1-2002

Discrete Predictive Analysis in Probabilistic Safety Assessment

Paul Kvam
University of Richmond, pkvam@richmond.edu

J. Glenn Miller

Recommended Citation
Kvam, Paul H., and J. Glenn Miller. "Discrete Predictive Analysis in Probabilistic Safety Assessment." *Journal of Quality Technology* 34, no. 1 (January 2002): 106-117.

Discrete Predictive Analysis in Probabilistic Safety Assessment

PAUL H. KVAM and J. GLENN MILLER

Georgia Institute of Technology, Atlanta, Georgia 30332-0205

This paper presents methods for predicting future numbers of component failures for probabilistic safety assessments (PSAs). The research is motivated and illustrated by discrete failure data from the nuclear industry, including failure counts for emergency diesel generators, pumps, and motor operated valves. Failure counts are modeled with Poisson and binomial distributions. Multiple-failure environments create extra problems for predictive inference, and are a primary focus of this paper. Common cause failures (CCFs), in particular, refer to the simultaneous failure of system components due to an external event. CCF prediction is investigated, and approximate inference methods are derived for various CCF models.

Introduction

STATISTICAL modeling is at the heart of probabilistic safety assessment (PSA) for complex systems, such as those in the aerospace, nuclear, and chemical industries. A nuclear power plant, where several mechanical and electronic component groups work interdependently, represents an especially important PSA application and helps to motivate much of our paper. In short, a PSA identifies potential hazards to a system along with corresponding accident sequences, and then categorizes relevant consequences to the system and its surrounding environment. In a nuclear power plant, PSA is a tool used to estimate the plant's core damage frequency, which is typically less than 10^{-4}/year. The analyst records all potential component failures that act as precursors to more harmful accident sequences leading to core damage.

Much of the effort needed to complete a system assessment goes into the qualitative analysis, which involves constructing fault trees and event trees to describe accident sequences and consequences. The qualitative component to PSA starts with a thorough investigation of past failure events and potential future risks. Risks include accidents within the plant (component failure, maintenance faults, operator error) and events outside the plant (natural disasters, accidents in transport, unexpected local peaks in power demand).

The quantitative aspect of PSA includes reliability analyses of component groups within the system. System reliability is estimated from the component reliability estimates and the system configuration. The PSA uses failure data and other system information (such as component degradation information or expert opinion) to quantify failure frequencies of initiating events and the correspondent accidents sequences. Models that translate failure data into estimated accident rates tend to be elaborate due to the network of complex accident sequences that comprise the PSA. Analytical solutions to predictions, estimates, and uncertainties are rarely attainable, leaving the practitioner to make inferences based on approximations and Monte Carlo methods. This is especially apparent in the nuclear industry, where failure data are sparse, and simulations provide more convenient solutions compared to traditional statistical methods. The Monte Carlo methods rely on statistical parameter estimates from reliability analyses to generate simulated failure events, and relevant accident sequences are selected based on the accident frequencies and consequences.

In any PSA simulation, methods of parametric inference are applied to infer knowledge about compo-

Dr. Kvam is an Associate Professor in the School of Industrial & Systems Engineering. His e-mail address is pkvam@isye.gatech.edu.

Mr. Miller is a Ph.D. student in the School of Industrial & Systems Engineering.

nent lifetimes, degradation models, or external failure events. With the available failure data, the uncertainties in estimating the unknown parameters from the final model are summarized through statistical confidence intervals. However, the ultimate goal of most safety assessments is to predict *future* failure events from the estimated reliability model. That is, the future number of component failures are more important to the PSA than the parameters that characterize their failure distributions, which sometimes have little meaning to the user. Instead of applying only standard confidence intervals for unknown parameters, prediction regions for future failure frequencies should be a major consideration of PSA.

By construction, Monte Carlo simulation methods already include uncertainty due to prediction, and this allows the user to model prediction uncertainty without having to understand the statistical issues of prediction, which are sometimes more complicated than standard estimation problems. However, simulation methods cannot help us realize general prediction uncertainty. Traditional methods based on standard reliability estimation will underestimate the variability of future system failure frequency if the problem of prediction is not addressed directly.

In this paper, we focus on prediction intervals for discrete failure events in a PSA. Most current predictive inference in the statistical literature is derived for continuous distributions such as the normal distribution for linear regression. However, a great amount of PSA data are discrete failure counts. Lacking the needed individual component testing programs, analysts rely more on success/failure data produced in operating environments. Degradation measurements and times on test are less common luxuries.

Basic methods for constructing prediction intervals in conventional settings are described in Cox and Hinkley (1974), Faulkenberry (1973), Hahn and Meeker (1991), and Knüsel (1994), and these papers include prediction problems for general discrete distributions. In the following two sections, we discuss the construction of exact one-sided prediction bounds using Poisson and binomial data, which commonly appear in safety assessments. Poisson failure counts for pumps from a U.S. nuclear power plant are examined, and prediction intervals are computed based on fixed future test times. We examine binomial trials of emergency diesel generators (EDGs) using data collected from seven domestic power plants. We also discuss a special concern for PSA, the construction of prediction bounds with zero-failure data.

A primary goal in this paper is to develop predictive inference techniques for multiple failure models. For this case, exact prediction bounds are not generally available. We derive approximate prediction bounds for systems that experience common cause failures (CCFs), which are external events that cause the simultaneous failure of two or more components in the system. Two particular models are used to illustrate multiple-failure prediction, but other models can be easily substituted. Bayesian methods have become increasingly popular in PSA, and Bayesian extensions to predictive analysis are outlined later in this paper. We analyze data from a recent report for the Nuclear Regulatory Commission found in IEEL (1997) on the CCF of generators at U.S. power plants.

TABLE 1. 95% Upper Prediction Bound for Future Pump Failures over the Course of One Reactor-Year. Time is Measured in 1000s of Hours

System Number	Pump Failures	Time on Test	$\widehat{\theta}$	95% UPB ($T_2 = 1$)	95% UPB ($T_2 = T_1$)
1	5	94.32	0.0530	1	12
2	1	15.72	0.0636	1	6
3	5	62.88	0.0795	1	12
4	14	125.76	0.1113	1	25
5	3	5.24	0.5725	3	9
6	19	31.44	0.6043	2	31
7	1	1.05	0.9525	6	6
8	1	1.05	0.9524	6	6
9	4	2.10	1.9048	6	11
10	22	10.48	2.0992	5	35

TABLE 2. 95% Upper Prediction Bound for Future EDG Failures over the Course of One Reactor-Year. Time is Measured in 1000s of Hours

Plant	EDG demands	EDG failures	$\widehat{\theta}$	95% UPB $n_2 = 100$	95% UPB $n_1 = n_2$	n_2 for UPB = 0
A	2017	35	0.0174	4	51	2
B	301	16	0.0532	10	27	0
C	793	18	0.0227	5	30	2
D	206	9	0.0437	10	18	1
E	1176	13	0.0111	3	23	4
G	283	8	0.0283	7	16	1
H	192	11	0.0573	12	20	0

Exact Prediction Bounds for Poisson Data

Data for many individual plant components are in the form of failure counts aggregated over fixed time intervals. For reliable components such as cooling fans, motor-operated valves, check valves, and auxiliary feedwater pumps, failure rates are low and in most cases are relatively constant between scheduled maintenance operations (see Sanzo et al. (1994)). With special exceptions, a homogeneous Poisson process is assumed to govern the failure events for these components. In this section, we illustrate predictive analysis in PSA using pump failure data from the U.S. Farley-1 nuclear power plant. The data originated in a plant report at the Electric Power Research Institute by Worledge et al. (1982) and are reproduced in Gaver and O'Muircheartaigh (1987). A summary of the failure data from 10 pump systems in the Farley-1 plant is listed in Table 1. The time on test, T_1, is listed in thousands of hours.

For the Poisson prediction interval, consider the following model. From a homogeneous Poisson process with unknown recurrence rate $\theta_1 > 0$, let X represent the number of occurrences in time interval $(0, T_1)$, and denote by Y the future number of occurrences of an independent process with unknown recurrence rate $\theta_2 > 0$ in the time interval (T_1, T_1+T_2). It is our goal to predict $Y \sim \text{Poisson}(\theta_2 T_2)$ using $X \sim \text{Poisson}(\theta_1 T_1)$, assuming the processes have identical recurrence rates; i.e., $\theta_1 = \theta_2$. Consider the hypothesis test for $H_0 : \theta_1 = \theta_2$ vs. $H_a : \theta_1 < \theta_2$. The rejection region for the α-level test of H_0 has form $\{(x, y) : y > U(x)\}$, where $U(x)$ serves as an upper prediction bound (UPB) for Y, given X. Lower prediction bounds are constructed similarly by switching the inequality in the alternative hypothesis.

Under H_0, the conditional distribution of X given $N = X + Y$ is binomial with parameters $N = n$ and $p = T_1/(T_1 + T_2)$; thus, $X \mid N = n$ forms a pivotal quantity, and we can use this binomial distribution to construct exact prediction intervals for Y. Let $b_\alpha(n, p) = b_\alpha$ be the lower α^{th} quantile of the binomial distribution; i.e., b_α is the largest integer such that $P(X < b_\alpha \mid X + Y = n) \leq \alpha$. The rejection region for the one-sided test of hypothesis consists of all pairs (x, y) for which x is considered too small, relative to y. Accordingly, the one-sided prediction interval for Y, given $X = x$, has the form $[0, y^*]$, where we choose y^* to be the largest integer such that $x \geq b_\alpha(x + y^*, p)$. This ensures that $y^* + 1$ is the smallest value of Y (in a fixed $X+Y$) for which X is considered too small (relative to $X + Y$) to fail to reject H_0. Thus, given that $X = x$, $[0, y^*]$ is the narrowest interval for which $P(Y \leq y^*) \geq 1 - \alpha$.

As a simple example from the nuclear industry, suppose that in one (reactor) year, a group of containment cooling fans at a power plant experiences no failures or significant degradation during continuous service. Common risks for cooling fans include broken blades, fan belt slippage, and loss of motor-bearing lubrication, which causes degradation and low flow. We can form a 95% UPB for the number of future failures experienced by the cooling fans in the next (reactor) year if we assume the failure rate remains the same (that is, the Poisson process for failure events is homogeneous). If we let $X = 0$ be the number of observed failures and Y as the (Poisson distributed) number of future failures, the pivotal statistic is distributed binomial with $n = 0 + Y$, and $p = T_1/(T_1+T_2) = 0.5$. The lower quantiles of the binomial distribution are $b_{0.05}(3, 0.5) = b_{0.05}(4, 0.5) = 0$, and $b_{0.05}(5, 0.5) = 1$. Thus, $y^* = 4$ is the largest value of y for which $x > b_{0.05}(x + y, 0.5)$.

While one-sided prediction bounds are the main focus for safety assessments in which component failures are rare, two sided prediction intervals can be constructed in the same manner by simultaneously considering lower bounds and upper bounds, each based on the quantiles $b_{\frac{\alpha}{2}}$ and $b_{1-\frac{\alpha}{2}}$. For the pump failure data, the 95% UPBs are calculated in Table 1. In column 4 of the table, the upper bound refers to the case in which we predict future failures based on $T_2 = 1.0$ thousand hours on test. The results in column 5 refer to the case in which the future time on test T_2 is fixed to be the same as T_1.

Methods for standard prediction limits are described in various sources for the reliability practitioner, including Nelson (1982) and Cox and Hinkley (1974), where quicker, approximate methods are usually prescribed. A normal approximation, described in Chapter 6 of Nelson (1982), is straightforward and somewhat effective for this Poisson prediction problem. For example, the $(1-\alpha)$-UPB for Y in the case $T_1 = T_2$ is $x + z_{1-\alpha}\sqrt{2x}$, where z_q is the q^{th} quantile of the standard normal distribution. However, the bounds fail to be accurate unless the mean values $\theta_i T_i$ are large enough. Obviously, these approximate bounds will not be effective for problems with few or no observed failures.

Exact Prediction Bounds for Binomial Data

In a Brookhaven National Laboratory report for the NRC by Lofgren and Gregory (1991), emergency diesel generators (EDGs) were tested individually in separate maintenance programs for seven different nuclear power plants. Between 1976 and 1991, EDG demands and failures were recorded at each of these plants, which we treat as binomial outcomes summarized in Table 2. Failure events can include the EDG's failure to start, its failure to run according to standards once started, or the occurrence of problems with an EDG that might lead to an incipient failure in future start-ups, such as poor maintenance or worn components.

Along with Poisson data, binomial failure counts provide the bulk of discrete PSA data. Predictive inferences with binomial observations are derived in the same manner as with Poisson data. Suppose we observe $X \sim \text{binomial}(n_1, \theta)$, with only n_1 being known. Let $Y \sim \text{binomial}(n_2, \theta)$ denote future failure observations, with n_2 being known. Now, $X+Y$ has a binomial distribution, and the conditional distribution of X given $X + Y = n$ is hypergeometric with probability mass function

$$p_N(x \mid n) = \frac{\binom{n_1}{x}\binom{n_2}{n-x}}{\binom{n_1+n_2}{n}},$$
$$x = \max\{0, n - n_2\}, ..., n_1. \qquad (1)$$

To determine a $(1-\alpha)100\%$ UPB, we choose y^* to be large enough to ensure $x \le h_\alpha(x + y^*, n_1, n_2)$, where h_α represents the lower α^{th} quantile of the hypergeometric distribution in Equation (1). Then, $y^* = U_\alpha(x)$ is the largest value of Y such that $P(X \le x \mid x + y = n) > \alpha$.

For the EDG trials, the 95% UPBs are calculated in Table 2. Prediction intervals in column 5 are computed for a future year in which $n_2 = 100$ demands are placed upon each EDG. Prediction intervals in column 6 are for the case in which $n_1 = n_2$. Column 7 lists the largest value of $n_2 \ge 0$ for which the 95% prediction intervals are computed as y^*=0. The information provided in column 7 has value in expensive trials for extremely reliable components where small n_2 and zero-failure data are not uncommon. Except for Plant E (which has the highest observed EDG reliability), no plant obtains 95% certainty of zero-failures if more than two trials are run.

Table 3 exhibits 95% binomial prediction UPBs for the case in which $n_1 = n_2$ and $X = 0, 1, 2$. In column 1, the UPB for Y is listed. Alongside the UPB are the smallest values of $n_1 = n_2$, given X, required to achieve that UPB. For example, if $X = 0$, we need $n_1 = n_2 \ge 16$ in order for $y^* = 4$ to serve as a 95% UPB. Furthermore, there is no $n_1 = n_2$ for which $y^* = 5$ serves as an UPB.

Prediction Intervals Based on Zero Failures

In demonstration testing at power plants, electronics manufacturing plants, and military test cen-

TABLE 3. Smallest Values of $n_1 = n_2$ at Which We Obtain the Following 95% Upper Prediction Bounds for $X = 0, 1, 2$

UPB	$X = 0$	$X = 1$	$X = 2$
1	1	1	•
2	2	2	2
3	4	3	3
4	16	4	4
5		7	5
6		26	7
7			14
8			95

TABLE 4. Upper $(1-\alpha)100\%$ Prediction Bounds for Zero-Failure Poisson Data, Where the Ratio $a = T_2/T_1$ Represents the Minimum Value Needed to Obtain the Listed UPB

$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	UPB
0.0001	0.0001	0.0001	0
0.1112	0.0527	0.0102	1
0.4625	0.2881	0.1112	2
0.8663	0.5833	0.2747	3
1.2849	0.8971	0.4625	4
1.7098	1.2187	0.6615	5
2.1377	1.5443	0.8663	6
2.5675	1.8723	1.0745	7
2.9984	2.2016	1.2849	8
3.4300	2.5320	1.4968	9
3.8622	2.8631	1.7098	10
4.2947	3.1946	1.9235	11
4.7276	3.5265	2.1377	12
5.1606	3.8587	2.3524	13
5.5939	4.1912	2.5675	14
6.0273	4.5238	2.7828	15
6.4608	4.8566	2.9984	16
6.8943	5.1895	3.2141	17
7.3280	5.5225	3.4300	18
7.7617	5.8555	3.6460	19
8.1955	6.1887	3.8622	20
10.3651	7.8552	4.9441	25
12.5353	9.5226	6.0273	30
6.8766	12.8586	8.1955	40
21.2186	16.1955	10.3651	50

ters, highly reliable test items in non-accelerated (working) environments often lead to zero failures of the items being tested. For this reason, the special case of zero-failure data warrants extra attention. Statistical inference, especially nonparametric techniques, can lead to ambiguous or misleading answers (e.g., estimated failure rates of zero). Maximum likelihood theory, as one example, fails to produce adequate measures of uncertainty to correspond to zero estimates for rates of failure for Poisson or binomial data. However, prediction problems are not so negatively affected.

Table 4 contains information for obtaining frequentist UPBs in zero-failure problems using Poisson data. If we observe $X = 0$ failures in time T_1, then we obtain a $(1-\alpha)100\%$ prediction interval for the future number of failures in time T_2 by first computing the ratio $a = T_2/T_1$. The table lists the minimum value of the ratio needed to obtain the UPB listed in the fourth column. That is, for values of a between two table values, the UPB is the smaller a value. For example, suppose we observe no failures with $T_1 = 2.0$, and we wish to construct a 95% UPB for the future number of failures based on $T_2 = 3.0$. With $a = T_2/T_1 = 1.5$, which falls between $a = 1.2187$ ($y^* = 5$) and $a = 1.5443$ ($y^* = 6$) on the table, we generate the UPB as $y^* = 5$.

Due to the inadequacy of maximum likelihood estimates in the zero-failure problem, most reliability testing theory for zero-failure problems apply Bayesian analyses (see Martz and Waller (1979), Mao et al. (1993), and Tang and Mao (1993), for examples). One can argue that Bayesian methods have gained popular support in the engineering sciences due in part to this convenience. In the context of PSA, predictive inference within the Bayes framework is discussed later.

Prediction for Multiple-Failure Models

If components in a group can fail at once due to a common cause, we might be interested in predicting, for a given time interval, the number of occurring catastrophic failures; i.e., the number of times all the components in the group fail simultaneously. This is a critical event in nuclear power stations, where system reliability is increased by adding redundant components to back up safety systems in case of component failure. Many current PSAs employ parametric models based on fundamental assumptions about component dependencies and common cause failure. If a stochastic relationship between components exists and is known, a simple prediction interval for a future number of catastrophic failures based only on the number of catastrophic failures observed in a test period will fail to exploit the information garnered from this relationship. If no useful relationship exists, it is often the case that no catastrophic failures have been observed, and PSAs should rely on prediction intervals based on zero failures, as discussed earlier. In this section, we use parametric CCF models to demonstrate prediction analysis for multiple failures.

There exists no strong consensus for parametric models dealing with multiple failure events. We will apply the binomial failure rate (BFR) model, introduced by Vesley (1977), in order to illustrate the construction of a prediction interval for a general parametric failure model. The BFR model represents one of many CCF models developed for PSA. In this case, it is assumed that the system has m identical compo-

nents and that system shocks occur to the component group according to a homogeneous Poisson process. Once a shock hits the system, each component has an independent and equally likely chance ($0 \leq p \leq 1$) of failure. Thus, the number of components failing due to a system shock has a binomial distribution. This model and several alternative models in CCF analysis are described fully in Mosleh et al. (1988).

Depending on the specific application, other models described in Mosleh et al. (1988) might be preferred to the BFR, which is chosen here for its somewhat frequent application and ease of illustration. Methodological problems with the model are discussed in Kvam (1993), and problems with model fit are discussed in Section 3.3 of Mosleh et al. (1988). Improvements made to the BFR model by Atwood (1986) increase the domain of problems that can be well approximated in PSA, and Kvam (1998a) extends the BFR technique to a mixture model that increases this domain further. In the following subsections, prediction for multiple-failure models is illustrated using the BFR model and the BFR mixture model. Generalizations to other CCF models will be apparent to the reader.

By applying a parametric failure model, we cannot effectively construct exact prediction bounds based on a single pivotal statistic. There is no unique method for forming the approximate UPB. We will outline an intuitive method based on calibrating the naïve UPB, which is based on ignoring the uncertainty associated with estimation of the failure rate parameters.

Calibration Technique for Computing UPB

To illustrate the calibration technique with a multiple failure model, suppose a system of five components is observed, and simultaneous failures occur according to the BFR model. Also suppose that, in one year of testing, we observe 37 CCF events of one component (out of five), 34 failures involving two components, 13 involving three components, and 4 failures involving four out of the five components. No catastrophic events that cause the failure of all five components are observed. We can summarize the failure events in an aggregate *impact vector* $\mathbf{x} = (37, 34, 13, 4, 0)$. Estimation of the BFR parameters (p = probability of component failure given a system shock, μ = Poisson rate for shock process) is not straightforward because we cannot observe x_0, the number of shocks that cause no failures. With some small samples, the method of maximum likelihood produces boundary solutions, and another estimation method, such as the method of moments technique used in Kvam (1996), would be preferred. An outline for constructing BFR parameter estimators is given in the Appendix.

According to the BFR model, the rate of catastrophic failure is μp^5. For a catastrophic failure to occur, all of the five system components must fail (with the same probability p) after a given common-cause shock. A naïve $(1 - \alpha)$ UPB for the number of catastrophic failures is the $(1 - \alpha)$ quantile of its multiple failure distribution, which is estimated to be Poisson with rate $\widehat{\mu}\widehat{p}^5$. That is, UPB is the smallest integer x for which $F(x; \lambda = \widehat{\mu}\widehat{p}^5) \geq 1 - \alpha$, where F is the Poisson cumulative distribution function. This upper bound is naïve in the sense that it ignores uncertainty associated with estimating the Poisson rate parameter.

Cox (1975), Atwood (1984), and Beran (1990) derived prediction bounds by finding a *calibration level* $(1 - \alpha_c)$ that produces the desired (true) confidence level $(1 - \alpha)$ for the prediction uncertainty once the uncertainty of the parameter estimates is included. Beran (1990) derived asymptotic properties needed for prediction calibration, and showed that the naïve UPB, based on substituting estimates $\widehat{\mu}\widehat{p}^5$ for μp^5, approaches the true UPB asymptotically, thus the need for calibration decreases with increasing sample size. For a given sample, the calibration level is approximated using Taylor series expansions of the maximum likelihood estimates when solving $F(x; \lambda = \widehat{\mu}\widehat{p}^5) = 1 - \alpha$. Unfortunately, this calibration method is very difficult to implement, and applications toward multiple failure models are unrealistic.

Alternatively, Escobar and Meeker (1999) use simulation methods to calibrate naïve prediction intervals. The application of the simulation method to complex prediction models is straightforward and can be implemented using the following steps:

1. Choose a confidence level $1 - \alpha_c$.
2. Compute estimates of the unknown parameters (e.g., via maximum likelihood) $\widehat{\phi} = (\widehat{p}, \widehat{\mu})$. Note that the UPB based on choosing the $1 - \alpha_c$ quantile from $F(x; \widehat{\mu}\widehat{p}^5)$ will have (actual) coverage probability of less than $1 - \alpha_c$, because the uncertainty of the estimates $(\widehat{p}, \widehat{\mu})$ is not reflected in the UPB.
3. Simulate an impact vector from the multiple failure model for which the parameter values

are $\phi = \widehat{\phi}$, so that the expected failure count is the same as the actual failure count observed in the data.

4. Compute estimates of the parameters based on the simulated data (call this $\widetilde{\phi} = (\widetilde{p}, \widetilde{\mu})$).
5. Compute the UPB for the number of catastrophic failures from the Poisson distribution; i.e., find the smallest value of x for which $F(x; \widetilde{\mu}\widetilde{p}^5) \geq 1 - \alpha_c$.
6. Generate a random value x^* from $F(x|\widehat{\mu}\widehat{p}^5)$.
7. In repeated iterations of steps [3]–[6], count the proportion of times in which the UPB in step [5] is smaller than x^*. This proportion is the estimated coverage probability $1-\widehat{\alpha}$ corresponding to the nominal coverage $1 - \alpha_c$.
8. Repeat steps [1]-[7] for different values of $1-\alpha_c$.

If we choose enough repetitions in step 7, the estimated coverage probability will approach the true coverage probability. The calibration needed to construct an accurate UPB is implied by the graph of $1 - \alpha_c$ *vs.* $1 - \widehat{\alpha}$. For the BFR example, this calibration curve is plotted in Figure 1. The dashed line indicates the *nominal* confidence level, and it serves to show how much confidence is lost due to parameter uncertainty. The curve is based on 10,000 simulations (in steps [3]-[6]) at 2 different values of $1-\alpha_c$ between 0.90 and 1.00. For example, if we specify a 0.90 confidence level when constructing a naïve UPB, the actual level is approximately 0.74. To achieve a true 0.90 upper bound, we need to specify a level of $1 - \alpha_c = 0.9733$. From the data, the maximum likelihood estimators for (p, μ) are (0.3031,105.31), and the estimated rate of catastrophic failure is $\widehat{\mu}\widehat{p}^5 = 0.2694$. The 0.9733 quantile of the Poisson distribution (with rate parameter $\lambda = 0.2694$) is 2, thus our 90% upper prediction bound for the number of catastrophic failures for the system of five components is 2.

Mixture Model Example

We can directly apply the calibration method for prediction to more complicated multiple-failure models without going into the detail of the model estimation. Hokstad (1988) suggested using a mixing distribution $G(p)$ for the parameter p to make BFR model applications more realistic. By doing this, we acknowledge the chance that system shocks can be generated from different sources which produce shocks of varying strength. Details on deriving the mixture distribution are given in the Appendix. We can choose a parametric form for G, such as beta(α, β) used by Kvam (1998b), or a nonparametric structure, as in Kvam (1998a), where the mixing distribution G is estimated as a discrete measure on $[0, 1]$ with between one and three points that have positive probability mass.

The mixture model was fit to CCF data involving one or more EDGs at various domestic nuclear power plants. The failure data for CCF events between 1980 and 1995, listed in a report for the U.S. Nuclear Regulatory Commission, IEEL (1997), are summarized in Table 6. Failures include natural disasters, shared design flaws, and machine maintenance errors, to name a few. The database also lists incipient failures, which include observations where a single diesel failed, but it was determined that its cause of failure could potentially cause the failure of other EDGs in the near future. In such cases, these potential failures are conservatively listed as actual CCFs. Component group sizes range from 2 to 5. Each group size refers to one or more nuclear power plants, and they are not considered identical. For each group size, we assign a separate (unknown) BFR rate parameter to obtain $(\mu_1, \ldots, \mu_5)$.

The time on test for each group is not available for public disclosure. As a consequence, we focus on computing the probability of a catastrophic failure given that a CCF event has occurred. For illustrative purposes, we will consider a future power plant that will employ four EDGs to back up the main power, and we will seek a 90% UPB based on the same amount of test time experienced by the power plants that used group sizes of four in Table 6.

The best fitting mixture model from Kvam (1998a) is a 2-point distribution for p with mass at $p_1 = 0.3103$ and $p_2 = 0.6788$. The estimated mixing probabilities are $P_G(p_1) \equiv \pi_1 = 0.6788$ and $P_G(p_2) \equiv \pi_2 = 0.3212$, respectively. For a future system of m components, the conditional probability of catastrophic failure given a CCF event is $\pi_1 {p_1}^m + (1 - \pi_1){p_2}^m$. Given a CCF event, the probability of catastrophic failure for a future group of four EDGs is estimated to be 0.1656. There are $n_T = 34$ CCF events listed for plants with group size four. For the mixture model, the probability of no failure (given a CCF event) is 0.1538, and we estimate the CCF shock rate for the group as $n_t/P(CCF) = 34/(1 - 0.1538) = 40.18$; thus, the Poisson rate of catastrophic failure for the group of size four is $40.18 \times 0.1656 = 6.6538$.

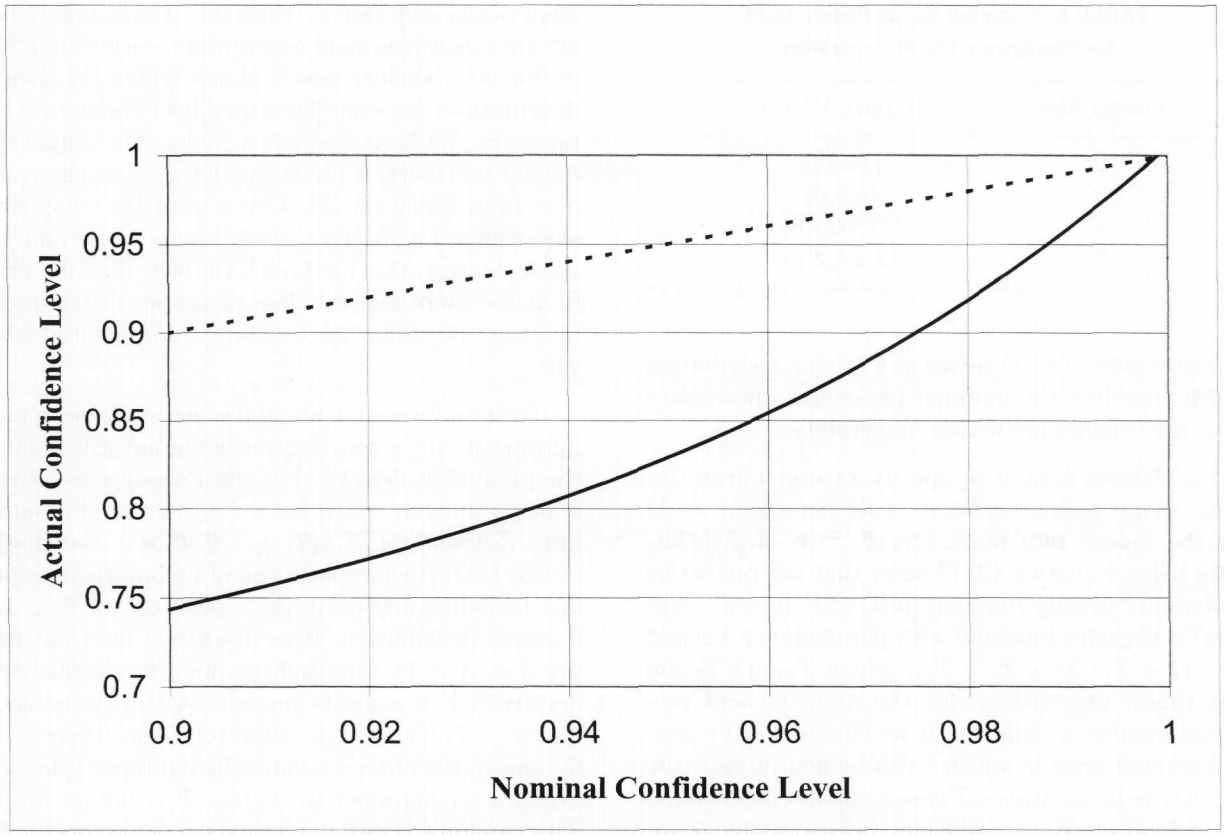

FIGURE 1. Calibration Curve for UPB on Catastrophic Common Cause Failures to a Five-Component System.

The calibration curve for the EDG upper prediction bound is plotted in Figure 2. Again, the dashed line represents the nominal confidence level. Along with the UPB for a future common cause group of size $m = 4$, Figure 2 provides calibration curves for group sizes of $m = 2$, 3, 4, and 5. Note that for the case $m = 5$, the figure implies that using a naïve interval for the UPB leads to gross underestimation of the predictive uncertainty. As before, the curves are each based on 10,000 simulations (in steps [3]-[6]) at 21 different values of $1 - \alpha_c$ between 0.90 and 1.00. To achieve a 90% UPB for the future number of catastrophic failures from a group of $m = 4$, we need to use $1 - \alpha_c = 0.9881$. The 0.9881 quantile of the Poisson distribution (with rate parameter $\lambda = 6.6538$) is 13, thus our 90% upper prediction bound for the number of catastrophic failures for the system of four EDGs is 13.

Bayesian Prediction

The goal of this paper is to present statistical methods for constructing prediction intervals in PSA. Up to this point we have presented the frequentist approach for prediction inference, but Bayesian methods for predictive inference are also applicable in certain PSAs. Furthermore, Bayes methods have become increasingly popular in risk assessment. For the Bayesian approach, a distribution $\pi(\theta)$ is assigned to θ in order to reflect user uncertainty in the parameter. This allows subjective prior information to affect the outcome of data analysis through $\pi(\theta)$. If the density of the observable data x is denoted $f(x \mid \theta)$, and the parameter θ has prior distribution $\pi(\theta)$, information is updated (using Bayes rule) through the conditional posterior distribution

$$\pi(\theta \mid x) = \frac{\pi(\theta) f(x \mid \theta)}{\int \pi(\theta) f(x \mid \theta) d\theta}.$$

Unlike the frequentist approach, predictive inference is straightforward using a Bayesian framework; we construct the *predictive density* of a new observation Y from $f(x \mid \theta)$ as

$$P_{Y|X}(y \mid x) = \int_{-\infty}^{\infty} f(y \mid \theta) \pi(\theta \mid x) d\theta. \qquad (2)$$

TABLE 6. Common Cause Failure Data for Emergency Diesel Generators

Group Size	Impact Vector
2	(17,14)
3	(9,5,6)
4	(11,10,7,6)
5	(2,2,1,2,1)

The posterior $\pi(\theta \mid x)$ serves as a mixing distribution which combines the updated parameter uncertainty with the original prediction uncertainty.

For Poisson data, if we specify a gamma prior (integer shape parameter $= r$, scale parameter $= \lambda$) for the failure rate parameter θ, it is straightforward using Equation (2) to show that the predictive probability density function (pdf) of Y (given x failures) is negative-binomial with parameters $r+x$ and $p = (\lambda + T_1)/(\lambda + T_1 + T_2)$, where T_1 and T_2 are test times, respectively, for the observed and predicted number of failures. If we choose a more general gamma prior in which r can be non-integer, the predictive distribution is Poisson-gamma (see Section 8.3 of Johnson, Kotz, and Kemp (1993)) with parameters $r+x$, $\lambda + T_1$, and T_2. Of course, other reasonable prior distributions can be chosen for θ. We chose the gamma distribution for illustration because it is a *conjugate family* for the Poisson distribution; that is, by using a Poisson likelihood with a gamma prior, the posterior distribution is also gamma.

With binomial data, we can choose the conjugate prior $\theta \sim \text{beta}(\alpha_0, \beta_0)$. From Equation (2), the predictive density for $Y \sim \text{binomial}(n_2, \theta)$ given $X \sim \text{binomial}(n_1, \theta)$ can be calculated as beta-binomial$(\alpha + x, \beta + n_1 - x, n_2)$ with

$$E(Y \mid x) = \frac{n_2(\alpha + x)}{(\alpha + \beta + n_1)},$$

$$\text{Var}(Y \mid x) = \frac{n_2(\alpha + x)(\beta + n_1 - x)(n_1 + n_2 + \alpha + \beta)}{(\alpha + \beta + n_1)^2(\alpha + \beta + n_1 + 1)}.$$

Unlike prediction intervals outlined in the previous section, one-sided Bayesian prediction bounds are made simply by using appropriate percentiles from the predictive pdf. Two-sided prediction intervals are constructed in the same manner as regular Bayesian credible sets.

If one can appropriately apply prior information with a conjugate prior distribution, the resulting Bayesian inferential analysis is straightforward and analytically convenient. However, nonconjugate priors are sometimes more appropriate, such as in PSAs of domestic nuclear power plants where log-normal distributions are sometimes used for Poisson rate parameters. Without conjugacy, we lose the simple and elegant solutions for predictive densities which we expect from Equation (2). Given available computing algorithms (i.e., Markov chain Monte Carlo and numerical integration methods), the selection of a prior from the more general class of proper distributions no longer precludes the feasibility of inferential analysis.

If prior information about θ is not available, a noninformative prior distribution can be substituted into the prediction density. For the Poisson case, θ is a scale parameter, which has a Jeffries noninformative prior distribution of $\pi(\theta) = 1/\theta$, $\theta > 0$ (see Berger (1985) for further explanation of noninformative priors, including Jeffries priors). Note that $\pi(\theta)$ is not a proper distribution. If we observe at least one failure (i.e., $X \geq 1$), the resulting predictive probability density of Y is negative binomial with parameters x and $p = T_1/(T_1 + T_2)$. However, if we observe no failures in the observed data, the improper prior will produce an improper predictive distribution for Y. This can limit Bayesian methods in environments for which zero-failure experiments are not uncommon, as discussed in the last section.

As an alternative, the analyst may choose *maximum entropy priors*, as defined in Kapur (1989), for which the prior $\pi(\theta)$ maximizes an entropy measure subject to constraint, such as a fixed prior mean. For Poisson data, we specify the Poisson parameter θ in terms of the prior distribution with $E[\theta] = 1/\lambda$, where λ is known. If we define entropy as

$$H_\pi = -\int_{-\infty}^{\infty} \pi(\theta)\ell\text{n}(\pi(\theta))d\theta,$$

then the maximum entropy prior for $\theta \in (0, \infty)$ is the exponential distribution with mean λ^{-1} (see Kapur (1989)). Furthermore, it can be shown that without this constraint, the maximum entropy prior does not exist. Using an exponential distribution (with failure rate parameter λ), the posterior distribution $\pi(\theta \mid X = 0)$ is gamma with parameters $(1, \lambda + 1)$, and the predictive density of Equation (3) is negative binomial with parameters 1 and $p = (\lambda + 1)/(\lambda + 2)$. In the binomial data case, θ has support on $(0, 1)$, the Jeffries noninformative prior is beta(1/2,1/2), and the maximum entropy prior is the uniform distribution, or beta(1,1).

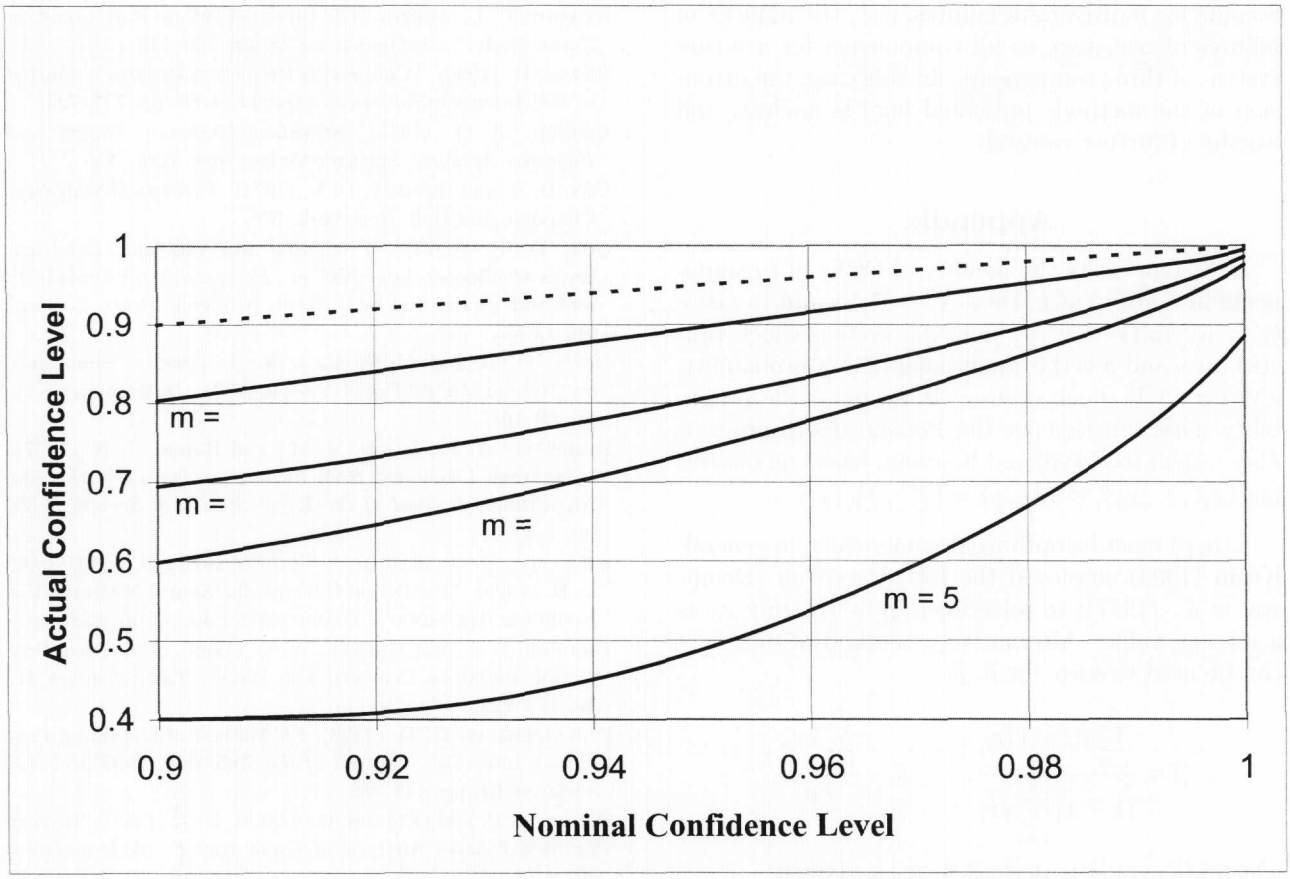

FIGURE 2. Calibration Curves for UPB on Catastrophic Common Cause Failures to EDG in Group Sizes of $m = 2, 3, 4, 5$.

Discussion

Engineering practitioners have been apprehensive about applying predictive inference to problems that involve the estimation of future failures. Statistical research for predictive inference in the physical sciences, however, has surged in recent years. See, for instance, De Veaux et al. (1998), Davis and McNichols (1999), and Escobar and Meeker (1999). Engineering examples for which predictive inference was not considered in the initial design of the experiment are referenced in those papers, suggesting that predictive inference might serve as an afterthought in some engineering sciences. Inference for only the unobservable parameters can be unnecessary, if not misguided, in many such applications. This is especially true if the parameters in the statistical model have no direct operational meaning to the analyst. For the simple discrete prediction problems described earlier, the pivotal statistic used in each case leads to the derivation of predictive models that are free of unknown parameters. In a PSA, this allows one to focus on future failures without the necessity of estimating failure rates.

In multiple failure models, prediction bounds based on simple pivotal statistics are not generally obtainable. Approximate prediction bounds derived in this paper lack the simplicity or the aesthetic appeal of exact bounds, but they can be especially useful in nuclear power plant studies, where CCFs are a crucial part of the PSA. Although the simulation-based calibration method suggested in this paper is computationally intensive, its implementation is straightforward, and it can guarantee an accurate estimate of the bound if enough simulations are executed. For general confidence intervals based on simulation methods, 2000 runs are often suggested (see Escobar and Meeker (1999), for example) but we ran 10,000 for the results in Figures 1 and 2.

Here we considered the problem of predicting catastrophic failures for common cause failure environments. Future problems in probabilistic safety assessment may include simultaneous prediction

bounds for multivariate failures; e.g., the number of failures of one, two, or all components for a future system of three components. In this case, the extension of the methods presented here is unclear, and worthy of further research.

Appendix

If X_i represents the number of CCFs of i components in a group of k, then $X_i \sim \text{Poisson}(\theta_i)$, where $\theta_i = \mu\binom{k}{i}p^i(1-p)^{k-i}$, μ is the system shock process rate, and p is the conditional failure probability, given a CCF shock occurs. If we write the probability mass function for the Poisson distribution as $P_\theta(x)$, then the likelihood function, based on observing $(x_1, \ldots, x_k)$, is $L(\mu, p) = \prod_{i=1}^{k} P_{\theta_i}(x_i)$.

$L(\mu, p)$ must be optimized numerically, in general. Kvam (1993) employed the EM Algorithm (Dempster et al. (1977)) to solve for $\widehat{\mu}$, $\widehat{p}$ by treating x_0 as a missing value. Alternatively, method of moments can be used to solve for $\widehat{\mu}$, $\widehat{p}$:

$$\widehat{p} = \frac{\sum_{i=1}^{k} i(i-1)x_i}{(k-1)\sum_{i=1}^{k} ix_i}, \qquad \widehat{\mu} = \frac{\sum_{i=1}^{k} ix_i}{k\widehat{p}},$$

which are explicitly derived in Kvam (1993).

For the BFR mixture model, a mixing distribution $G(p)$ is assigned to the parameter p, so the distribution for X_i is now

$$P(X_i) = \int_0^1 P_{\theta_i}(X_i \mid p)dG(p),$$

where $G(p)$ can be specified as a beta distribution (Kvam (1998b)) or unspecified as a nonparametric mixing distribution (Kvam (1998a)). Both references offer iterative methods that can be used to estimate G.

Acknowledgment

The authors are grateful for the comments by the referees that led to significant improvements in the paper. The research of both authors was supported by National Science Foundation Grant DMI-9700527.

References

ATWOOD, C. L. (1984). "Approximate Tolerance Intervals Based on Maximum Likelihood Estimates". *Journal of the American Statistical Association* 79, pp. 459–465.

ATWOOD, C. L., (1986). "The Binomial Failure Rate Common Cause Model". *Technometrics* 28, pp. 139–148.

BERAN, R. (1990). "Calibrating Prediction Regions". *Journal of the American Statistical Association* 85, pp. 715–723.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, NY.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, New York, NY.

COX, D. R. (1975). "Prediction Intervals and Empirical Bayes Confidence Intervals" in *Perspectives in Probability and Statistics* edited by J. Gani, Academic Press, London, pp. 47 55.

DAVIS, C. B and MCNICHOLS, R. J. (1999). "Simultaneous Nonparametric Prediction Models". *Technometrics* 41, pp. 89 101.

DEMPSTER, A. P.; LAIRD, N. M.; and RUBIN, D. B. (1977). "Maximum Likelihood from Incomplete Data Via the EM Algorithm". *Journal of the Royal Statistical Society* B 39, pp. 1–39.

DE VEAUX, R. D.; SCHUMI, J.: SCHWEINSBERG, J. and UNGAR, L. H. (1998). "Prediction Intervals for Neural Networks Via Nonlinear Regression". *Technometrics* 40, pp. 273 282.

ESCOBAR, L. A. and MEEKER, W. Q. (1999). "Statistical Prediction Based on Censored Life Data". *Technometrics* 41, pp. 113–124.

FAULKENBERRY, G. D. (1973). "A Method of Obtaining Prediction Intervals". *Journal of the American Statistical Association* 68, pp. 433–435.

GAVER, D. P. and O'MUIRCHEARTAIGH, I. G. (1987). "Robust Empirical Bayes Analysis of Event Rates". *Technometrics* 29, pp. 1–15.

HAHN, G. J. and MEEKER, W. Q. (1991). *Statistical Intervals: A Guide for Practitioners*. John Wiley and Sons, New York, NY.

HOKSTAD, P. (1988). "A Shock Model for Common Cause Failures". *Reliability Engineering & System Safety* 23, pp. 127–145.

IDAHO ENGINEERING AND ENVIRONMENTAL LABORATORY (1997). *Common Cause Failure Database and Analysis System, Volume 3-Data Collection and Event Coding* INEEL/EXT-97/00696, June, 1997.

JOHNSON, N. L.; KOTZ, S.; and KEMP, A. W.: (1993). *Univariate Discrete Distributions*, 2nd ed. John Wiley and Sons, New York, NY.

KAPUR, J. N. (1989). *Maximum Entropy Models in Science and Engineering*. Wiley Eastern Limited, New Delhi.

KNÜSEL, L. (1994). "The Prediction Problem as the Dual Form of the Two-Sample Problem with Applications to the Poisson and the Binomial Distribution". *The American Statistician* 48, pp. 214 219.

KVAM, P. H. (1993). "Computational Problems with the Binomial Failure Rate Model and Incomplete Common Cause Failure Reliability Data" in *Proceedings of the 25th Symposium on the Interface : Computing Science and Statistics* edited by M. E. Tartar and M. D. Lock, Interface Foundation of North America, pp. 107–111.

KVAM, P. H. (1996). "Estimation Techniques for Common Cause Failure Data with Different System Sizes". *Technometrics* 38, pp. 382–388.

KVAM, P. H. (1998a). "The Binomial Failure Rate Mixture Model for Common Cause Failure Data from U.S. Nuclear Power Plants". *Applied Statistics: Journal of the Royal Statistical Society* C 47, pp. 49–61.

KVAM, P. H. (1998b). "A Parametric Mixture Model for Common Cause Failure Data". *IEEE Transactions on Reliability* 47, pp. 337–352.

LOFGREN, E. V. and GREGORY, S. H. (1991). "Issues and Approaches for Using Equipment Reliability Alert Levels". *Report BNL-NUREG-52251 (NUREG/CR-5611)*, Brookhaven National Laboratory, Long Island, New York.

MAO, S.; XIA, J.; and WENQI, G. (1993). "The Reliability Analysis of Zero-Failure Data for Bearing Life Test". *Chinese Journal of Applied Probability and Statistics* 9, pp. 326–331.

MARTZ, H. F. and WALLER, R. A. (1979). "A Bayesian Zero-Failure (BAZE) Reliability Demonstration Testing Procedure". *Journal of Quality Technology* 11, pp. 128–138.

MOSLEH, A.; FLEMING, K. N.; PARRY, G. W.; PAULA, H. M.; WORLEDGE, D. H.; and RASMUSON, D. M. (1988). "Procedures for Treating Common Cause Failures in Safety and Reliability Studies". *NUREG/CR-4780, EPRI NP–5613, Volume 1 and Volume 2*, prepared for the Nuclear Regulatory Commission and the Electric Power Research Institute by Pickard, Lowe, and Garrick, Inc., January 1988.

NELSON, W. (1982). *Applied Life Data Analysis*. John Wiley & Sons, New York, NY.

TANG, D. and MAO, S. (1993). "Bayesian Analysis for Exponential Zero-Failure Data". *Proceedings of the Third Pacific Area Statistical Congress* STMA V35 1447, pp. 513–520.

SANZO, D. L.; KVAM, P. H.; APOSTOLAKIS, G.; WU, J. K.; MILLICI, T.; and GHONIEM, N. (1994). "Survey and Evaluation of Aging Risk Assessment Methods and Applications". NUREG/CR-6157, prepared for the U. S. Nuclear Regulatory Commission.

VESLEY, W. E. (1977). "Estimating Common Cause Failure Probabilities in Reliability and Risk Analysis : Marshall-Olkin Specializations" in *Nuclear Systems Reliability Engineering and Risk Assessment* edited by J. B. Fussell and G. R. Burdick, Philadelphia : Society for Industrial and Applied Mathematics, pp. 314–341.

WORLEDGE, D. H.; STRIGHAM, R. S.; and MCCLYMONT, A. S. (1982). "PWR Power Plant Reliability Data". *Interim Report NP-2592*, Electric Power Research Institute.

Key Words: *Bayesian Prediction, Binomial Distribution, Calibration, Common Cause Failure, Mixture Distributions, Poisson Distribution.*