutions usually involve specially computed tables. Therefore in this paper we will restrict attention to a simpler definition. The interval I is a <u>tolerance interval with expected content</u> $\beta$ if

$$E \{P [Y \in I(X_1,\ldots,X_n) \mid X_1,\ldots,X_n]\} = \beta \quad . \tag{2}$$

In words, this says that the average coverage is $\beta$ of the distribution of Y. Since the expectation of a conditional probability is an unconditional probability, (2) can be rewritten as

$$P [Y \in I(X_1,\ldots,X_n)] = \beta \tag{3}$$

This simply says that the interval contains Y with probability $\beta$, i.e., I is a <u>prediction interval</u> for Y. In this last probability statement, both Y and the $X_i$'s are considered random.

Equations (1) and (2) are related, since an interval satisfying (2) satisfies (1) with $\alpha \doteq 1/2$. This approximation rests on the approximate equality of the mean and the median of the probability in (2). For a fuller treatment of these concepts, see Guttman (1970).

Again, this paper only considers intervals satisfying (2) or, equivalently, (3).

The method for finding an approximate tolerance interval is as follows. Let $F_\theta$ be the cumulative distribution function (c.d.f.) of Y, determined by an unknown parameter $\theta$. Here Y is one-dimensional, but $\theta$ may be multidimensional. Based on $X_1,\ldots,X_n$, find the maximum likelihood estimate, $\hat{\theta}$. Then $F_{\hat{\theta}}$ estimates the distribution of Y. Now let $\gamma$ be some probability of interest, such as 0.05 or 0.95. Let $a_\gamma(\hat{\theta})$ be defined by

$$F_{\hat{\theta}}[a_\gamma(\hat{\theta})] = \gamma \quad .$$

A naive interval would use $a_\gamma(\hat{\theta})$ as one end point of the interval I. Such an interval ignores the uncertainty in $\hat{\theta}$. It incorporates the random variability of Y, but not the random variability of the $X_i$'s. Therefore it is generally too short, and does not have an expected content as large as claimed.

To get a better interval, use the asymptotic theory for the maximum likelihood estimator, to get a correction term of order 1/n in the equation

$$E\,\{P[Y \leq a_\gamma(\hat{\theta}) \mid \hat{\theta}]\} \equiv E\,\{F_\theta[a_\gamma(\hat{\theta})]\} \doteq \gamma + \text{correction} \quad .$$

This gives an approximation to the true expected content of the naive interval. Iteration on $\gamma$ yields an interval so that ($\gamma$ + correction) is the desired size, e.g. 0.95. For this $\gamma$, $a_\gamma(\hat{\theta})$ is one end point of a tolerance interval with approximately the desired expected content.

## 2. RESULTS

### 2.1 Notation

All the notation needed for the statements of the results is given here, for convenient reference. The upper case letters U, X, and Y represent random variables.

Let $g_m$ be the density or discrete probability function of $X_m$.

Define

$$U_{i;m} = (\partial/\partial\theta_i) \log g_m(X_m)$$

$$U_{ij;m} = (\partial^2/\partial\theta_i\, \partial\theta_j) \log g_m(X_m)$$

$J$ = the matrix having $-\Sigma_m EU_{ij;m}$ as element ij

$J^{ij}$ = the element ij of $J^{-1}$.

The matrix J is the Fisher information. The letter J is used because I has already been used for "interval." Let $F_\theta(y)$ equal $P(Y \leq y|\theta)$, the cumulative distribution function of Y. Define

$a_\gamma(\theta)$ = $F_\theta^{-1}(\gamma)$, i.e. $a_\gamma(\theta)$ is such that $F_\theta[a_\gamma(\theta)] = \gamma$

$F_{10}$ = $(\partial/\partial y)\, F_\theta(y)$

$F_{20}$ = $(\partial^2/\partial y^2)\, F_\theta(y)$

$F_{01}$ = the vector having $(\partial/\partial\theta_i)\, F_\theta(y)$ as element i

$F_{11}$ = the vector having $(\partial^2/\partial\theta_i\, \partial y)\, F_\theta(y)$ as element i

$F_{02}$ = the matrix having $(\partial^2/\mu\theta_i\ \partial\theta_j)\ F_\theta(y)$ as element ij

Unless it is stated otherwise, the derivatives of F are all evaluated at $y = a_\gamma(\theta)$. A superscript T will be used to denote a transpose.

Let $\Theta$ be a neighborhood of the true $\theta$.

## 2.2 Assumptions on $X_1,\ldots,X_n$

The random variables $X_1,\ldots,X_r$ are independent. The usual regularity conditions hold so that the maximum likelihood estimator is asymptotically normal with mean $\theta$ and nonsingular covariance matrix $J^{-1}$. [See, for example, Cox and Hinkley (1974) or Cramer (1946).] Also, for all $\theta\epsilon\Theta$ and all i, j, k and m,

$$E\ [(\partial^3/\partial\theta_i\ \partial\theta_j\ \partial\theta_k)\ g_m(X_m)\ /\ g_m(X_m)] = 0. \tag{4}$$

A sufficient condition for (4) is: the third mixed partial derivative is continuous almost surely, and there exists some integrable function $h_{ijkm}$, independent of $\theta$, with

$$|(\partial^3/\partial\theta_i\ \partial\theta_j\ \partial\theta_k)\ g_m(x)|\ \leq h_{ijkm}(x)\ \text{for all } x\ .$$

## 2.3 Assumptions on Y

The range of Y, R, is an interval, possibly infinite, which does not depend on $\theta$. The c.d.f. $F_\theta(y)$ has continuous derivatives with respect to y and all the components of $\theta$, up through derivatives of order 3, for all $y\epsilon R$ and $\theta\epsilon\Theta$. For y in the interior of R and $\theta$ in $\Theta$, $(\partial/\partial y)\ F_\theta(y)$ is strictly positive.

## 2.4 Statement of Results

Theorem 1. Let $\hat{\theta}_t$ be the t th component of $\hat{\theta}$. Under the assumptions on $X_1,\ldots,X_n$, the bias of $\hat{\theta}_t$ is given by

$$E\,(\hat{\theta}_t - \theta_t) = -\frac{1}{2}\,\Sigma_m\, E\,[\Sigma_k\, J^{jk}\,(U_{jk;m} + U_{j;m}\,U_{k;m})\;\Sigma_i J^{it} U_{i;m}] + o(1/n) \quad .$$

Anderson and Richardson (1979) obtain an expression for the bias which uses third derivatives ($U_{ijk;m}$). The expression given here, using only second-order derivatives, is made possible by the assumption on $X_1,\ldots,X_n$ given by (4).

Theorem 2. Under the assumptions on $X_1,\ldots, X_n$ and Y, we have

$$E\,\{F_\theta[a_\gamma(\hat{\theta})]\} = \gamma - E(\hat{\theta}-\theta)^T F_{01} - \frac{1}{2}\mathrm{tr}\, F_{02}\, J^{-1} + F_{01}{}^T J^{-1} F_{11}/F_{10} + o(1/n) \quad .$$

The proofs are in the appendix.

These two theorems suggest a way to estimate $E\{F_\theta[a_\gamma(\hat{\theta})]\}$. If the expected values in the expressions for J and $E(\hat{\theta}-\theta)$ can be written as explicit functions of $\theta$, estimate them by substituting $\hat{\theta}$ for $\theta$. Otherwise, replace each expected sum by the corresponding sum of observed values, to obtain estimates of J and $E(\hat{\theta}-\theta)$. Both estimators of $n^{-1}J$ and $E(\hat{\theta}-\theta)$ are consistent, but if $\hat{\theta}$ is a sufficient statistic, then the first estimators are preferable because they depend only on the sufficient statistic.

Similarly, estimate $F_{10}$, $F_{01}$, $F_{11}$, and $F_{02}$, by evaluating them at $\hat{\theta} = \theta$ and $y = a_\gamma(\hat{\theta})$. Use these to estimate the quantity

$$\gamma - E(\hat{\theta}-\theta)^T F_{01} - \frac{1}{2}\,\mathrm{tr}\, F_{02}\, J^{-1} + F_{01}{}^T J^{-1}\, F_{11}/F_{10}$$

This estimate equals $E\{F_\theta[a_\gamma(\hat{\theta})]\}$ plus $o_p(1/n)$. Iterate on $\gamma$ until the estimate of $E\{F_\theta[a_\gamma(\hat{\theta})]\}$ equals a desired probability, such as 0.95. Then use $a_\gamma(\hat{\theta})$ as one end point of the interval I.

## 3. EXAMPLES

### 3.1 The Normal Example

In this example, $X_1, \ldots, X_n$ and $Y$ are independent normal$(\mu, \sigma^2)$. Let $\theta = (\mu, \sigma)^T$. Then direct calculation shows that

$$J^{-1} = (\sigma^2/n) \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$E(\hat{\theta}_1 - \theta_1) = 0$$

$$E(\hat{\theta}_2 - \theta_2) = -3\sigma/4n.$$

Let $z = [a_\gamma(\theta) - \mu]/\sigma$, and let $\phi$ denote the standard normal density,

$$\phi(z) = (2\pi)^{-1/2} e^{-z^2/2} .$$

Then more calculation shows that

$$E\,\{F_\theta[a_\gamma(\hat{\theta})]\} = \gamma - z\phi(z)(5 + z^2)/4n + o(1/n) \quad .$$

The right hand side is of the form $\gamma$ + correction + $o(1/n)$. Now $z = \Phi^{-1}(\gamma)$, where $\Phi$ is the standard normal c.d.f. Therefore the correction term is determined by $\gamma$, up to $o(1/n)$, and no estimation is needed. The expectation $E\{F_\theta[a_\gamma(\hat{\theta})]\}$ is

$$\gamma - z\,\phi(z)(5 + z^2)/4n$$

up to $o(1/n)$.

The accuracy of this approximation can be investigated, because in this example exact calculations are possible. These calculations are based on the easily verified fact that

$$[(n - 1)/(n + 1)]^{1/2} (Y - \overline{X})/\sigma$$

has a Student's t distribution with n-1 degrees of freedom. Here $\sigma$ is the maximum likelihood estimator, based on the biased estimator

$$\hat{\sigma}^2 = (X_i - \overline{X})^2/n.$$

For example, if the desired expected content of the tolerance interval is to be 95%, the naive interval would be

$$\overline{X} \pm 1.96\,\hat{\sigma},$$

since $\Phi(1.96) = 0.975$. The approximate interval would be

$$\overline{X} \pm z\,\hat{\sigma},$$

where z is such that $\Phi(z) = \gamma$ and

$$\gamma - z\phi(z)(5 + z^2)/4n = 0.975.$$

The exact interval would be

$$\overline{X} \pm [(n + 1)/(n - 1)]^{1/2}\, t\,\hat{\sigma},$$

where t is such that the cumulative Student's t distribution (with n-1 degrees of freedom) equals 0.975 there.

All three intervals are of the form $\overline{X} \pm c\hat{\sigma}$, so they may be compared easily. Table 1 shows the three kinds of intervals for a number of examples. In these examples, it can be seen that the approximate intervals are closer to the exact intervals than to the naive intervals, even for quite small n. It is also evident that in these examples the approximation never goes far enough, but always undershoots.

3.2 The Gamma-Poisson Example, with Monte Carlo Data

Here $\lambda$ has a gamma($\alpha,\beta$) density

$$f(\lambda) = \lambda^{\alpha-1} e^{-\lambda/\beta} / [\beta^{\alpha}\Gamma(\alpha)].$$

Conditional on $\lambda_i$, $X_i$ has a Poisson($\lambda_i t_i$) distribution, with $t_i$ known. Therefore, the unconditional distribution of $X_i$ is negative binomial ($\alpha,\beta t_i$), with probability distribution function

$$P[X_i = k] = [\alpha \cdots (\alpha + k - 1)/k!] \; (\beta t_i)^k \; (1 + \beta t_i)^{-\alpha-k} .$$

The expected value of $X_i$ is $\alpha\beta t_i$. (See Johnson and Kotz (1969) for a summary of the basic facts.)

This natural parametrization turns out not to work well in practice. The intuitive reason is that $\alpha\beta$--essentially the mean of X--can be estimated rather well, but then $\hat{\alpha}$ and $\hat{\beta}$ have a strong negative correlation. In data considered recently at EG&G Idaho, correlations of around -0.9 are typical. Therefore the information matrix J is poorly conditioned, so any statistical error in estimating J is greatly magnified in $J^{-1}$. In some Monte Carlo runs, described more fully below, the estimates were so unstable that they were useless. The average coverage probability of the resulting approximate tolerance intervals was no better than that of the naive intervals.

TABLE 1. COMPARISON OF TOLERANCE INTERVALS FOR THE NORMAL EXAMPLE[a]

| n | Naive Interval | Approximate Interval | Exact Interval |
|---|---|---|---|
| Desired Expected Content = 0.9000 | | | |
| 5 | 1.645<br>0.7496 | 2.227<br>0.8568 | 2.611<br>0.9000 |
| 10 | 1.645<br>0.8291 | 1.964<br>0.8906 | 2.027<br>0.9001 |
| 30 | 1.645<br>0.8776 | 1.752<br>0.8991 | 1.757<br>0.9000 |
| 100 | 1.645<br>0.8934 | 1.677<br>0.9000 | 1.677<br>0.9000 |
| Desired Expected Content = 0.9500 | | | |
| 5 | 1.960<br>0.8152 | 2.609<br>0.8998 | 3.400<br>0.9500 |
| 10 | 1.960<br>0.8900 | 2.350<br>0.9375 | 2.501<br>0.9500 |
| 30 | 1.960<br>0.9320 | 2.102<br>0.9487 | 2.115<br>0.9500 |
| 100 | 1.960<br>0.9448 | 2.003<br>0.9499 | 2.004<br>0.9500 |

a. In each column labeled "Interval," the pairs of numbers have the following meaning. The upper number is c, where the interval is of the form $\bar{X} \pm c\hat{\sigma}$. The lower number is the true expected content of the interval.

To remedy this, parametrize the distributions in terms of $\alpha$ and $\mu = \alpha\beta$. Then the information matrix J is diagonal, with

$$J_{11} = \sum_m E \sum_{j=0}^{X_m - 1} (\alpha + j)^{-2} - \sum_m \mu t_m / [\alpha(\alpha + \mu t_m)]$$

$$J_{12} = J_{21} = 0$$

$$J_{22} = \Sigma_m \, \alpha t_m / [\mu(\alpha + \mu t_m)].$$

The two sums that do not involve $X_m$ can be estimated by substituting $\hat{\alpha}$ and $\hat{\mu}$ for $\alpha$ and $\mu$. The term

$$\sum_m E \sum_{j=0}^{X_m - 1} (\alpha + j)^{-2}$$

can be estimated either by

$$\sum_m \sum_{j=0}^{x_m - 1} (\hat{\alpha} + j)^{-2}$$

or else by numerically summing the series for the expections, using the estimated values of the parameters. In the Monte Carlo runs described below, these two estimation methods gave almost the same estimates. The first method is simpler.

To complete the example, the derivatives of F must be found. Since $F_\theta(y)$ equals

$$\int_0^{\alpha y/\mu} x^{\alpha - 1} e^{-x} \, dx / \Gamma(\alpha),$$

it is straightforward to find both the derivatives with respect to $\mu$, and the mixed partial derivatives. The derivatives with respect to $\alpha$ are best found numerically. For example, approximate the integral in a plausible range of $\alpha$ by a smooth function such as a cubic spline, and find the derivatives of the smooth function. Differentiation under the integral sign with respect to $\alpha$ is not valid for $\alpha < 1$, and only valid once for $\alpha < 2$. (Nonetheless, all the assumptions on $X_1,\ldots,X_n$ and Y of Sections 2.2 and 2.3 are satisfied.)

Three Monte Carlo runs were performed, with

$\alpha = 1, \quad \mu = 1, \quad n = 15$

$\alpha = 1, \quad \mu = 1, \quad n = 60$

$\alpha = 3, \quad \mu = 9, \quad n = 60$

In each run, $t_m$ was 1 for $m = 1,\ldots,n$. Except for t, which is essentially just a scale factor, these values were chosen to be not too different from values given by real data. In each run, 1000 sets of $X_1,\ldots,X_n$ were generated, using the IMSL (1980) program MDNB. For some data sets, the search for the maximum likelihood estimates did not converge. This happens if the distribution of $\lambda$ appears to degenerate to a constant. Then that constant is $\mu$, and $\alpha$ is infinite. However, when the maximum likelihood estimates were found, the approximate tolerance intervals of this paper were also found, with desired expected content of 0.90.

Since the true values of the parameters are known, exact 90% intervals for $\lambda$ can also be found, for comparison. Summary statistics for the naive intervals and the approximate intervals are shown in Table 2, where they are compared with the exact intervals.

In this table, the average upper and lower end points are printed. Of course the intervals vary greatly from data set to data set, so these average end points are only suggestive of the behavior of the methods for obtaining tolerance intervals. The mean tail probabilities and mean contents are shown. Ideally, these should equal 0.05, 0.05, and 0.90,

TABLE 2. MONTE CARLO COMPARISON OF TOLERANCE INTERVALS FOR THE GAMMA-POISSON EXAMPLE

| Quantity | Interval: Naive | Approximate | Exact |
|---|---|---|---|
| | $\alpha = 1, \mu = 1, n = 15$ (847 trials)[a] | | |
| Mean Interval | (0.138, 3.05) | (0.083, 4.98) | (0.051, 3.00) |
| Mean Tail Probabilities | 0.116, 0.086 | 0.040, 0.016 | 0.050, 0.050 |
| Mean Content (standard (error) | 0.80 (0.0062) | 0.94 (0.0054) | 0.90 |
| | $\alpha = 1, \mu = 1, n = 60$ (998 trials)[a] | | |
| Mean Interval | (0.089, 2.29) | (0.043, 3.25) | (0.051, 3.00) |
| Mean Tail Probabilities | 0.081, 0.068 | 0.041, 0.047 | 0.050, 0.050 |
| Mean Content (standard error) | 0.85 (0.0035) | 0.91 (0.0017) | 0.90 |
| | $\alpha = 3, \mu = 9, n = 60$ (1000 trials)[a] | | |
| Mean Interval | (2.59, 18.72) | (2.40, 19.35) | (2.45, 18.89) |
| Mean Tail Probabilities | 0.061, 0.058 | 0.051, 0.051 | 0.050, 0.050 |
| Mean Content (standard error) | 0.88 (0.0016) | 0.90 (0.0014) | 0.90 |

a. The number of trials listed is the number of data sets, out of a possible 1000, for which maximum likelihood estimates could be found.

respectively. The standard error is also shown for each mean content. They are small enough to show that the Monte Carlo mean contents probably differ from the corresponding true expected contents by at most 0.01.

In these three examples, Table 2 shows that the naive interval is on the average too short, while the approximate interval is on the average too long. Of the two methods, the approximate interval comes closer to the target than does the naive interval, in the sense that its expected content is closer to 0.90.

### 3.3 A Gamma-Poisson Example with Real Data

Failure data on 128 diesel generators in 58 commercial nuclear power plants for 1976-1978 are presented by Atwood and Steverson (1982). The diesel generators serve as emergency back-up power sources to the plant equipment. They are normally not running, but are started when other power sources are lost, and for periodic tests. Since the number of demands on each diesel generator is unknown, the observed number of calendar hours for the plant is used as $t_i$ for the ith diesel generator. Most of the plants operated for the full three years ($t_i$ = 26304 hours) although one was observed for only 600 hours, and others had intermediate values of $t_i$.

Assume that the number of reported failures of a diesel generator is a Poisson($\lambda_i t_i$) random variable, with $\lambda_i$ an unknown failure rate associated with the diesel generator. Because of changes in personnel, reporting policy, rate of demands, and equipment, $\lambda_i$ probably varies over time, so the Poisson model does not perfectly correspond to reality.

Assume that the $\lambda_i$'s come from a gamma population. The gamma family is chosen for mathematical convenience, and lack of a better model.

Two kinds of failures are considered: "individual failures," in which there is no mechanism synchronizing the failure of one diesel generator with that of another, and "common cause events," when there is a shock, external to the diesel generators, that can at least potentially cause several simultaneous failures. If we ignore certain technicalities and details, the data contain reports of 350 individual failures and 22 common cause events.

Based on the individual failures for 128 diesel generators, $\alpha$ and $\mu$ are estimated to be 1.387 and 8.904E-3, so the naive 90% interval for the individual failure rate is (1.265E-5, 3.302E-4). One of the diesel generators had enough recurrent failures so that it is a borderline outlier. However the upper end point of the interval only changes by 10% if that diesel generator is not counted, so it is left in.

The method of this paper, as detailed in Section 3.2, gives an approximate 90% interval of (1.148E-5, 3.440E-4). Compared to the width of the interval, the effect of the correction is very small.

Based on the 22 common cause events at 58 plants, $\alpha$ and $\mu$ are estimated to be 0.5830 and 2.944E-5. Because the estimate of $\alpha$ is less than 1.0, the estimated gamma density approaches infinity at $\lambda = 0$. Therefore the left end point of the naive interval is very small. The naive 90% interval for the common cause event rate is (1.423E-7, 6.241E-5). The correction of this paper produces an approximate 90% interval of (2.073E-8, 6.576E-5). The change in the lower end point is large compared to the lower end point, but small compared to the length of the interval. The change in the upper end point is small by either standard.

Atwood and Steverson simply present the naive intervals. Their reasons are the following. The data are not of high quality, since there is probably not a completely consistent reporting policy from plant to plant, and since at least a few events probably have not been reported. There is also some evidence of lack of fit to the model: occasional strings of recurrent failures are reported; one diesel generator may be an outlier; and fitting a gamma distribution to the rate of common cause events makes the most likely rate exactly zero. Because of these considerations, it does not seem worthwhile to go to the trouble of using the method of this paper, in order to change the length of the interval by a few percent. The value of the method of this paper, with the diesel generator data, is to confirm that the naive interval is good enough, that the uncertainty due to lack of failure data is small compared to the inherent variability in $\lambda$.

## 4. PROOFS

Proof of Theorem 1 Following Cox and Hinkley (1974), pp. 309-310, Anderson and Richardson (1979) show that the asymptotic bias of the maximum likelihood estimator is given by

$$E(\hat{\theta}_t - \theta_t) = \frac{1}{2} \Sigma_m \Sigma_{i,j,k} J^{it} J^{jk} \{2 E\, U_{j;m} U_{ik;m} + E\, U_{ijk;m}\} + o(1/n). \quad (5)$$

(Anderson and Richardson do not make essential use of their assumption that $X_1, \ldots, X_n$ are identically distributed.) For compactness, denote $g_m(X_m)$ by $G$, denote $(\partial/\partial\theta_i) g_m(X_m)$ by $G_i$, and let $G_{ij}$ and $G_{ijk}$ be similarly defined. Then the expression in curly brackets in (5) can be rewritten as

$$E\, [G_j G_{ik}/G^2 - G_k G_{ij}/G^2 - G_i G_{jk}/G^2 + G_{ijk}/G].$$

The expection of $G_{ijk}/G$ is 0, by the assumption on $X_1, \ldots, X_n$ given by (4). Moreover, by the symmetry of the information matrix J, the first two terms cancel in

$$\Sigma_{jk} J^{jk} E\, [G_j G_{ik}/G^2 - G_k G_{ij}/G^2 - G_i G_{jk}/G^2].$$

Therefore (5) can be written

$$E(\hat{\theta}_t - \theta_t) = -\frac{1}{2} \Sigma_m \Sigma_{i,j,k} J^{it} J^{jk} E[(G_i/G)(G_{jk}/G)] + o(1/n).$$

This can then be rewritten as in Theorem 1. □

Proof of Theorem 2 Let $\delta = a_\gamma(\hat{\theta}) - a_\gamma(\theta)$. Then, by Taylor's theorem,

$$F_\theta[a_\gamma(\hat{\theta})] = F_\theta[a_\gamma(\theta)] + \delta F_{10} + \frac{1}{2} \delta^2 F_{20} + o(\delta^2). \quad (6)$$

Here and below, $F_{10}$, $F_{20}$, $F_{01}$, and $F_{11}$ are always evaluated at $a_\gamma(\theta)$. The first term on the right hand side of (6) is $\gamma$. To evaluate the other terms, an expression for $\delta$ is required. Such an expression is given by Taylor's theorem:

$$\delta = (\hat{\theta} - \theta)^T a'_\gamma(\theta) + \frac{1}{2} (\hat{\theta} - \theta)^T a''_\gamma(\theta)(\hat{\theta} - \theta) + o(|\hat{\theta} - \theta|^2)$$

Here $a'_\gamma(\theta)$ is the vector with element i equal to $(\partial/\partial\theta_i)a_\gamma(\theta)$, and $a''_\gamma(\theta)$ is the matrix with element ij equal to $(\partial^2/\partial\theta_i\, \partial\theta_j)a_\gamma(\theta)$. To evaluate $a'_\gamma(\theta)$ and $a''_\gamma(\theta)$, differentiate the identity

$$F_\theta[a_\gamma(\theta)] \equiv \gamma.$$

This yields the expressions

$$a'_\gamma(\theta) = -(1/F_{10})F_{01}$$

$$F_{10}\, a''_\gamma(\theta) = -F_{02} + (1/F_{10})\, [F_{11}\, F_{01}^T + F_{01}\, F_{11}^T] - (F_{20}/F_{10}^2)\, F_{01}\, F_{01}^T.$$

Substitute the expressions for $\delta$, $a'_\gamma(\theta)$, and $a''_\gamma(\theta)$ into (6). After simplification, this yields

$$F_\theta[a_\gamma(\hat{\theta})] = \gamma - (\hat{\theta}-\theta)^T F_{01}$$

$$- \frac{1}{2} \operatorname{tr} F_{02}\, (\hat{\theta} - \theta)\, (\hat{\theta} - \theta)^T$$

$$+ (1/F_{10})\, F_{01}^T\, (\hat{\theta}-\theta)\, (\hat{\theta}-\theta)^T\, F_{11} + o(|\hat{\theta}-\theta|^2)$$

Take expectations on both sides, to obtain the conclusion of the theorem. □

3

## REFERENCES


1. ATWOOD, C. L., and STEVERSON, J. A. (1982) Common Cause Fault Rates for Diesel Generators: Estimates Based on Licensee Event Reports at U.S. Commercial Nuclear Power Plants, 1976-1978 NUREG/CR-2099.

2. ANDERSON, J. A., and RICHARDSON, S. C. (1979), "Logistic Discrimination and Bias Correction in Maximum Likelihood Estimation," Technometrics, 21, 71-78.

3. COX, D. R., and HINKLEY, D. V. (1974), Theoretical Statistics, London: Chapman and Hall, (U.S. distributor, Halsted Press).

4. CRAMER, H. (1946), Mathematical Methods of Statistics, Princeton: Princeton Univ. Press.

5. GUTTMAN, I. (1970), Statistical Tolerance Regions: Classical and Bayesian, Darien, CT: Hafner.

6. IMSL Library: Reference Manual (1980), International Mathematical and Statistical Libraries, Inc., 7500 Bellaire Blvd., Houston, TX.

7. JOHNSON, N. L., and KOTZ, S. (1969), Discrete Distributions, New York: John Wiley & Sons, pp. 122-126.