CONF-930116--35

BNL-NUREG--48226

DE93 006389

ON THE PRIOR PROBABILITIES FOR TWO-STAGE BAYESIAN ESTIMATES*

P. KOHUT
BROOKHAVEN NATIONAL LABORATORY
UPTON, NY 11973

ABSTRACT

The method of Bayesian inference is reexamined for its applicability and for the required underlying assumptions in obtaining and using prior probability estimates. Two different approaches are suggested to determine the first-stage priors in the two-stage Bayesian analysis which avoid certain assumptions required for other techniques. In the first scheme, the prior is obtained through a true frequency based distribution generated at selected intervals utilizing actual sampling of the failure rate distributions. The population variability distribution is generated as the weighed average of the frequency distributions. The second method is based on a non-parametric Bayesian approach using the Maximum Entropy Principle. Specific features such as integral properties or selected parameters of prior distributions may be obtained with minimal assumptions. It is indicated how various quantiles may also be generated with a least square technique.

INTRODUCTION

The method of Bayesian inference is widely used to estimate specific system failure rates and to incorporate experience on similar systems. Two somewhat different methods have been developed, one is the most widely used two-stage process of Kaplan[1], and the other is a more conventional technique using parametric Bayesian estimation[2]. Each of these approaches tries to extract information regarding population variability from data obtained with similar systems and incorporate it through the judicial selection of appropriate prior distributions.

The object of this paper is to illuminate the problem associated with the determination of the prior and point out certain difficulties which arise in some of the techniques. In addition it suggests two new approaches, which avoid some of the assumptions of other models.

METHOD - A

In the two-stage method, the experiences of similar plants are incorporated through a population variability distribution. Since this is not known, a family of lognormal distributions are laid over the discretized parameter space (μ_i and σ_i). The operating experience, events k_i, and exposure time t_i, satisfy a Poisson-distribution, $p(k_i|t_i, \lambda)$ with λ=failure rate, which may be combined with the assumed lognormal form of the population variability $p(\lambda|\mu, \sigma)$ to obtain a likelihood function.

Kaplan's original procedure was criticized by Frohner[3] on the basis of using uniform initial priors, discretization of a continuous problem and for complexity. His suggested technique of the analytical approach may seem simpler, but as Kaplan pointed out in a comment[4], the conceptual derivation of the first-stage prior through simple averaging of the $p(\lambda_n|E_1, E_2)$ distributions (where E_1 stands for design etc., and E_2 for plant experience) does not contain the true information on the frequency distribution or population variability of the failure rates.

The potential problems of the two-stage approach lies in part in the discretization of the population space. The assumption that the variability distribution is

* This work was performed under the auspices of the U. S. Nuclear Regulatory Commission.

lognormal $p(\lambda/\mu, \sigma)$ is not unique and is not considered by this author as a serious problem. Nonunimodal distributions are probably better handled as separate populations. However, by discretizing the parameter space, μ_i, σ_i, it is assumed that the resulting likelihood function is smooth and well behaved. Therefore the discretization will not "miss" certain irregular or rapidly changing portions of the surface.

There is very little *a priori* information on how to perform the discretization and how to limit the number of parametric priors. The process also involves discrete summation over infinite ranges. In practice, these considerations are generally dealt with in some approximate manner and usually no major problems arise.

The basic problem involves the estimation of the failure rate λ_i for a component or system i given observed k_i event in exposure time t_i, where $i=1,2..N$ represents the variation over the population. The events k_i satisfy the Poisson-distribution with an MLE estimate of $\lambda_i = k_i/t_i$.

Now λ_i is a sample from the system probability distribution, which may be derived as a gamma distribution $g(\lambda/k_i+1, t_i)$, if one uses a uniform prior (Frohner used $1/\lambda$ prior to get $g(\lambda/k_i, t_i)$). The $g(\lambda/k_i+1, t_i)$ distribution is a true probability and reflects our degree of confidence in λ_i given k_i, t_i. Summing and averaging these distributions (as Frohner does) has little information content on the frequency or variability distribution of the λ_i's, except for very large samples when the distribution becomes very narrow around $\lambda_i \sim \langle\lambda\rangle$.

Instead, the following procedure may be used to extract the frequency or variability information. The $g(\lambda/k_i, t_i)$ distribution may be inverted to find λ for a given degree of confidence, i.e.,

$$p(\lambda)d\lambda = g(\lambda/k_i, t_i) = \text{const.},$$

leading to the associated $\lambda_j(p_j(\lambda)d\lambda = \text{const.})$. Now, the *pdf* may be discretized $[..p_{j-1}d\lambda, p_j d\lambda, ..]$, with each $p_j d\lambda$ defining a confidence level. Effectively, the distribution $g(\lambda/k_i, t_i)$ is inverted for each $i=1,2..N$ to find a sequence of λ_j's.

This set of λ's for fixed j's form a sample of the variability distribution which is assumed to be a gamma distribution $g_j(\alpha,\beta/\lambda_j)$ (one could select lognormal distributions just as well) and as such a simple frequency plot may be established to obtain the parameters for example using sample moment estimators.

The population variability distribution is simply the mean value of $g_j(\alpha,\beta/\lambda_j)$ with $j=1,2..J$

$$p(\lambda/E_1, E_2) \propto \sum_j w_j\, p(\lambda/A, B)\, g_j(\alpha,\beta/\lambda_j)$$

where E_1 and E_2 represents the "engineering knowledge" and past experience of similar systems and $p(\lambda/A, B)$ is the prior for the population variability distribution. The weights w_j reflect the statistical confidence attached to the frequency plot. The distribution $p(\lambda/E_1, E_2)$ may then be used as the prior for the final update or plant specific data containing information about similar plants.

The main difference between the proposed method and the usual two-stage approach is that a true frequency distribution is established at selected intervals. The parameter space (α, β) is not arbitrarily discretized, but obtained only at selected values reflecting the actual sampling of the failure rate distribution.

The construction of actual frequency plots at discretized intervals assures that the information contained in the data from similar system is used in the final prior. This is also the main difference between the two-stage and the empirical Bayesian methods, where effectively only the mean value obtained from the experience (k_i, t_i) is utilized usually through a moment matching scheme to generate in one step the prior distribution $p(\lambda/E_1, E_2)$.

A computer program implementing the above described technique was developed and Figure 1 indicates, for one example, the results obtained for a problem treated in Ref. 1.

METHOD - B

The estimation of component or system failure rates usually comprises of the construction of appropriate probability distributions based on observed events and collected data. The data normally consist of k_i observed events associated with exposure time t_i with $i=1,2..N$. The sequence of observation or pairs of (k_i, t_i) may represent a series of measurements or pooled data from similar systems or components.

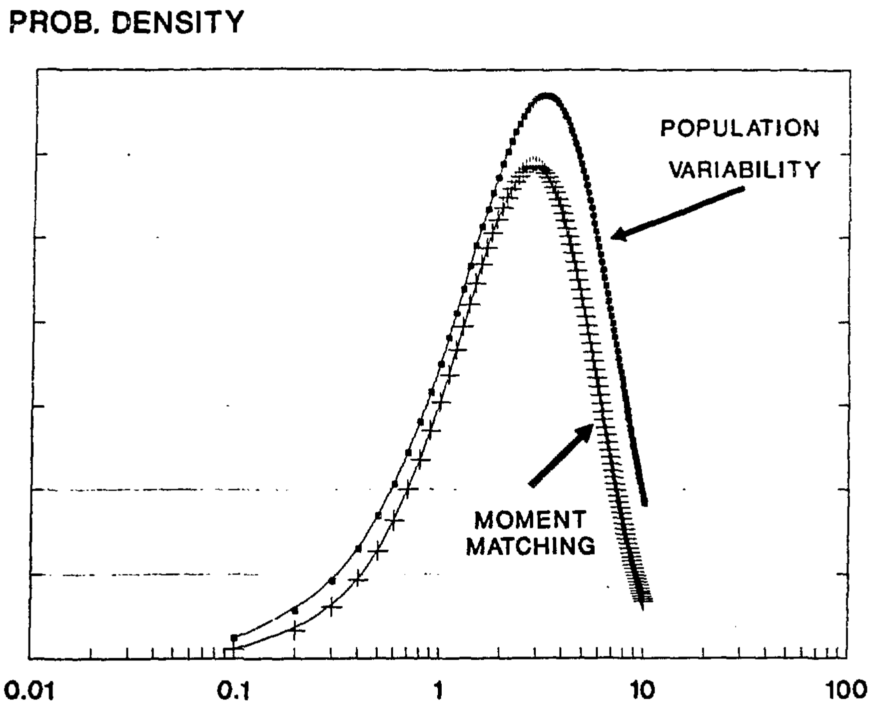

FIGURE 1 FREQUENCY, OCCURRENCE PER YEAR

The present analysis is restricted to normal Poisson-type processes that is the observable events k_i satisfy the Poisson-distribution $p(k_i/t_i, \lambda)$, where λ is the failure rate with MLE estimate of $<\lambda>=k_i/t_i$. The basic problem with pooling data of similar components or systems is that the individual failure rates are not identical, but are distributed in some manner.

In the usual two-stage Bayesian approach[1,3] or moment matching methods[2] an assumption is required for the probability distribution using free parameters. These parameters are then determined by some adjustments to the data. However, there is no *a priori* information on the underlying density function and naturally the analyst is free to choose between a large class of suitable distributions.

These type of problems are fairly common and has given impetus for the use of non-parametric techniques. In these methods the principle of maximum-entropy[5] is utilized to obtain some information on the *pdf* underlying probability density function (*pdf*) using minimal assumptions. The independence from the functional form of the underlying *pdf* is usually achieved by working with the cumulative probability distribution $q(x)=\int pdf(x)dx$.

Combining the maximum-entropy principles with Bayesian analysis has resulted in powerful new techniques. The primary purpose of this Section is to introduce and apply the non-parametric Bayesian approach (following Ref.6) to establish certain characteristic features of the population variability distribution or first stage prior of the usual two-stage updating process and potentially determine the most suitable parametric form.

The Maximum Entropy Principle (ME) states that the most optimum choice of the density function $pdf(x)$ should be the one which has the maximum entropy defined as

$$H = - \int pdf(x) \log(pdf(x) \frac{1}{v(x)})dx$$

under predefined constraints of the moments of the $pdf(x)$. For example to determine the most likely median of the distribution the constraint would take the form of $q(x)=\int pdf(x)dx=1/2$. From the Bayesian perspective this may be viewed as our initial information and may serve as the starting point to establish a posterior distribution.

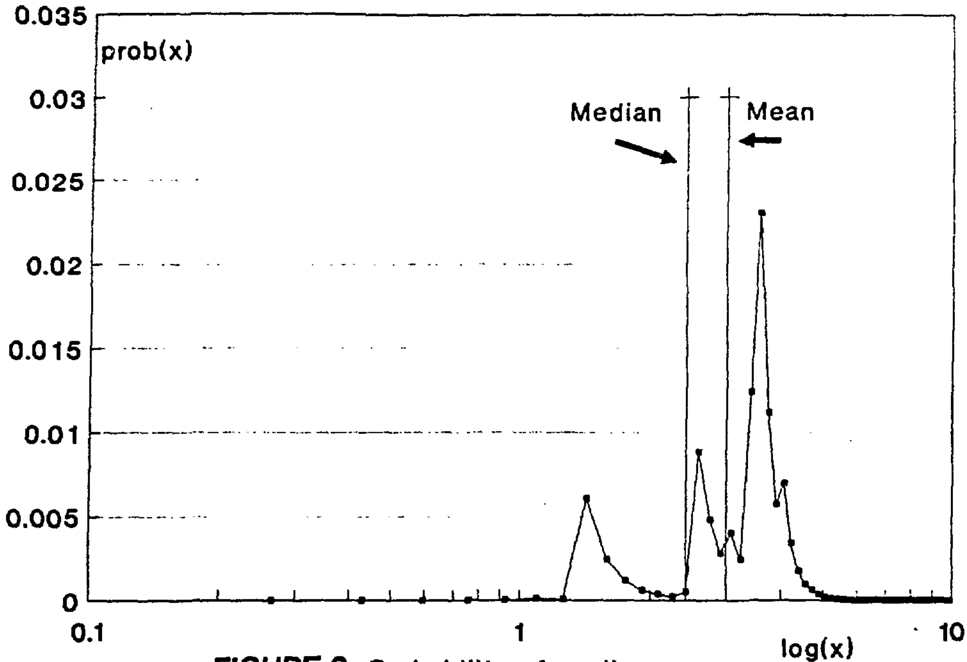

FIGURE 2 Probability of median

Given a set of data points $\{\lambda_i\}$ on the interval $[a,b]$ the distribution of the median is assumed to be uniform over $[a,b]$ (uniform prior). It turns out that the final answer is not very sensitive to this particular selection of the prior. The appropriate form of the pdf maximizing the entropy with the given constraint has the form (for one sample) $pdf(\lambda')=1/2(\lambda-a)$ when $\lambda'<\lambda$ and $pdf(\lambda')=1/2(b-\lambda)$ when $\lambda'>\lambda$, where λ is the assumed value of the median. The likelihood function is just the product of these pdf's for each data point.

With the uniform initial prior the posterior distribution becomes

$$pdf(\lambda \mid q=1/2) = \left(\frac{1}{2(\lambda-a)}\right)^{n_1} \left(\frac{1}{2(b-\lambda)}\right)^{n_2}$$

with $n_1+n_2 = N$, where n_1 and n_2 are the number of data points in the respective interval. The above expression may easily be generalized to arbitrary quantiles by replacing $1/2$ by q when $\lambda'<\lambda$ and $1-q$ for $\lambda'>\lambda$. It must be emphasized that this process gives information about the characteristics of the underlying pdf and not simply gives an estimate of the sample median.

Figure 2 indicates the results or the probability distribution of the median for the Loss of Feedwater problem from Ref. 1. Note that the distribution is not smooth and purposely not fitted by a curve. This is in stark contrast with the usual parametric picture where the distribution is smoothly fitted through the data. The two additional points are the median and mean of the parametric lognormal distribution fitted by a moment method to the data.

The non-parametric estimate seems to predict a median similar to the best estimate results of Ref. 1 which is somewhat higher than the maximum likelihood value. Figure 3 indicates the distribution of various quantiles (.05, .5, and .95) and again the results are very reasonable along the lines of intuitive expectations.

The non-parametric Bayesian technique may be applied to other questions such as determining the mean or combination of mean and variance of the *pdf*. The question of the mean leads to an undetermined constrain for a uniform prior, which may be handled by restricting the solution to finite interval leading to transcendental equations. If the first two moments are known, the entropy is maximized by a normal distribution.

The ME methodology is useful in selecting and examining the properties of the data and helps to

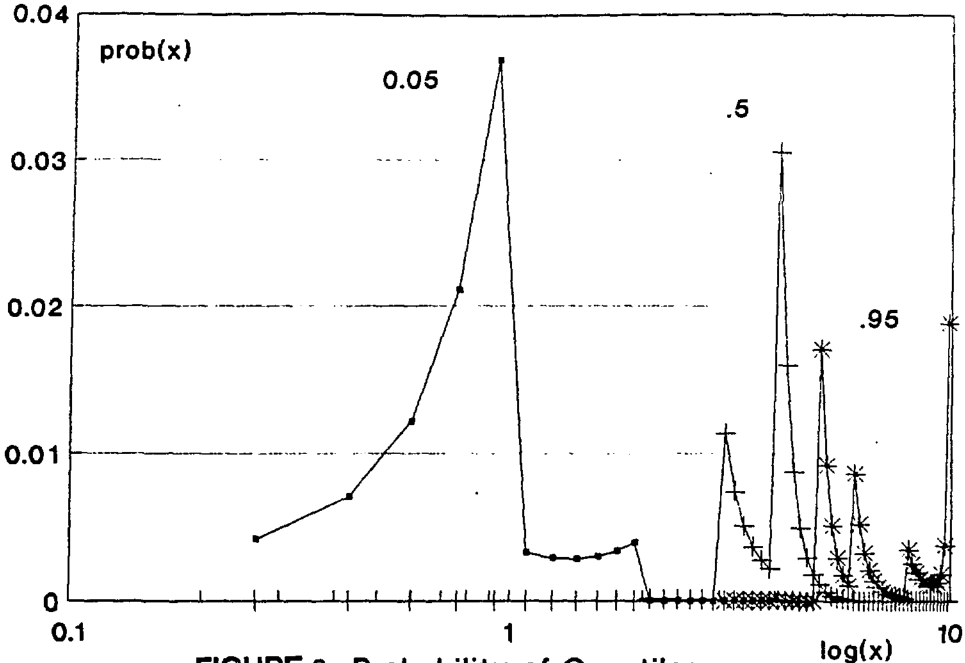

FIGURE 3 Probability of Quantiles

determine for the analyst the nature of the underlying distribution. It also gives sufficiently robust information on the integral properties helping the analyst to select the appropriate parametric distributions if required. For example, by generating the various quantiles of the underlying *pdf* a suitable parametric distribution may be selected, if required.

SUMMARY

The present paper suggests the utilization of novel approaches to determine the first-stage prior in the two-stage Bayesian analysis by avoiding certain assumptions required for other methods. The first method is similar to the usual first-stage of the two-stage process with the exception that it relies on true frequency distributions in obtaining the population variability distribution.

The second technique is a non-parametric approach utilizing the ME principle. It is capable of producing acceptable and reasonable quantiles to various quantities and may further be extended to generate the parameters for distributions in a least square sense.

REFERENCES

1. S. Kaplan, "Bayesian Methods in PRA," *Proceedings of the Conference on Advances in Reactor Computations,* Vol. II., ANS, Salt Lake City, 1983.

2. J.K. Vaurio, "On Analytic Empirical Bayes Estimation of Failure Rates," *Risk Analysis,* Vol. 7, No. 3, 1987.

3. F.H. Frohner, "Analytic Bayesian Solution of the Two-Stage Poisson-Type Problem in Probabilistic Risk Analysis," *Risk Analysis,* 5(3), 1985.

4. S. Kaplan, Comment on Ref. 3, see *Risk Analysis,* Vol. 5, No. 3, pg. 227, 1985.

5. E.T. Jaynes, "Prior Probabilities," *IEEE Trans. SSC-4,* 1968.

6. S.F. Gull, J. Fielden, "Bayesian Non-Parametric Statistics," *Proceedings of the Fourth Maximum Entropy Workshop,* Ed., J.H. Justice, University of Calgary, (1984).

DISCLAIMER