BAYESIAN INFERENCE IN RELIABILITY

THE RELATIONSHIP BETWEEN RELIABILITY AND PROBABILITY

Reliability engineering as a technical discipline is concerned with ensuring that systems perform their function successfully. It aims at identifying failures, determining their causes, and assessing their probabilities. Therefore, one of the major topics in reliability deals with assessing the level of reliability, defined as the probability that an item (system, component, part, etc.) performs its specified function. Formally,

$$\text{Reliability} = R = \text{Prob (system functions as specified)}$$

Being a probability, the value of R satisfies the condition $0 \leq R \leq 1$. Sometimes R is estimated *directly* based on data and/or other available information, but often its value is determined *indirectly* through a function that represents a mathematical, logical, or physical model. Such models are typically written as a function of a set of parameters Θ:

$$R = f(\Theta)$$

Whether directly estimated or developed as a function of other parameters, estimation of R involves assessing unknown quantities using relevant information, such as data from field operation, tests, or engineering analysis and judgment. Bayesian reliability refers to a set of probabilistic and statistical concepts and methods that are used to estimate the reliability function and its parameters by using available information. This article describes some of the basic principles of Bayesian statistical inference and their relationship to reliability. Several examples show how these principles are applied to typical reliability estimation problems. Further reading on the subject can be found in Refs. 1, 2, and 3, among a vast literature on reliability applications of Bayesian methods.

AN INTRODUCTION TO BAYESIAN STATISTICS

Meaning of Probability

There are two schools of thought regarding the meaning, and consequently the application, of probability: (1) frequentist and (2) subjectivist. These are also known as the classical and Bayesian schools, respectively. According to the frequentist or classical interpretation, probability is the limiting relative frequency of occurrence of an event when the experiment or trial, in which the event in question is an outcome, is repeated a large number of times. Formally, the probability of event E is given by

$$\Pr(E) = \lim_{N\to\infty} \frac{N_E}{N}$$

where N_E is the total number of times that event E occurs, $N_{\bar{E}}$ is the total number of trials in which E does not occur, $N = N_E + N_{\bar{E}}$ is the total number of trials, and $\Pr(E)$ is the limiting ratio if such limit exists.

According to the subjectivist (Bayesian) school, probability is the degree of *confidence* in the truth of a *proposition*. A proposition is a statement which can, in principle, be proven true or false. For example the statements, “It will rain tomorrow” and “2 + 2 = 4” are both valid propositions subject to verification by observation, logical reasoning, or experimental or theoretical verification. On the other hand, the statement “It may rain tomorrow” is not a proposition because its truth or falsehood can never be shown. In general, a proposition is a statement with a yes or no answer. The *degree of confidence* is a measure of personal belief or an indication of how much one knows about the proposition or event in question. As such, it is subjective and personal. It is a measure of uncertainty (or certainty) and thus a representation of a state of mind and *not* the outside world. However, such a measure is objective in the sense that it encodes an objective state of a person's mind and represents a degree of knowledge.

According to the subjectivist school, probability as an objective entity outside our minds does not exist. However, any two individuals with the same totality of knowledge, information, and biases will assign the same probability value to the truth of a proposition. The only requirement is coherence, that is, one's subjective probability of an event must be consistent with one's body of knowledge and must obey the calculus of probability, that is, it must satisfy the axioms of the theory of probability.

Bayes's Theorem

A simple but extremely powerful relationship known as Bayes's theorem (4) is developed on the basis of the notion of conditional probability:

$$\Pr(A|E) = \frac{\Pr(E|A)\Pr(A)}{\Pr(E)}$$

$\Pr(A|E)$ is the *posterior* probability of A given evidence that event E has occurred, and $\Pr(E|A)/\Pr(E)$ is the relative *likelihood* of the evidence (or occurrence of event E) assuming the occurrence of A, and $\Pr(A)$ is the *prior* probability of event A. In this formulation $\Pr(E) = \Pr(E|A)\Pr(A) + \Pr(E|\bar{A})\Pr(\bar{A})$.

Interpreted in the language of a subjectivist, $\Pr(A|E)$ is the *posterior* or "updated" degree of confidence in the occurrence of A knowing that E has occurred, that is, the degree to which one believes that A is true when proposition E is true. In this context $\Pr(A)$ is the degree of confidence *prior* to receiving or incorporating evidence E (proposition E is true). Therefore, Bayes's theorem provides the mechanism for updating one's degree of knowledge about a proposition (e.g., occurrence of an event) in the light of new evidence.

Updating Probability Distributions

The current state of knowledge or degree of belief about an unknown quantity X represented by a probability distribution may change in light of new evidence E. Bayes's theorem is used to obtain the "updated" or posterior state of knowledge, given the new information.

Updating Discrete Probability Distributions. If the prior discrete probability distribution of an unknown quantity X is $\Pr_0(x_i)$ for $i = 1, \ldots, m$, then the posterior probability distribution, given evidence E, is obtained from

$$\Pr(x_i|E) = \frac{\Pr(x_i|E)\Pr_0(x_i)}{\Pr(E)} \quad \text{where } i = 1, \ldots, m$$

and $\Pr(E|x_i)$ is the likelihood of the evidence when the random variable takes the value x_i. The quantity

$$\Pr(E) = \sum_{i=1}^{m} \Pr(E|x_i)\Pr_0(x_i)$$

is the total probability of E *based on* the prior distribution of X.

EXAMPLE. Transistors used by a company are supplied by three suppliers M_1, M_2, and M_3. The quantities supplied by each supplier and the corresponding fraction of defective transistors are as follows:

M_1 produces 20% of the supply with a defective rate of 0.01,
M_2 produces 30% of the supply with a defective rate of 0.02, and
M_3 produces 50% of the supply with a defective rate of 0.05,

If a transistor is randomly selected from the supply and is defective, what is the probability that it has been supplied by M_i?

To answer this question we use Bayes's theorem as follows:

$$\Pr(M_i|D) = \frac{\Pr(D|M_i)\Pr(M_i)}{\Pr(D)} \quad i = 1, 2, 3$$

where $\Pr(D) = \sum_{i=1}^{3} \Pr(D|M_i)\Pr(M_i)$, where $\Pr(M_i)$ is the prior probability that a randomly selected transistor is supplied by supplier i. Based on the information provided, $\Pr(M_1) = 0.2$, $\Pr(M_2) = 0.3$, and $\Pr(M_3) = 0.5$.

$\Pr(D|M_i)$ is the probability that a randomly selected transistor is defective when it is known that it is supplied by M_i.

The data provided give $\Pr(D|M_1) = 0.01$, $\Pr(D|M_2) = 0.02$, and $\Pr(D|M_3) = 0.05$. Therefore,

$$\begin{aligned}\Pr(D) &= \Pr(D|M_1)\Pr(M_1) + \Pr(D|M_2)\Pr(M_2) \\ &\quad + \Pr(D|M_3)\Pr(M_3) \\ &= (0.01)(0.2) + (0.02)(0.3) + (0.05)(0.5) = 0.033\end{aligned}$$

Thus according to Bayes's theorem,

$$\Pr(M_1|D) = \frac{(0.01)(0.2)}{(0.033)} = 0.061$$

Similarly

$$\Pr(M_2|D) = \frac{(0.02)(0.3)}{(0.033)} = 0.182$$

and

$$\Pr(M_3|D) = \frac{(0.05)(0.5)}{(0.033)} = 0.757$$

Note that $\Pr(M_1|D) + \Pr(M_2|D) + \Pr(M_3|D) = 1$, as expected.

Updating a Continuous Probability Distribution. The prior probability distribution of a continuous unknown quantity $\Pr_0(x)$ can be updated to incorporate new evidence E as follows:

$$\Pr(x|E) = \frac{1}{k} L(E|x)\Pr_0(x)$$

where $\Pr(x|E)$ is the posterior or updated probability distribution of the unknown quantity X, given evidence E, $L(E|x)$ is the probability of the evidence assuming the value of the unknown quantity is x, and

$$k = \int L(E|x)\Pr_0(x)\,dx$$

EXAMPLE. A reliability engineer's initial assessment of the range of possible values of the failure rate of a component λ is summarized in form of the following probability distribution:

$$\Pr_0(\lambda) = (2000)^2 \lambda e^{-2000\lambda} \quad 0 \le \lambda \le \infty$$

The component is operated for 10,000 h during which it fails once:

$$E = \{1 \text{ failure in } 10{,}000 \text{ h}\}$$

To see how this new data changes the reliability engineer's estimate of λ, we use Bayes's theorem as follows:

$$\Pr(\lambda|E) = \frac{L(\lambda|E)\Pr_0(\lambda)}{L(E)}$$

A logical model for the likelihood function $L(E|\lambda)$ in this case is the Poisson distribution:

$$L(1 \text{ failure in } 10{,}000 \text{ h}|\lambda) = (10,000)\lambda e^{-(10,000)\lambda}$$

Using this likelihood function and the prior distribution, the quantity $L(E)$ in the denominator of Bayes's theorem is calculated as follows:

$$\begin{aligned} L(E) &= \int_0^\infty L(E|\lambda)\Pr_0(\lambda)\,d\lambda \\ &= \int_0^\infty (4 \times 10^{10})\lambda^2 e^{-12,000\lambda} = 0.0463 \end{aligned}$$

The resulting posterior distribution using Bayes's theorem is given by

$$\Pr(\lambda|1 \text{ failure in } 10{,}000 \text{ h}) = \frac{(12{,}000)^3\lambda^2}{2} e^{-12,000\lambda}$$

The mean value of this distribution is $\bar{\lambda} = (3/12{,}000) = 0.00025$ failure/h, compared with the mean value of the prior distribution at $\bar{\lambda} = (2/2{,}000) = 0.001$ failure/h.

APPLYING BAYESIAN THINKING IN PROBABILISTIC INFERENCE

A number of useful concepts and implications of the Bayesian way of addressing probabilistic situations are demonstrated by the following example.

Suppose that one is interested in assessing the probability of heads H in a specific trial of a coin flipping experiment when the result of a recent experiment with the coin in question shows 5 heads in 20 trials. The answer to this question can be developed step-by-step and through answers to several more fundamental questions:

What is the Quantity of Interest? The quantity of interest is the probability of heads in a specific case of flipping a specific coin. There are two possible outcomes: $X = \{H, T\}$. The corresponding probabilities are $\Pr(H)$ and $\Pr(T)$. Let us assume that $\Pr(H) = p$. Clearly, to be coherent, we must have $\Pr(T) = 1 - p$. The quantity p is our degree of belief about the event $X = H$.

Depending on the prior information, our assessment of the value of p varies. Note that $0 \le p \le 1$. Our state of knowledge, or degree of belief, or uncertainty, is represented by our probability distribution $\pi(p)$ on possible values of p. Some possible situations are illustrated in the following:

1. If we believe that the coin is fair, then $\Pr(H) = \Pr(T) = 0.5$ with 1 (or 100%) as the level of confidence in $p = .5$ [see Fig. 1(a)].
2. If someone tells us that the coin is very likely (but not certain) to be fair, then our prior distribution of p may look like the curve shown in Fig. 1(b).
3. If on the other hand, we are told that the coin is biased (same face on both sides) without being told about the direction of bias, our prior distribution would look more like Fig. 1(c). In other words $\Pr(H) = 1$, if both sides are H. When both sides are T, then $\Pr(H) = 0$. We note that the expected value of p in this case is $\bar{p} = .05(0) + .05(1) = .05$.
4. If we are told that the coin is most likely biased (meaning that there is some likelihood that it is not biased), then a possible form of the prior distribution of p is what is shown in Fig. 1(d).
5. If we do not know anything about the coin, we probably assign a flat (equal likelihood, or noninformative) prior distribution to express our opinion about likely values of p, as shown in Fig. 1(e).
6. Finally, with *perfect information,* we would assign a δ-function centered about $p = p_0$ [see Fig. 1(f)]. As we will see later, this should be the long-term frequency of heads in a very large number of trials.

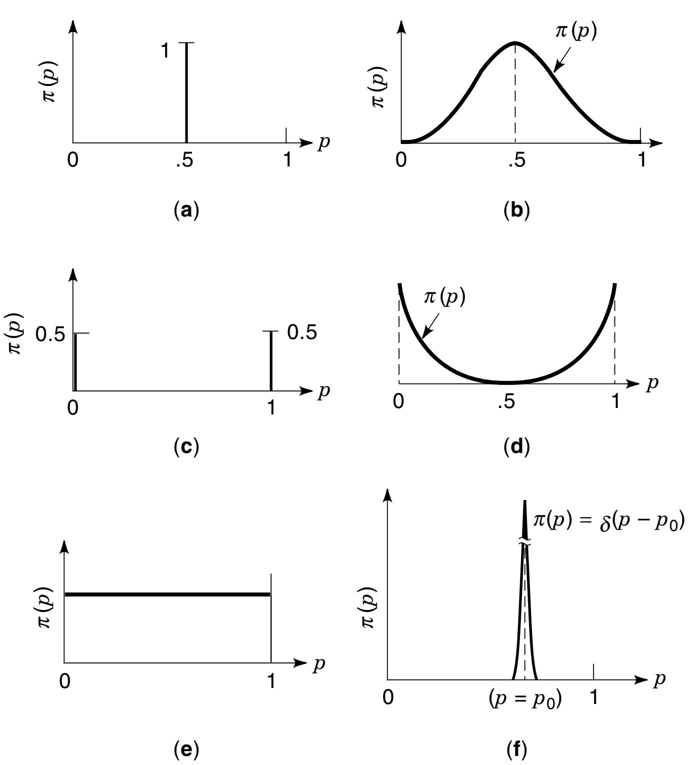

Figure 1. Different prior distributions repressing different states of knowledge.

In all these cases [(1) through (6)] the mean value of p is the same (.5), but the difference is in the uncertainty distribution representing different states of knowledge.

What is the Evidence? Other than our prior information, we are using the fact that in 20 trials of the coin we have observed five heads. The evidence is, therefore, E = {5 heads in 20 trials}. To state the problem more generally, let us assume that the evidence is $E = \{N_H \text{ in } N \text{ trials}\}$ where $N = N_H + N_T$, where N_H and N_T are the number of cases where the outcome is heads or tails, respectively.

How Can the Evidence Be Used to Update Our State of Knowledge? To answer this, we use Bayes's theorem:

$$\pi(p|E) = \frac{L(E|p)\pi_0(p)}{\int L(E|p)\pi_0(p)\,dp}$$

where $\pi(p|E)$ is the posterior updated distribution of p, given evidence E, $L(E|p)$ is the likelihood function or probability of observing evidence E, given p, and $\pi_0(p)$ is the prior probability distribution of p.

We note that according to classical statistics, $p = \lim_{N\to\infty} (N_H/N)$, and a "point estimate" for p (for limited values of N) is given by $\hat{p} = (N_H/N)$.

What Is the Form of the Likelihood? The likelihood $L(E|p)$ is the probability of observing N_H heads in N trials, if p is the probability of observing H in any given trial. The total probability of all sequences of outcomes involving N_H heads and N_T tails is given by the binomial distribution (1):

$$L(E|p) = \Pr(N_H|N,p) = \begin{bmatrix} N \\ N_H \end{bmatrix} p^{N_H}(1-p)^{N-N_H}$$

What Is Our New State of Knowledge Regarding *p*, Given the Evidence? Let us start with the assumption that before the experiment we had no information about the characteristics of the coin, that is, our prior distribution of p is flat, i.e., $\pi_0(p) = C$ = constant. Since the prior distribution is constant,

$$\pi(p|E) = \frac{L(E|p)(C)}{\int L(E|p)(C)\,dp} = \frac{\Pr(N_H|N,p)}{\int \Pr(N_H|N,p)\,dp}$$

The resulting posterior distribution is given by

$$\pi(p|N_H,N) = \frac{(N+1)!}{(N_H+1)!(N-N_H)!}\,p^{N_H}(1-p)^{N-N_H}$$

Some Observations about the Results.

1. The form of the posterior density is different depending on the values of N_H and N.
2. The *mode* or *most likely value* of p is determined by maximizing $\pi(p)$ with respect to p, resulting in $\hat{p} = N_H/N$.
3. The *average* or *expected value* is the value which the assessor expects to see, considering all uncertainties in an aggregated way, and is given by

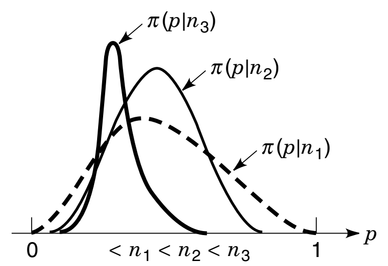

Figure 2. The effect of data strength (sample size) on the spread of posterior distribution.

$$\overline{p} = \int_0^1 p\pi(p|E)\,dp$$

4. Variance of the posterior distribution is given by

$$\text{var}(p) = \frac{(N_H+1)(N-N_H+1)}{(N+2)^2(N+3)}$$

   As we can see, as $N \to \infty$, var(p) $\to$ 0. This is shown graphically in Fig. 2 for three different values of N.

5. By keeping N_H/N constant and increasing N, we can see that

$$\lim_{N\to\infty} \pi(p|E) \to \delta(p - \phi_\infty)$$

   where

$$\phi_\infty = \lim_{N\to\infty} \frac{N_H}{N}$$

Therefore, by increasing the number of trials N, uncertainty is decreased. Because the spread of distribution reflects our uncertainty about the value of the unknown quantity, a delta function represents "perfect knowledge." In this case the results tell us that our estimate of p should be ϕ_∞, that is, in the limit of overwhelmingly strong evidence, the limiting frequency ϕ_∞, is the probability of the event H assessed by a coherent assessor of probabilities. This establishes the link between the frequentist (classical) and subjectivist (Bayesian) interpretations of probability.

CONJUGATE PRIORS

There are a number parametric families of distributions which, when used as priors and in conjunction with a particular type of likelihood function in Bayes's theorem, result in a posterior distribution from the same family. For such distributions the parameters of the resulting posterior are simple functions of the parameters of the prior distribution and the likelihood function. Consequently the computations in Bayes's theorem are simplified. These distributions are called *conju-*

Table 1. Conjugate Distributions

Prior Distribution	Likelihood Function	Posterior Distribution
Beta	Binomial	Beta
Gamma	Poisson	Gamma
Normal	Normal	Normal
Lognormal	Lognormal	Lognormal

gate. Most commonly used conjugate distributions are listed in Table 1.

The conjugate properties of some of these distributions are used in the following sections.

EXAMPLES OF APPLICATION OF BAYESIAN METHODS IN RELIABILITY

Treatment of Data from Homogeneous Populations

By a homogeneous population we mean identical with respect to the characteristic of interest. For instance, if all members of a population of components have the same failure rate, the population is regarded as homogeneous in failure rate. One can perform a sequential updating in which data from one subset of the population are used to update the state of knowledge obtained based on data from another subset of the population. This process continues until data from all subpopulations are incorporated into the state-of-knowledge distribution on the quantity of interest. By contrast, the data from all subpopulations can be added together and used in one application of Bayes's theorem to obtain the posterior state of knowledge representing the cumulative information. The two results are identical provided that the data represent the same characteristics of subpopulations of a homogeneous population.

EXAMPLE. Consider the case where 50 identical engines are tested by five independent laboratories. The test results are summarized in Table 2.

Assume that all components tested have the same failure rate. To estimate this rate, we start with a prior distribution that expresses our initial state of knowledge based on information other than the test data. Let this prior be a beta distribution with parameters $a_0 = 2$ and $b_0 = 150$:

$$\text{Prior: } B(p|a_0 = 2, b_0 = 150)$$

where p is the probability of failures per test. We note that the mean of p is given as

Table 2. Summary of Engine Tests

Lab.	Engines tested	Engine Starts (N_i)	Number of Failures (k_i)
1	10	1000	2
2	2	250	1
3	20	3000	1
4	10	600	2
5	8	1000	1

Table 3. Some Characteristics of Prior and Posterior Distributions for the Homogeneous Data Example

	5th Percentile	Mean	95th Percentile
Prior	0.00235	0.0132	0.0311
Posterior	0.000785	0.00150	0.00240

$$\overline{p}_0 = \frac{a_0}{a_0 + b_0} = \frac{2}{150 + 2} = 0.0132 \text{ failure/start}$$

Because the components are assumed to have the same failure rate (homogeneous population), the data from various tests can be pooled. Therefore the evidence is

$$k = \text{Total number of failures} = \sum_{i=1}^{5} k_i = 7$$

$$N = \text{Total number of tests} = \sum_{i=1}^{5} N_i = 5850$$

Now this data can be used in Bayes's theorem to obtain the posterior state-of-knowledge distribution of p. Using the binomial distribution as the likelihood function, we can benefit from the simplicity of a conjugate beta prior and obtain a beta posterior distribution $B(p|a, b)$ with parameters

$$a = a_0 + k = 2 + 7 = 9$$

and

$$b = b_0 + N - k = 150 + 5850 - 7 = 5993$$

The posterior mean is given by

$$\overline{p} = \frac{a}{a + b} = \frac{9}{9 + 5993} = 0.0015 \text{ failures/engine start}$$

The main characteristics of both the prior and posterior probability distributions are shown in Table 3.

Note that by adding more data from subpopulations or by increasing the number of subpopulations, the posterior distribution becomes narrower and centered about the point

$$p_t = \frac{k}{N}$$

That is, by assuming that all data are relevant and that there is one underlying parameter p_t (the true, but unknown probability of failure), which is the same for each of the components in the population, the spread of the posterior distribution decreases as the amount of data (k_i's and N_i's) increase. In the limit, the true value of p (i.e., p_t) becomes known because the posterior distribution becomes a δ-function about p_t. Again this is true only for homogeneous populations.

Treatment of Data from Nonhomogeneous Populations

Basic Definitions and Concepts. Often, because of several reasons such as environmental factors, design differences, and operational variations, the components or systems in dif-

Table 4. Assumed True Failure Probabilities for the Nonhomogeneous Data Example

Subpopulation (i)	Fraction of Engines in Subpopulation (%)	True Probability of Failure to Start (p_i)
1	20	0.001 /start
2	4	0.008 /start
3	40	0.0003/start
4	20	0.002 /start
5	16	0.0006/start

ferent subpopulations exhibit different reliability characteristics, that is, they may have different failure modes, failure rates, and repair times.

In such cases, it is not realistic to assume that all members of a population composed of different subpopulations have the same reliability parameters. In other words, when the population at hand is nonhomogeneous, a given reliability parameter, for example, failure rate, will have an inherent variability, that is, the failure rate is inherently different for one subpopulation compared with others. This is called *population variability* of the failure rate (or any other reliability parameters of interest).

In the previous example if we assume the existence of a population variability among the five subpopulations tested by different laboratories, then instead of one failure probability we would need to estimate five different failure probabilities, each corresponding to one subpopulation. We note that no amount of information eliminates this variability. It changes only if the actual reliability characteristics of the engines change.

To help better understand the meaning and implications of population variability, let us assume that the failure probabilities of each of the five subpopulations in this example is perfectly known with no uncertainty. Let these probabilities be values shown in Table 4.

In other words if we select an engine at random from this population, its failure rate is one of the tabular values. The probability that its failure rate is p_i is, of course, the fraction of engines that have that failure rate, which is the fraction of all engines in the ith subpopulation.

Consequently, for the failure rate of a randomly selected engine from this nonhomogeneous population, we have the probability distribution shown in Fig. 3.

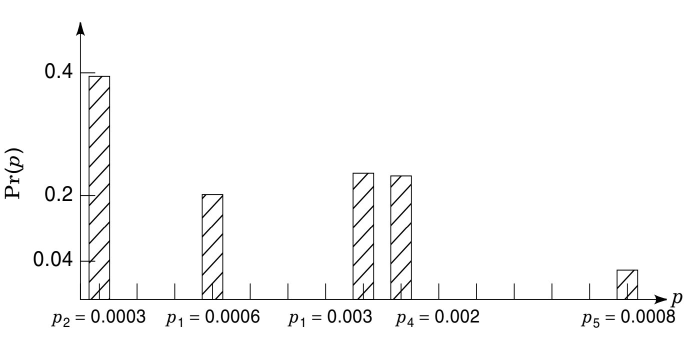

Figure 3. Example. Population variability distribution of failure probability.

Again the location of p_i's and the magnitude of the corresponding probabilities are dictated by the reliability characteristics and the size of the subpopulations. As such, this probability distribution is real and in principle measurable, that is, by knowing the exact values of p_i's and the fraction of components in each subpopulation, everything will be known about this distribution, which we call the *population variability distribution.*

Given the previous population variability distribution, the expected value of the failure probability of an engine selected at random from these engines is given by

$$\overline{p} = \sum_{i=1}^{5} p_i P(p_i) = 0.00114 \text{ failure/start}$$

The most probable value of p is 0.0003 because the corresponding probability, .4, is the highest among the five probabilities.

Now consider the case where only estimates, and not the exact values, of failures frequencies are available for some but not all of the subpopulations involved. For instance the database in the engine example here gives us limited number of failures in limited number of tests from each of the laboratories. Consequently we can obtain only an estimate of the subpopulation failure probabilities. For instance, in the case of subpopulation 1, this estimate based on the maximum likelihood method is

$$\hat{p}_1 = \frac{k_1}{N_1} = \frac{2}{1000} = 0.002/\text{start}$$

which is different from the true value (assumed to be 0.001) for p_1. With this limited state of knowledge, obviously we cannot know the exact form of the population variability distribution. The question is how this more limited information can be used to estimate the population variability distribution. We demonstrate the method through an example.

Application to Failure Rate Estimation. Consider the case where the following data are available about the performance of a particular type of component $E = \{(k_i, T_i), i = 1, \ldots, N\}$ where k_i = number of failures in subpopulation i, T_i = total number of hours of operation in subpopulation i, and N = number of subpopulations.

Each (k_i, T_i) pair represents the experience of a subpopulation. It is important to note that the subpopulations are not necessarily different. The objective is to find $\phi(\lambda)$, the population variability distribution of λ, the failure rate.

To simplify matters we assume that $\phi(\lambda)$ is a member of a parametric family of distributions, such as beta, gamma, or lognormal. Let $\underline{\theta} = \{\theta_1, \theta_2, \ldots, \theta_m\}$ be the set of m parameters of $\phi(\lambda)$, that is, $\phi(\lambda) = \phi(\lambda|\underline{\theta})$.

For example, for normal distribution, $\underline{\theta} = \{\mu, \sigma\}$ and

$$\phi(\lambda) = \phi(\lambda|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\lambda-\mu}{\sigma}\right)^2}$$

Uncertainty distribution (state of knowledge) over the space of ϕ's is the same as uncertainty distribution over values of $\underline{\theta}$. For each value of $\underline{\theta}$ a unique $\phi(\lambda|\underline{\theta})$ exists and vice versa. Now our goal of estimating $\phi(\lambda)$ is reduced to that of estimating $\underline{\theta}$. Given the information available, E, and a prior distribution on $\underline{\theta}$, we can use Bayes's theorem to find our state-of-knowledge probability distribution over $\underline{\theta}$:

$$\pi(\underline{\theta}|E) = \frac{L(E|\underline{\theta})\pi_0(\theta)}{\int_{\theta_1}\int_{\theta_2}\cdots\int_{\theta_\mathrm{m}} L(E|\underline{\theta})\pi_0(\underline{\theta})\,d\underline{\theta}}$$

where $\pi_0(\underline{\theta})$ = prior distribution of $\underline{\theta}$, and $\pi(\underline{\theta}|E)$ = posterior distribution of $\underline{\theta}$, given evidence E. We note that $\pi(\underline{\theta}|E)$ is an n-dimensional joint probability distribution over values of θ_1, $\theta_2, \ldots, \theta_n$, and $L(E|\underline{\theta})$ = likelihood of evidence E, given $\underline{\theta}$.

Once this distribution is determined, we can obtain various estimates for $\phi(\lambda)$. For instance, the *expected (average) distribution* is given by

$$\overline{\phi}(\lambda) = \int_{\theta_1}\cdots\int_{\theta_\mathrm{m}} (\lambda|\theta_1, \theta_2, \ldots, \theta_m)\pi(\lambda|\theta_1, \theta_2, \ldots, \theta_m|E)\,d\underline{\theta}$$

The expected value of $\underline{\theta}$ is obtained from

$$E(\underline{\theta}) = \int_{\theta_1}\cdots\int_{\theta_\mathrm{m}} \underline{\theta}\pi(\theta|E)\,d\underline{\theta}$$

Using the expected value of $\underline{\theta}$ as the set of parameters of ϕ gives us another "point estimate" of ϕ, that is, $\phi[\lambda|E(\underline{\theta})]$, is the distribution with mean value parameters. We note that $\overline{\phi}(\lambda) \neq \phi[\lambda|E(\underline{\theta})]$.

Similarly the *most likely distribution* is obtained by finding the values of $\theta_1, \theta_2, \ldots, \theta_m$ such that $\pi(\theta_1, \theta_2, \ldots, \theta_m|E)$ is maximized:

$$\left.\frac{\partial\pi(\theta_1, \theta_2, \theta_3, \ldots, \theta_m|E)}{\partial\theta_i}\right|_{\theta_i=\hat{\theta}_i} = 0 \qquad \text{where } i = 1,2,\ldots,m$$

Then the *"most likely"* distribution is that member of the parametric family $\phi(\lambda|\theta_1, \theta_2, \ldots, \theta_m)$ for which $\theta_1 = \hat{\theta}_1$, $\theta_2 = \hat{\theta}_2, \ldots, \theta_m = \hat{\theta}_m$.

Note that the *likelihood function* $L(E|\underline{\theta})$ is the probability of observing the data E, given that the set of the parameters of the population variability distribution is $\underline{\theta}$. Now we assume that the data from the individual subpopulations are independent, that is, the process of generating the data in one subpopulation is not influenced by the process in another subpopulation. Therefore, the elements of $E = \{E_1, E_2, \ldots, E_\mathrm{N}\}$, defined as $E_i = \{k_i, T_i\}$ are independent events. In this case the likelihood function can be written as the product of subpopulation likelihoods:

$$L(E|\underline{\theta}) = \prod_{i=1}^{N} L(E_i|\underline{\theta})$$

$$L[(k_i|T_i), \quad i = 1, \ldots, N|\underline{\theta}] = \prod_{i=1}^{N} L(k_i, T_i|\underline{\theta})$$

If the failure rate λ_i for the ith subpopulation is known exactly, say $\lambda_i = \lambda$, then using the Poisson model, the likelihood of observing k_i failures in T_i units of time can be calculated from

$$P(k_i = T_i|\lambda) = \frac{(\lambda T_i)^{k_i}}{k_i!}e^{-\lambda T_i}$$

However, λ is not known. All we know is that λ is one of many possible values represented by $\phi(\lambda|\underline{\theta})$. Therefore we can find the probability of the data unconditional on the (unknown) value of λ. This is done by averaging the likelihood over all possible values of λ:

$$L(k_i, T|\underline{\theta}) = \int_0^\infty L(k_i, T_i|\lambda)\phi(\lambda|\theta_i)\,d\lambda$$

or

$$L(k_i, T_i|\underline{\theta}) = \int_0^\infty \frac{(\lambda T_i)^{k_i}}{k_i!}e^{-\lambda T_i}\phi(\lambda|\theta_i)\,d\lambda$$

Note that now the likelihood becomes conditional on $\underline{\theta}$. This is in fact the desired form of the likelihood for use in Bayes's theorem for estimating $\underline{\theta}$.

Depending on the parametric family chosen to represent $\phi(\lambda|\underline{\theta})$, the integration in the previous equation can be carried out either analytically or by numerical methods.

For example, if $\phi(\lambda|\underline{\theta})$ is a gamma distribution with parameters $\underline{\theta} = \{\alpha, \beta\}$, that is, if

$$\phi(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\beta\lambda}$$

then

$$\begin{aligned} L(k_i, T_i|\underline{\theta}) &= L(k_i, T_i|\underline{\alpha, \beta}) \\ &= \int_0^\infty \frac{(\lambda T_i)^{k_i}}{k_i!}e^{\lambda T_i}\frac{\lambda^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta\lambda}\beta^\alpha\,d\lambda \\ &= \frac{T_i^{k_i}}{k_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+k_i)}{(\beta+T_i)^{\alpha+k_i}} \end{aligned}$$

The joint or total likelihood is given by

$$\begin{aligned} L\{(k_i, T_i), i = 1, \ldots, N|\alpha, \beta\} &= \prod L(k_i, T_i|\alpha, \beta) \\ &= \frac{\beta^\alpha}{[\Gamma(\alpha)]^N}\prod_{i=1}^{N}\frac{T_i^{k_i}}{k_i!}\frac{\Gamma(\alpha+k_i)}{(\beta+T_i)^{\alpha+k_i}} \end{aligned}$$

Table 5. Simulated Failure Data Using a Gamma Distribution as the Population of Underlying Failure Rates

Sample Failure Data			
Time	Failures	Time	Failure
1000	130	10	1
1000	311	1000	107
100	22	100	27
10	2	100	41
100	13	1000	163
1000	110	10000	1653
100	22	100	13
100	24	100	37
10	2	1000	170
10	1	100	14

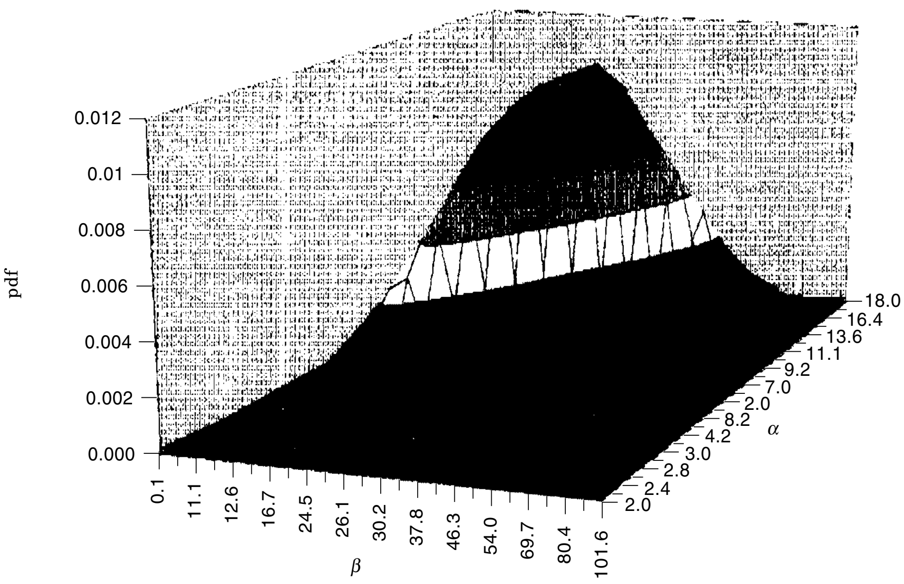

Figure 4. Posterior joint probability density function for the unknown gamma distribution parameters.

Now we can use this likelihood in Bayes's theorem to find the joint distribution of α and β:

$$\pi(\alpha\beta|\{k_i, T_i\}, i = 1, \ldots, N) = \frac{L(\{k_i, T_i\},\ i = 1, \ldots, N|\alpha, \beta)\pi_0(\alpha, \beta)}{\int_\alpha \int_\beta L(\{k_i, T_i\},\ i = 1, \ldots, N|\alpha, \beta)\pi_0(\alpha, \beta)\, d\beta\, d\alpha}$$

which is a two-dimensional probability distribution over α and β.

The *expected distribution* is obtained as follows:

$$\overline{\pi}(\lambda|E) = \int_\alpha \int_\beta \pi(\alpha,\ \beta|\left\{k_i,\ T_i\right\},\ i = 1,\ \cdots,\ N)\phi(\lambda|\alpha,\ \beta)\, d\alpha\, d\beta$$

Table 5 lists the data used to illustrate the technique. The data were generated by Monte Carlo sampling from a gamma distribution as the population variability. Figure 4 shows plots of the posterior distribution of parameters, and Fig. 5 shows the resulting expected posterior population variability pdf compared with the theoretical (correct) population variability distribution (i.e., the original gamma distribution).

It is evident that computational aspects of using Bayes's theorem in many practical situations are quite involved. In fact computational complexity in the past was one of the main reasons for relatively limited use of Bayesian methods in reliability. However, recent advances in computer technology have removed some of the practical barriers, and the past few years have witnessed a significant rise in interest in Bayesian methods among design and reliability engineers.

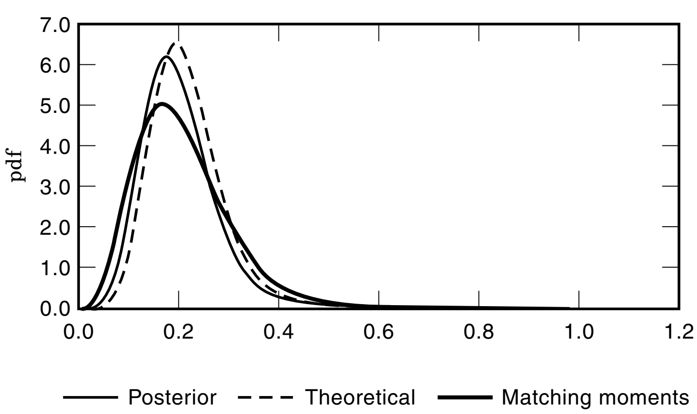

Figure 5. Population variability pdf for example application.

BIBLIOGRAPHY

1. H. F. Martz and R. A. Waller, *Bayesian Reliability Analysis,* New York: Wiley, 1982.
2. S. French and J. Q. Smith (eds.), *The Practice of Bayesian Analysis,* London: Arnold, 1997.
3. J. M. Bernardo and A. F. M. Smith, *Bayesian Theory,* New York: Wiley, 1994.
4. T. Bayes, Essay Towards Solving a Problem in the Doctrine of Change, *Biometrika,* **45**: 298–315, 1964.

ALI MOSLEH
University of Maryland