B. BASICS OF STATISTICS

B.1 Random Samples

When sampling from a distribution (or population), it is usually assumed that the n observations are taken at random, in the following sense. It is assumed that the n random variables $X_1, X_2, ..., X_n$ are independent. That is, the sample $X_1, X_2, ..., X_n$ taken from a distribution $f(x)$ has the joint p.d.f. h satisfying

$$h(x_1, x_2, ..., x_n) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n).$$

This follows the definition of independent random variables given in Section A.4.5. A sample taken in this way is called a **random sample**. (As elsewhere in this handbook, upper case letters denote random variables and lower case letters denote particular values, number.)

The random variables $X_1, X_2, ..., X_n$ forming such a random sample are referred to as being independent and identically distributed. If n is large enough, the sampled values will represent the distribution well enough to permit inference about the true distribution.

B.2 Sample Moments

Mathematical expectation and moments provide characteristics of distributions of random variables. These ideas can also be used with observations from a random sample from a distribution to provide **estimates** of the parameters that characterize that distribution.

A **statistic** is a function of one or more random variables that does not depend on any unknown parameters. A function of random variables that can be computed from the collected data sample is thus a statistic. Note that a function of random variables is also a random variable that has its own probability distribution and associated characteristics.

If $X_1, X_2, ..., X_n$ denote a random sample of size n from a distribution $f(x)$, the statistic

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

is the **mean of the random sample**, or the **sample mean** and the statistic

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

is the **variance of the random sample**. Note that $n - 1$ is used as the denominator in the S^2 statistic to make the statistic an *unbiased* estimator of the population variance, σ^2 (unbiased estimators are discussed in Section B.4.1). Although not used as much as the sample mean and sample variance, the **sample skewness** is occasionally of interest. The definition can vary in detail, but one, used by SAS (1988) is

$$\frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} (X_i - \overline{X})^3 \Big/ S^3 .$$

Similarly, the statistics defined by

$$m_r = \sum_{i=1}^{n} \frac{X_i^r}{n},$$

for $r = 1, 2, ...$, are called the **sample moments**.

One of the common uses of statistics is estimating the unknown parameters of the distribution from which the sample was generated. The sample mean, or average, $\overline{X}$, is used to estimate the distribution mean, or population mean, μ, the sample variance, S^2, is used to estimate the population variance, σ^2, and so forth.

B.3 Statistical Inference

Since values of the parameters of a distribution are rarely known, the distribution of a random variable is rarely completely known. However, with some assumptions and information based on a random sample of observations from the distribution or population, values of the unknown parameters can often be estimated. Probabilities can then be calculated from the corresponding distribution using these parameter estimates.

Statistical inference is the area of statistics concerned with using sample data to answer questions and make

statements about the distribution of a random variable from which the sample data were obtained. Parameter **estimators** are functions of sample data that are used to estimate the distribution parameters. Statements about parameter values are inferred from the specific sample to the general distribution of the random variable or population. This inference cannot be perfect; all inference techniques involve uncertainty. Understanding the performance properties of various estimators has received much attention in the statistics field.

For the purposes of this handbook, statistical inference procedures can be classified as follows:

- parameter estimation
  - estimation by a point value
  - estimation by an interval
- hypothesis testing
  - tests concerning parameter values
  - goodness-of-fit tests and other model-validation tests.

Parametric statistical inference assumes that the sample data come from a particular, specified family of distributions, with only the parameter values unknown. However, not all statistical inference is based on parametric families. In many cases, in addition to not knowing the distribution parameter values, the form of the parametric family of distributions is unknown. **Distribution-free**, also called **nonparametric**, techniques are applicable no matter what form the distribution may have. **Goodness-of-fit tests** are an important type of nonparametric tests that can be used to test whether a data set follows a hypothesized distribution.

For statistical inference, two major approaches exist, the **frequentist** approach and the **Bayesian** approach. The two resulting sets of inference tools are summarized in Sections B.4 and B.5. In PRA work, Bayesian estimators are normally used for parameter estimation. See, for example, NUREG-1489 (NRC 1994). However, frequentist hypothesis tests are often used for model validation, especially when the hypothesis to be tested does not involve a simple parameter. This use of Bayesian techniques for estimation and frequentist techniques for model validation is also recommended by Box (1980).

NUREG-1489 (NRC 1994) lists a number of "advantages" and "disadvantages" for each of the Bayesian and frequentist approaches. "Advantage" is often in the eye of the beholder. For example, is it an advantage or disadvantage that frequentist methods use only the data at hand, not external or prior information? Therefore, the lists from that report are presented in modified and augmented form in Table B.1, where the points are not called advantages or disadvantages, but simply "features," which the reader may attach values to.

B.4 Frequentist Inference

Frequentist estimation of distribution parameters uses only the information contained in the data sample and assumptions about a model for the sample data. In contrast to **Bayesian estimation** (discussed in Section B.5), degree of belief is not incorporated into the estimation process of frequentist estimation.

In the **frequentist** approach to estimation, a distribution parameter is treated as an unknown constant and the data to be used for estimation are assumed to have resulted from a random sample. Information outside that contained in the sample data is used minimally. The random variability in the sample data is assumed to be due directly to the process under study. Thus, the frequentist approach addresses variation in parameter estimates and how far estimates are from the true parameter values.

Frequentist testing of a hypothesis follows the same spirit. The hypothesis is assumed, and the data are compared to what would have been expected or predicted by the hypothesis. The frequentist analyst asks whether the observed values come from the likely part of the distribution or from the extreme tails, and decides in this way whether the data are consistent with the hypothesis.

B.4.1 Point Estimation

Many situations arise in statistics where a random variable X has a p.d.f. that is of known functional form but depends on an unknown parameter θ that can take on any value in a set. The different values for θ produce a family of distributions. One member of the family corresponds to each possible value of θ. **Estimators** of the distribution parameter are functions

of sample data that are used to estimate the distribution parameters. Thus, estimators are themselves random variables. The specific value of the estimator computed from a random sample provides an **estimate** of the distribution parameter.

Table B.1 Features of Bayesian and Frequentist Approaches

Bayesian Approach	Frequentist Approach
Bayesian methods allow the formal introduction of prior information and knowledge into the analysis, which can be especially useful when sample data are scarce, such as for rare events. For the nuclear industry, this knowledge often exists in the form of **industry-wide generic data**. Thus, Bayesian estimation allows the use of various types of relevant generic data in PRA.	Results depend only on the data sample. Including relevant information about a parameter that is external to the random sample is complicated.
If the prior distribution accurately reflects the uncertainty about a parameter, Bayesian parameter estimates are better than classical estimates.	
Bayesian estimation can be sensitive to the choice of a prior distribution. Therefore: Identifying suitable prior distributions and justifying and gaining acceptance for their use can be difficult. The choice of a prior distribution is open to criticism that the choice is self-serving and may reflect inappropriate, biased, or incorrect views.	
Because Bayesian probability intervals can be interpreted as probability statements about a parameter, they are easily combined with other sources of uncertainty in a PRA using the laws of probability.	A confidence interval cannot be directly interpreted as a probability that the parameter lies in the interval.
Bayesian distributions can be propagated through fault trees, event trees, and other logic models.	It is difficult or impossible to propagate frequentist confidence intervals through fault and event tree models common in PRA to produce corresponding interval estimates on output quantities of interest.
Using Bayes' Theorem, Bayesian estimation provides a method to update the state of knowledge about a parameter as additional data become available.	Frequentist methods can update an earlier analysis if the original data are still available or can be reconstructed.
In complicated settings, Bayesian methods require software to produce samples from the distributions.	In complicated settings, frequentist methods must use approximations. In some cases they may be unable to analyze the data at all.
Bayesian hypothesis tests are commonly used only with hypotheses about a parameter value.	A well-developed body of hypothesis tests exists, useful for model validation. These are appropriate for investigating goodness of fit, poolability of data sources, and similar questions that do not involve a simple parameter.

Both Approaches
When the quantity of data is large, both approaches produce good estimates.
Both types of computation are straightforward when estimating a parameter in a simple setting.

An estimate of a distribution parameter in the form of a single number is called a **point estimate** of that parameter. The sample mean is a point estimate of the mean of the distribution and the sample variance is a point estimate of the variance of the distribution. For another sample drawn from the same population, a different sample mean and variance would be calculated. In fact, these sample statistics are specific values of random variables and, thus, have their own **sampling distributions**. For example, it can be shown that $\overline{X}$ has mean μ and variance σ^2/n, regardless of the distribution from which the samples are drawn.

Different techniques exist for obtaining point estimates for unknown distribution characteristics or parameters. Two of the most common methods are presented here [see Hogg and Craig (1995) for more information]: maximum likelihood estimation and the method of moments.

A distribution of a random variable X that depends on an unknown parameter θ will be denoted $f(x; \theta)$. If X_1, X_2, ..., X_n is a random sample from $f(x; \theta)$, the joint p.d.f. of X_1, X_2, ..., X_n is $f(x_1; \theta) \cdot f(x_2; \theta) \cdot \ldots \cdot f(x_n; \theta)$. This joint p.d.f. may be viewed as a function of the unknown parameter θ and, when so viewed, is called the **likelihood function**, L, of the random sample. Thus, the likelihood function is the joint p.d.f. of X_1, X_2, ..., X_n, denoted

$$L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta) ,$$

viewed as a function of θ. The **maximum likelihood estimate** of θ is defined as the value $\hat{\theta}$ such that $L(\hat{\theta}; x_1, x_2, ..., x_n) \geq L(\theta; x_1, x_2, ..., x_n)$ for every value of θ. That is, the maximum likelihood estimate of θ is the value $\hat{\theta}$ that maximizes the likelihood function. In many cases, this maximum will be unique and can often be obtained through differentiation. Note that solving the derivative set to zero for θ may be easier using $\ln(L)$, which is equivalent since a function and its natural logarithm are maximized at the same value of θ.

The maximum likelihood estimate is a function of the observed random sample $x_1, x_2, ..., x_n$. When $\hat{\theta}$ is considered to be a function of the random sample X_1, $X_2, ..., X_n$, then $\hat{\theta}$ is a random variable and is called the **maximum likelihood estimator** of θ.

Another method of point estimation is the **method of moments**, which involves setting the distribution moments equal to the sample moments:

$$M_r = E(X^r) = m_r = \sum_i x_i^r / n,$$

for $r = 1, 2, ..., k$, if the p.d.f. $f(x; \theta_1, \theta_2, ..., \theta_k)$ has k parameters. The k equations in k unknowns can be solved for the k unknowns $\theta_1, \theta_2, ..., \theta_k$ and the solutions $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are the method-of-moments estimators.

How "well" a point estimator estimates a parameter has received a large amount of attention. Numerous desirable properties of point estimators exist. One desirable property of estimators, alluded to previously in Section B.2, is **unbiasedness**. An **unbiased** estimator is one whose mean value is equal to the parameter being estimated. That is, an estimator $\hat{\theta}$ is unbiased for a parameter θ if $E(\hat{\theta}) = \theta$. For a random sample from a normal distribution, the sample mean, $\overline{X}$, and the sample variance, S^2, are unbiased estimators of μ and σ^2, respectively. However, the method of moments estimator of the variance is biased. The **bias** of an estimator $\hat{\theta}$ is defined as $E(\hat{\theta}) - \theta$.

Minimum variance is another desirable property of an estimator. An unbiased estimator is said to have minimum variance if its variance is less than or equal to the variance of every other unbiased statistic for θ. Such an estimator is referred to as an unbiased, minimum variance estimator.

Another desirable property of estimators is **sufficiency**. For a random sample $X_1, X_2, ..., X_n$ from $f(x; \theta_1, \theta_2, ..., \theta_m)$, and $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ functions (statistics) of the X_is, the statistics $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$ are **jointly sufficient statistics** if the conditional p.d.f. of the X_is given the statistics $\hat{\theta}_j$s, $g(x_1, x_2, ..., x_n | \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m)$, is independent of the parameters (Martz and Waller, 1991).

Sufficiency can be thought of as exhausting all the possible information about a parameter that is contained in the random sample. When a sufficient statistic exists, it may serve as the basis for a minimum variance or "best" estimator of the parameter. Sufficiency is also important because it simplifies Bayesian estimation methods.

Under certain commonly occurring conditions, as the sample size gets large, the maximum likelihood estimator is approximately normally distributed, approximately unbiased, and has approximately the minimum variance. It, therefore, is a very good estimator for large data sets. The maximum likelihood estimator is not necessarily good for small data sets.

Several other methods of estimation and desirable properties for estimators exist. Further information can be found in Hogg and Craig (1995) or Kendall and Stuart (1973).

B.4.2 Interval Estimation

Another way of estimating a parameter is to identify that it falls in some interval (*lcl*, *ucl*) with a specified degree of certainty, or confidence, where *lcl* denotes the lower confidence limit and *ucl* denotes the upper confidence limit. The interval (*lcl*, *ucl*) is referred to as an **interval estimate** of the parameter. The *lcl* and *ucl* values are calculated from the random sample from the given distribution. Associating a level of desired confidence with an interval estimate produces a **confidence interval**. The level of desired confidence is also referred to as the confidence coefficient.

Confidence intervals are based on estimators associated with a random sample (functions of the data), *LCL* for the lower confidence limit and *UCL* for the upper confidence limit, such that, prior to observing the random sample, the probability that the unknown parameter, θ, is contained in the interval [*LCL*, *UCL*] is known. That is,

$$\Pr[LCL \le \theta \le UCL] = 1 - \alpha$$

for $0 < \alpha < 1$.

Once the random sample has been generated, the functions *LCL* and *UCL* produce two values, *lcl* and *ucl*. The interval (*lcl*, *ucl*) is called a two-sided confidence interval with confidence level $1 - \alpha$, or equivalently, a $100(1 - \alpha)\%$ two-sided confidence interval. Similarly, upper one-sided confidence intervals or lower one-sided confidence intervals can be defined that produce only an upper or lower limit, respectively.

Since the true parameter value, although unknown, is some constant, the interval estimate either contains the true parameter value or it does not. A 95% confidence interval is interpreted to mean that, for a large number of random samples from the same distribution, 95% of the resulting intervals (one interval estimate of the same population parameter constructed the same way for each sample) would contain the true population parameter value, and 5% of the intervals would not. The $\alpha = .05$ risk of obtaining an interval that does not contain the parameter can be increased or decreased. Values for $1 - \alpha$ should be decided upon prior to obtaining the random sample, with .99, .95, and .90 being typical. Note that higher confidence levels result in wider interval estimates.

Confidence intervals cannot be interpreted as probability statements about the parameter being estimated, because the parameter is assumed to be an unknown constant and not a random variable. The level of confidence pertains to the percentage of intervals, each calculated from a different random sample from the same distribution, that are expected to contain the true parameter value. The confidence does not pertain to the specific calculated interval (it could be from the unlucky 5% of intervals that do not contain the true parameter value).

As an example, a confidence interval for the parameter μ can be produced from a random sample drawn from a normal(μ, σ^2) population by calculating the appropriate functions of the data. Recall that, if each sample value is drawn from a normal distribution, the sample mean $\bar{X}$ has a normal($\mu, \sigma^2/n$) distribution, where *n* is

the sample size. Even if the sample values are drawn from a distribution that is not normal, by the central limit theorem, $\overline{X}$ will be approximately normal(μ, σ^2/n) for sufficiently large n. Assuming that σ^2 is known (from previous data and experience), the standardized normal random variable

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

is normal(0, 1), and tabulated in Appendix C. From these tables, values of w can be found for which

$$\Pr[-w \le Z \le w] = 1 - \alpha. \qquad \text{(B.1)}$$

For example, for $\alpha = .05$, $w = 1.96$. In this case, w is the 97.5th percentile of the standard normal distribution, commonly denoted $z_{0.975}$, or $z_{1-\alpha/2}$ for $\alpha = .05$.

Substituting for Z in Equation B.1 above, along with some algebraic manipulation, produces

$$\Pr[\overline{X} - w\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + w\frac{\sigma}{\sqrt{n}}] = 1 - \alpha,$$

which defines a 100(1 – α)% confidence interval for the population mean μ, where

$$LCL = \overline{X} - w\frac{\sigma}{\sqrt{n}} \qquad \text{(B.2)}$$

and

$$UCL = \overline{X} + w\frac{\sigma}{\sqrt{n}}, \qquad \text{(B.3)}$$

with $w = z_{1-\alpha/2}$.

A random sample will yield a specific sample mean. The numbers w and n are known, and σ was assumed to be known. Therefore, for a preassigned confidence level, values for LCL and UCL can be calculated to produce a specific 100(1 – α)% confidence interval for μ. Each of the random variables LCL and UCL is a statistic, and the interval (LCL, UCL) is a random interval formed from these statistics.

Usually the value of σ is not known. In this case, the unbiased estimator of the population variance, S^2, can be used to produce S, which can be used in the above equations in place of σ. Thus, the following standardized random variable, T, can be formed:

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

For sufficiently large n (say 25 or 30), T follows a normal(0, 1) distribution. If n is not sufficiently large, T follows a **Student's *t* distribution**, for which tabulated probabilities exist in many statistics books, and in Appendix C. The Student's t distribution depends on a parameter called the **degrees of freedom**. In the present example this parameter equals n – 1. Confidence intervals for the population mean can then be calculated similarly to the case where σ is known, using either the Student's t distribution or, when n is large, the normal distribution.

Confidence intervals can also be constructed for differences of means and many other population parameters, such as variances, probabilities, quantiles, and distribution characteristics (see, for example, Hogg and Craig 1978).

B.4.3 Hypothesis Testing

Testing a statistical hypothesis is another major area of statistics. A **hypothesis** is a statement about the distribution of the observable random variable. Often this statement is expressed as a statement about one or more parameters of the distribution. As discussed previously, estimation uses information in the data from a random sample to infer something about the magnitude of a parameter value. Similar to estimation, **hypothesis testing** also uses information from the random sample. However, the objective of hypothesis testing is to determine whether the specific statement about the distribution is true.

The hypothesis to be tested is referred to as the **null hypothesis**, denoted by H_0. The alternative to the null hypothesis is referred to as the **alternative hypothesis**, denoted H_1 or H_a. A **test of a hypothesis** is a rule or procedure for deciding whether to reject or accept the null hypothesis. This rule or procedure is based upon information contained in the random sample and produces a single number, called a test statistic, which leads to a decision of whether the sample values do not

support H_0. The entire set of values that the test statistic may assume is divided into two regions, one corresponding to the **rejection region** and the other to the **acceptance region**.

If the test statistic computed from a particular sample has a value in the rejection region, H_0 is rejected. If the test statistic falls in the acceptance region, H_0 is said to be accepted, due to lack of evidence to reject. For each of the two possible cases for H_0, true or false, the test either rejects or does not reject H_0, producing four distinct possibilities. These possibilities (using conditional probability notation), along with some concepts and terms associated with hypothesis testing, are summarized in Table B.2 (Martz and Waller, 1991).

Table B.2 Possible Hypothesis Test Outcomes.

	H_0 True	**H_0 False**
Accept H_0	Pr(accept H_0 | H_0 is true) = $1-\alpha$ = Level of confidence	Pr(accept H_0 | H_0 is false) = β = Pr(Type II Error)
Reject H_0	Pr(reject H_0 | H_0 is true) = α = Level of significance = Pr(Type I Error)	Pr(reject H_0 | H_0 is false) = $1-\beta$ = Power

A stated null hypothesis is either true or false. One of two errors can occur in hypothesis testing:

1. rejection of the null hypothesis when it is true, referred to as the **Type I error**, and
2. acceptance of the null hypothesis when it is false, referred to as the **Type II error**.

The probability of making a Type I error, denoted by α, is referred to as the **significance level of the test**. Thus, $1 - \alpha$ is the probability of making a correct decision when H_0 is true. The probability of making a correct decision when H_0 is false, denoted $1 - \beta$, is referred to as the **power of the test**. The probability of making a Type II error is equal to one minus the power of the test, or β.

The goodness of a statistical hypothesis test is measured by the probabilities of making a Type I or a Type II error. Since α is the probability that the test statistic will fall in the rejection region, assuming H_0 to be true, increasing the size of the rejection region will increase α and simultaneously decrease β for a fixed sample size. Reducing the size of the rejection region will decrease α and increase β. If the sample size, n, is increased, more information will be available for use in making the decision and both α and β will decrease.

The probability of making a Type II error, β, varies depending on the true value of the population parameter. If the true population parameter is very close to the hypothesized value, a very large sample would be needed to detect such a difference. That is, the probability of accepting H_0 when H_0 is false, β, varies depending on the difference between the true value and the hypothesized value. For hypothesis tests, α is specified prior to conducting the random sample. This fixed α specifies the rejection region. For a deviation from the hypothesized value that is considered practical and that is wished to be detectable by the hypothesis test, a sample size can be selected that will produce an acceptable value of β.

Different alternative hypotheses will result in different rejection regions for the same H_0. This is seen most easily for a hypothesis that is expressed in terms of a parameter, for example, H_0: $\mu = \mu_0$ for some given value μ_0. In this case, there is an exact correspondence between one-sided and two-sided confidence intervals and rejection regions for one-sided and two-sided alternative hypotheses. If the hypothesized value falls outside a $100(1 - \alpha)\%$ confidence interval for the corresponding population parameter, the null hypothesis would be rejected with level of confidence equal to $1 - \alpha$.

For the example presented in the previous section, Section B.3.4.2, the $100(1 - \alpha)\%$ two-sided confidence

interval for a population mean is defined by the *LCL* and *UCL* in Equations B.2 and B.3. For the hypothesized value of the mean, say μ_0, if $\mu_0 < lcl$ or $\mu_0 > ucl$, H_0 would be rejected. Equivalently, the test statistic in Equation B.1 can be computed using $\mu = \mu_0$ and, for α = .05, if it is greater than 1.96 or less than -1.96, H_0 would be rejected with 95% level of confidence.

To further illustrate these concepts, a more detailed example is presented. Atwood et al. (1998) assert that for non-momentary losses of offsite power with plant-centered causes, the recovery times are lognormally distributed with median 29.6 minutes and error factor 10.6. This is equivalent to X being normally distributed with μ = ln(29.6) = 3.388 and σ = ln(10.6)/1.645 = 1.435, where X = ln(recovery time in minutes). Suppose that a plant of interest has experienced five such losses of offsite power in recent history. It is desired to test whether the plant's recovery times follow the claimed distribution.

To simplify the situation, the question is formulated in terms of μ only, assuming that σ = 1.435. The null hypothesis is

H_0: $\mu = 3.388$.

Because only long recovery times are of concern from a risk standpoint, the alternative hypothesis is defined as

H_1: $\mu > 3.388$.

That is, values <3.388 are possible, but are not of concern. The test statistic, based on n = 5 recovery times, is to reject H_0 if

$$Z = \frac{\overline{X} - 3.388}{1.435/\sqrt{5}} > w \quad .$$

To make α, the probability of Type I error, equal to 0.05, w is chosen to be the 95th percentile of the standard normal distribution, 1.645. Then the test can be re-expressed as rejecting H_0 if

$\overline{X} > 4.44$.

The upper part of Figure B.1 shows the density of $\overline{X}$ when μ = 3.388. The area to the right of 4.44 is

$\Pr(\overline{X} > 4.44 \mid H_0 \text{ is true})$,

which equals 0.05.

What if H_0 is false? For example, a median 60-minute recovery time corresponds to μ = ln(60) = 4.09. The lower part of Figure B.1 shows the density of $\overline{X}$ when μ = 4.09. The area to the right of 4.44 is

$\Pr(\overline{X} > 4.44 \mid \mu = 4.09)$,

which is equal to 0.29. This value represents the power of the hypothesis test when μ = 4.09 and is the probability of (correctly) rejecting H_0. The probability of a Type II error when μ = 4.09 is 1 – 0.29 = 0.71.

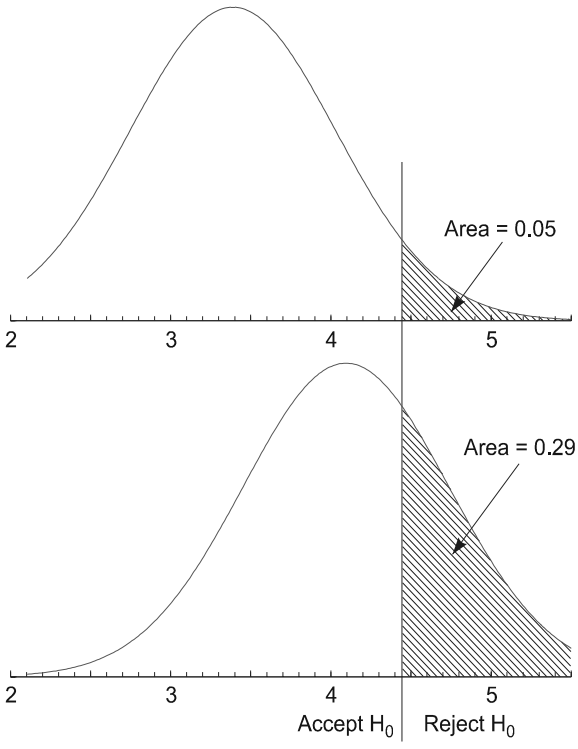

Figure B.1 Probability of rejecting H_0: μ = 3.388, if in fact H_0 is true (upper distribution), and if H_0 is false with μ = 4.09 (lower distribution).

It can be useful to plot the power as a function of μ. The plot is called a **power curve**. Figure B.2 shows two power curves, corresponding to n = 5 and n = 10. The probability of Type I error, that is, the probability of rejecting H_0 when H_0 is true, is shown as α. The

probability of Type II error, that is, the probability of accepting H_0 when H_0 is false, is shown as β for one value of μ, and equals 1 minus the power. The two tests, with $n = 5$ and $n = 10$, have both been calibrated so that $\alpha = 0.05$. The power, for any value of μ in H_1, is larger when $n = 10$ than when $n = 5$; equivalently, the probability of Type II error is smaller.

The interpretation of confidence in hypothesis testing is also the same as with confidence intervals. That is, the confidence is not in one specific test statistic. The confidence arises from the viewpoint that if the random sample was collected a number of times in the same way and if H_0 was true, $100(1 - \alpha)\%$ of the tests would result in not rejecting H_0.

As can be seen, interval estimation and hypothesis testing are closely related. Some experimenters prefer expressing inference as estimators. Others prefer to test a particular hypothesized value for the parameter of interest.

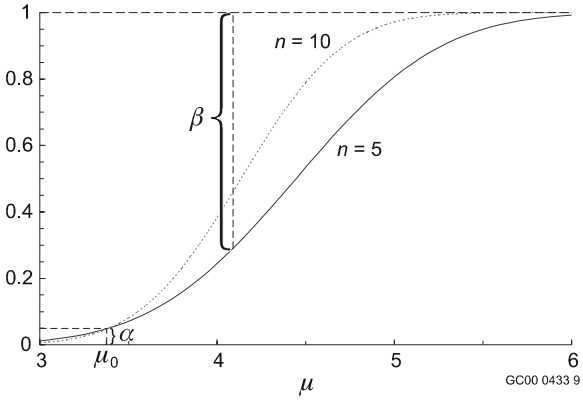

Figure B.2 Power curves when $n = 5$ and $n = 10$. The graph shows the probability of rejecting H_0, as a function of the true μ.

B.4.4 Goodness-of-Fit Tests

The methods presented above are concerned with estimating the parameters of a distribution, with the actual form of the distribution assumed to be known (or the central limit theorem applies with large n). Other hypothesis tests do not assume that only a parameter is unknown. In particular, **goodness-of-fit tests** are special hypothesis tests that can be used to check on the assumed distribution itself. Based on a random sample from some distribution, goodness-of-fit tests test the hypothesis that the data are distributed according to a specific distribution. In general, these tests are based on a comparison of how well the sample data agree with an expected set of data from the assumed distribution.

Perhaps the most familiar goodness-of-fit test is the **chi-square** test. The test statistic used for this statistical test has an approximate χ^2 distribution, leading to the name of the test. A random sample of n observations, $X_1, X_2, ..., X_n$, can be divided or binned into k groups or intervals, referred to as bins, producing an empirical distribution. The assumed distribution under the null hypothesis, $f_0(x)$, is used to calculate the probability that an observation would fall in each bin, with the probabilities denoted by $p_1, p_2, ..., p_k$.

These probabilities are frequently referred to as cell probabilities. The k bins are also called cells. The k bin intervals do not overlap and they completely cover the range of values of $f_0(x)$. It follows that $\sum_{i=1}^{k} p_i = 1$. The *expected* frequency of the ith bin, denoted e_i, is $e_i = np_i$, for $i = 1, 2, ..., k$. The e_i are commonly referred to as the expected cell counts. The *observed* frequencies for each of the k bins, denoted O_i, are referred to as observed cell counts.

The chi-square goodness-of-fit test compares the observed frequencies to the corresponding expected frequencies for each of the k groups by calculating the test statistic:

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}. \qquad \text{(B.5)}$$

If the observations come from some distribution other than that specified in the null hypothesis, the observed frequencies tend to agree poorly with the expected frequencies and the computed test statistic, X^2, becomes large.

The distribution of the quantity X^2 can be approximated by a chi-square distribution. The parameter that specifies the chi-square distribution is called the **degrees of freedom**. Its value depends on the number of unknown parameters and how they are estimated. When the null hypothesis distribution is normal with both μ and σ known, the degrees of freedom are $k - 1$. If $\bar{X}$ and S^2 from the sample are used to estimate μ and σ^2 when testing the distribution, the degrees of freedom

are between $k - 1$ and $k - 1 - m$, where m is the number of estimated parameters, 2. If the quantity X^2 is greater that the $1 - \alpha$ quantile of the $\chi^2(k - 1)$ distribution, the hypothesized probability distribution is rejected. If X^2 is less than the $1 - \alpha$ quantile of the $\chi^2(k - 1 - m)$ distribution, the data are concluded to be adequately modeled by $f_0(x)$.

When the sample size is small, the χ^2 distribution still applies as long as the expected frequencies are not too small. Larger expected cell counts make the chi-square distribution approximation better. The problem with small expected frequencies is that a single random observation falling in a group with a small expected frequency would result in that single value having a major contribution to the value of the test statistic, and thus, the test itself. In addition, small expected frequencies are likely to occur only in extreme cases. One rule of thumb is that no expected frequency should be less than 1 (see Snedecor and Cochran, 1989). Two expected frequencies can be near 1 if most of the other expected frequencies are greater than 5. Groups with expected frequencies below 1 should be combined or the groups should be redefined to comply with this rule. Note that k is the number of groups after such combination or redefinition.

Comparing how well sample data agree with an expected set of data leads to another common use of the chi-square test: testing whether two or more classification criteria, used to group subjects or objects, are independent of one another. Although not a goodness-of-fit test, the **chi-square test for independence** is similar to the chi-square goodness-of-fit test.

For two grouping criteria, the rows of a two-way **contingency table** can represent the classes of one of the criteria and the columns can represent the classes of the other criteria. To test the hypothesis that the rows and columns represent independent classifications, the expected number, e_{ij}, that would fall into each cell of the two-way table is calculated and used to compute the following chi-square test statistic:

$$X^2 = \sum_{i,\,j} \frac{(O_{ij} - e_{ij})^2}{e_{ij}},$$

where $i = 1, 2, ..., r$ (the number of rows); $j = 1, 2, ..., c$ (the number of columns); and O_{ij} is the number observed to belong to the ith row and jth column. The e_{ij} are calculated by

$$e_{ij} = \frac{R_i C_j}{n},$$

where R_i and C_j are the total observed in the ith row and jth column, respectively, and n is the total sample size ($n = \Sigma R_i = \Sigma C_j$).

For this test, the χ^2 test statistic follows a chi-square distribution with $(r - 1)(c - 1)$ degrees of freedom. If the calculated X^2 exceeds the $1 - \alpha$ quantile of the χ^2 distribution with $(r - 1)(c - 1)$ degrees of freedom, the null hypothesis of independence is rejected and the rows and columns are concluded to not represent independent classifications.

The **Kolmogorov goodness-of-fit test** tests the hypothesis that the observed random variable has c.d.f. $F_0(x)$ versus that the observed random variable does not have c.d.f. $F_0(x)$. It does this by comparing the sample c.d.f. (the empirical distribution function) to the hypothesized c.d.f. For a random sample of n observations, X_1, X_2, ..., X_n, the test statistic is defined as the maximum vertical distance between the empirical c.d.f., $\hat{F}(x)$ and $F_0(x)$. The actual procedure for calculating the test statistic can be found in many statistics texts, including Martz and Waller (1991) and Conover (1999). The test statistic is then compared to the $1 - \alpha$ quantile of tabled values for the Kolmogorov test, e.g. in Table C. If the calculated test statistic exceeds the $1 - \alpha$ quantile, the hypothesized c.d.f. is rejected. Otherwise, $F_0(x)$ is concluded to describe the data. The Kolmogorov goodness-of-fit test is based on each individual data point and therefore is equally effective for small or large samples.

As an example, consider the previous example of loss-of-offsite-power recovery times. Suppose that five recovery times have been observed at the plant: 7, 22, 94, 185, and 220 minutes. The corresponding values of x = ln(recovery time in minutes) are 1.95, 3.09, 4.54, 5.22, and 5.39. The null hypothesis and alternative hypothesis are

H_0: X is normal with $\mu = 3.388$, $\sigma = 1.435$
H_1: X has some other distribution .

Note, all possible alternative distributions are considered, not just normal distributions, or distributions with $\sigma = 1.435$.

Figure B.3 shows the distribution function specified by H_0 (the smooth curve) and the empirical distribution function specified by the data (the step function). The maximum distance between the two distributions is *D*, the Kolmogorov test statistic. If *D* is large, the test rejects H_0 in favor of H_1.

If the sample size is small, the Kolmogorov test may be preferred over the chi-square test. The Kolmogorov test is exact, even for small samples, while the chi-square test is an approximation that is better for larger sample sizes. The Kolmogorov statistic can also be used to construct a confidence region for the unknown distribution function.

The Kolmogorov goodness-of-fit test is sometimes called the Kolmogorov-Smirnov one-sample test. Statistics that are functions of the maximum vertical distance between $\hat{F}(x)$ and $F_0(x)$ are considered to be Kolmogorov-type statistics. Statistics that are functions of the maximum vertical distance between *two* empirical distribution functions are considered to be Smirnov-type statistics. A test of whether two samples have the same distribution function is the Smirnov test, which is a two-sample version of the Kolmogorov test presented above. This two-sample test is also called the **Kolmogorov-Smirnov two-sample test**. Conover (1999) presents additional information and tests.

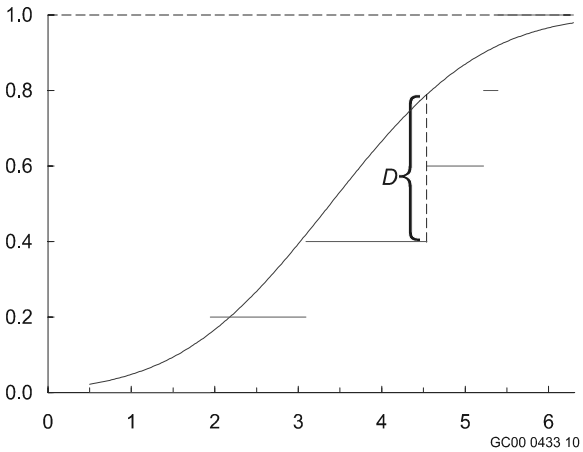

Figure B.3 The hypothesized distribution, the empirical distribution, and the Kolmogorov test statistic, *D*.

Another useful goodness-of-fit test is the **Anderson-Darling** goodness-of-fit test and test for normality. The Anderson-Darling test measures the squared difference between the empirical distribution function (EDF) of a sample and the theoretical distribution to be tested. It averages this squared difference over the entire range of the random variable, weighting the tails more heavily than the center. This statistic is recommended to guard against wayward observations in the tail and has generally good power.

Because many statistical methods require the assumption of normality, some assessment of whether data come from a normal population is helpful when considering appropriate analysis techniques. The Anderson-Darling statistic provides a measure of how much normal probability scores for the data (normal probability plot values) deviate from a straight line that would arise under normality. A computer package is often used to calculate this statistic and compare it to tabled values for the statistic. If the calculated statistic is too high, the deviations from the straight line are too large to be attributed to the variation due to sampling observations from a normal distribution. Thus, the hypothesis of normality is rejected. See the *Encyclopedia of Statistical Sciences*, 1982, for more information on the Anderson-darling goodness-of-fit test and Snedecor and Cochran (1989) for more information on the Anderson-Darling test used to test for normality.

Certain patterns of deviations from linearity in normal probability plots indicate common types of nonnormal characteristics, such as skewness or kurtosis (presence of long or short tails of the p.d.f.). Test for skewness or kurtosis are also available. See Snedecor and Cochran (1989) for more information on these tests.

B.5 Bayesian Estimation

B.5.1 Purpose and Use

Bayesian estimation is the other major class of statistical inference methods. Similar to frequentist estimation, both point and interval estimates can be obtained. However, Bayesian estimation is different from classical estimation in both practical and philosophical perspectives (NRC, 1994). Bayesian estimation incorporates degree of belief and information beyond

that contained in the data sample, forming the practical difference from classical estimation. The subjective interpretation of probability forms the philosophical difference from frequentist methods.

The prior belief about a parameter's value is contained in what is referred to as the **prior distribution**, which describes the state of knowledge (or subjective probability) about the parameter prior to obtaining the data sample. Therefore, in the Bayesian approach, the parameters of the sampling distribution have probability distributions. These probabilities do not model random variability of the parameters, but the analyst's uncertainty. Therefore, these distributions are sometimes called "uncertainty distributions," distributions that satisfies all the rules of probability.

Bayesian estimation consists of two main areas, both of which use the notion of subjective probability. The first area involves using available data to fit a subjective, prior distribution to a parameter, such as a failure rate. The degree of belief about the uncertainty in a parameter value is expressed in the prior distribution. This fitting of a prior distribution does not involve the use of Bayes' Theorem. The second area of Bayesian estimation involves using additional or new data to update an existing prior distribution. This is called **Bayesian updating**, and directly uses Bayes' Theorem.

Bayes' Theorem, presented in Section A.5, can be seen to transform the prior distribution by the effect of the sample data distribution to produce a **posterior distribution**. The sample data distribution, $f(x|\theta)$, can be viewed as a function of the unknown parameter instead of the observed data, x_i, producing a likelihood function, as discussed in Section B.4.1. Using the likelihood function, the fundamental relationship expressed by Bayes' Theorem is

$$\text{Posterior Distribution} = \frac{\text{Prior Distribution} \times \text{Likelihood}}{\text{Marginal Distribution}}.$$

The marginal distribution serves as a normalizing constant.

In Bayesian updating, the sampling distribution of the data provides new information about the parameter value. Bayes' Theorem provides a mathematical framework for processing new sample data as they become sequentially available over time. With the new information, the uncertainty of the parameter value has been reduced, but not eliminated. Bayes' Theorem is used to combine the prior and sampling distributions to form the posterior distribution, which then describes the updated state of knowledge (still in terms of subjective probability) about the parameter. **Point and interval estimates** of the parameter can then be obtained directly from the posterior distribution, which is viewed as containing the current knowledge about the parameter. This posterior distribution can then be used as the prior distribution when the next set of data becomes available. Thus, Bayesian updating is successively implemented using additional data in conjunction with Bayes' Theorem to obtain successively better posterior distributions that model plant-specific parameters.

Bayesian point and interval estimates are obtained from both the prior and posterior distributions. The interval estimates are **subjective probability intervals**, or **credible intervals**. The terminology is not yet universally standard. Berger (1985) and Bernardo and Smith (2000) both use the term **credible interval**, but Box and Tiao (1973) use **Bayes probability interval**, Lindley (1965) uses **Bayesian confidence interval**, and other authors have used other terms. A credible interval can be interpreted as a subjective probability statement about the parameter value, unlike classical interval estimates. That is, the interpretation of a 95% Bayesian posterior probability interval (a, b) is that, with 95% subjective probability, the parameter is contained in the interval (a, b), given the prior and sampling distributions.

B.5.2 Point and Interval Estimates

Bayesian parameter estimation involves four steps. The first step is identification of the parameter(s) to be estimated, which involves consideration of the assumed distribution of the data that will be collected. The second step is development of a prior distribution that appropriately quantifies the state of knowledge concerning the unknown parameter(s). The third step is collection of the data sample. The fourth and final step is combining the prior distribution with the data sample using Bayes' Theorem to produce the desired posterior distribution.

For PRA applications, determining the prior distribution is usually based on generic data and the data sample usually involves site-specific or plant-specific operating data. The resulting posterior distribution would then be the site-specific or plant-specific distribution of the parameter.

The plant-specific data collected are assumed to be a random sample from an assumed sampling distribution. The data are used to update the prior, producing the posterior distribution. Point estimates, such as the most likely value (the mode), the median, or (most commonly) the mean value, and probability interval estimates of the parameter can then be obtained. Other bounds and other point values can also be obtained with the Bayesian approach because the posterior parameter distribution is entirely known and represents the available knowledge about the parameter.

Bayesian interval estimation is more direct than classical interval estimation and is based solely on the posterior p.d.f.. A symmetric $100(1-\alpha)\%$ **two-sided Bayes probability interval estimate** of the parameter is easily obtained from the $\alpha/2$ and $1-\alpha/2$ quantiles of the posterior distribution. **Lower and upper one-sided Bayes probability interval estimates** can similarly be calculated. Again, note that the Bayes interval estimates are explicit probability statements of the true parameter being contained in the interval.

In some applications, such as a planned facility, plant-specific data do not exist. In these cases, Bayes' Theorem is not used. Only the generic data are used and parameter estimates are based solely on the assumed prior distribution. Investigation of the sensitivity of the results to the choice of the prior distribution is important for these cases.

B.5.3 Prior Distributions

The prior distribution is fundamental to any Bayesian analysis and represents all that is known or assumed about the parameter θ prior to collecting any data. The information summarized by the prior distribution can be objective, subjective, or both. Operational data and data from a previous but comparable experiment could be used as objective data. Subjective information could involve personal experience and opinions, expert judgement, assessments of degree of belief, and design information.

The selection of prior distributions can be seen to be somewhat subjective. A particular prior must be evaluated to determine the sensitivity of the choice of that prior on the parameter estimates. Consistency of the prior information and data with the prior distribution must be tested.

Choices for the initial prior distribution and techniques for handling various kinds of data are described in detail in several references, such as Martz and Waller (1991), Raiffa and Schlaifer (1961), and Siu and Kelly (1998) .

B.5.3.1 Noninformative Prior Distributions

One class of prior distributions that is widely used is termed **noninformative priors**, also referred to as priors of ignorance, or **reference priors** (Bernardo and Smith 1994). These names refer to the situation where very little *a priori* information about a parameter is available in comparison to the information expected to be provided by the data sample, or there is indifference about the range of values the parameter could assume.

One might think that this indifference could be expressed by a prior distribution that is uniformly distributed over the interval of interest. Every value in the interval is equally likely and no knowledge about any specific value over another value is imposed.

However, uniform distributions do not necessarily best reflect true noninformativeness (Box and Tiao 1973), because models can be parameterized in various ways. For example, if the time to failure, T, is exponentially distributed, it is common to write the density of T as

$$f(t) = \lambda e^{-\lambda t}$$

or alternatively as

$$f(t) = \frac{1}{\mu} e^{-t/\mu}.$$

The two parameters are related by $\lambda = 1/\mu$.

A uniform distribution cannot be said to automatically reflect ignorance and be used as a standard noninformative prior distribution. For the exponential example here, ignorance of λ implies ignorance of μ,

but λ and μ cannot both have a uniform distribution. In fact, suppose that λ has the uniform distribution in some finite range, say from a to b. Then μ has a density proportional to $1/\mu^2$ in the range from $1/b$ to $1/a$, as stated in Appendix A.4.7. The distribution of μ is *not* uniform.

Jeffreys' rule (Jeffreys 1961) is a rule that guides the choice of noninformative prior distributions and provides the Jeffreys prior distribution (Box and Tiao, 1973). The Jeffreys prior distribution is commonly used in PRA and involves using a specific parameterization of the model (distribution). Jeffreys' method is to transform the model into a parameterization that is in terms of a **location parameter**, a parameter that slides the distribution sideways without changing its shape. (See Box and Tiao 1978, Secs. 1.2.3 and 1.3.4). This method then uses the uniform distribution as the noninformative prior for the location parameter. It is reasonable to regard a uniform distribution as noninformative for a location parameter. The distribution for any other parameterization is then determined, and is called noninformative.

In the exponential example, working with log(time), let $\theta = \log(\mu)$, $S = \log(T)$, and $s = \log(t)$. Using algebraic formulas given in Section A.4.7 of Appendix A, it can be shown that the density in this parameterization is

$$f(s) = \exp(s - \theta)e^{-\exp(s-\theta)}.$$

Because θ only appears in the expression $s - \theta$, a change in θ simply slides the distribution sideways along the s axis. Therefore, θ is a location parameter. The Jeffreys noninformative prior is a uniform distribution for θ. This distribution translates to a density for λ which is proportional to $1/\lambda$, and a density for μ which is proportional to $1/\mu$. These are the Jeffreys noninformative prior distributions for λ and μ.

A further argument for Jeffreys prior distributions is that the resulting Bayes intervals are numerically equal to confidence intervals (Lindley 1958), and the confidence intervals are based on the data alone, not on prior belief. Unfortunately, the above approach cannot be followed exactly when the data come from a discrete distribution, such as binomial or Poisson. The original parameter can only approximately be converted to a location parameter. The resulting distribution is still called the Jeffreys prior, however, even though it only approximates the Jeffreys method.

To avoid the appearance of pulling prior distributions out of the air, the general formula for the Jeffreys prior is stated here, as explained by Box and Tiao (1973) and many others. All the particular cases given in this handbook can be found by working out the formula in those cases. Let θ denote the unknown parameter to be estimated. Let $L(\theta; x)$ denote the likelihood corresponding to a single observation. It is a function of θ, but it also depends on the data, x. For example, x is the number of Poisson events in a single unit of time, or the number of failures on a single demand, or the length of a single duration. Calculate

$$-\frac{d^2}{d\theta^2}\ln[L(\theta; x)] .$$

Now replace the number x by the random variable X, and evaluate the expectation of the calculated derivative:

$$E\left(-\frac{d^2}{d\theta^2}\ln[L(\theta; X)]\right) .$$

The Jeffreys noninformative prior is a function of θ proportional to the square root of this expectation.

B.5.3.2 Conjugate Prior Distributions

It is computationally convenient if the prior is a **conjugate prior distribution**. A conjugate prior distribution is a distribution that results in a posterior distribution that is a member of the same family of distributions as the prior. The methodology for obtaining conjugate priors is based on sufficient statistics and likelihood functions (see Martz and Waller, 1991).

The beta family of distributions is the conjugate family of prior distributions for the probability of failure of a component in a binomial sampling situation. The resulting posterior beta distribution can then be used to provide point and interval estimates of the failure probability.

A time-to-failure random variable is often assumed to follow an exponential distribution, with the failure events arising from a Poisson process. For this model, with either exponential or Poisson data, the gamma

family of distributions is the conjugate family of prior distributions for use in Bayesian reliability and failure rate analyses.

Figure B.4 is a schematic diagram showing the relation of the kinds of priors that have been mentioned so far. There are *very* many nonconjugate priors. A relatively small family of priors is conjugate, typically a single type such as the gamma distributions or beta distributions. Finally, the Jeffreys noninformative prior is a single distribution, shown in the diagram by a dot. In many cases, the Jeffreys prior is also conjugate, as indicated in the figure.

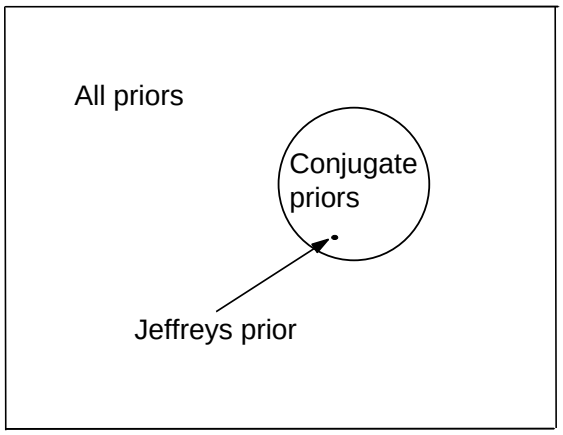

Figure B.4 Schematic diagram of types of priors.

A popular nonconjugate prior is the lognormal distribution. It can be used as a prior distribution for both the binomial sampling and Poisson process models above, although it is not conjugate.

Conjugate prior distributions provide convenience, but accurate modeling of prior degree of belief should not be sacrificed for mathematical convenience. However, when one expression of prior belief is viewed to be as correct as another, the more convenient expression is usually selected for use.

B.5.3.3 Other Prior Distribution Approaches

The prior distribution in Bayesian methods contains a subjective notion of probability since the frequencies of the values of the unknown parameter are seldom known. The prior distribution is thus the distribution of degree of belief before data that provide new information are obtained. Usually, the prior probabilities do not have a direct frequency interpretation and cannot be experimentally confirmed.

When the prior distribution does have a frequency interpretation, the observed data can be used to estimate the prior distribution. This situation represents another class of methods of statistical inference called **empirical Bayes** methods. The empirical Bayes prior distribution is empirically determined, for example, using observed plant-specific data for a given set of plants. Bayes' Theorem can then be applied to combine this prior with observed data from a specific plant to produce a posterior distribution. Thus, empirical Bayes methods are useful when data from similar, but not identical, sources exist. This situation also gives rise to the use of so-called **hierarchical Bayes** methods (see Gelman, et al., 1995, and Carlin and Louis, 1996).

Attempts have been made to remove some of the subjectivity present in selecting prior distributions, with the goal being to obtain *one* distribution for the same given information. That is, different analysts using the same information would decide upon the same prior distribution. The result has been development of the **method of maximum entropy**. If θ is a parameter with uncertainty distribution g, the entropy is defined as

$$-\int g(\theta)\ln[g(\theta 0]d\theta \ .$$

The distribution g that maximizes this expression is called the maximum entropy distribution. For finite ranges, the p.d.f. with the largest entropy is the uniform, or flat, distribution. Thus, entropy can be viewed as a measure of the variability in the height of a p.d.f., and a maximum entropy prior would be the one with the required mean that is as flat as possible. Siu and Kelly (1998, Table 2) give the maximum entropy distributions for a number of possible constraints.

Maximum entropy methods may see more use in the future, but still do not produce a systematic approach to selecting only one prior from a set of possible priors. In fact, the same problem that the Jeffreys' method attempts to address (Section B.5.3.1) is present with the maximum entropy approach: the same maximum entropy prior distribution cannot be used for different parameterizations and parameters of the same model, even though ignorance of each of the different parameters is viewed as equal.

To address this lack of invariance, **constrained noninformative priors** are obtained. They are based on the maximum entropy approach in conjunction with Jeffreys' method. That parameterization is used for which the parameter is a location parameter. Giving maximum entropy to this parameter produces a distribution called the constrained noninformative prior distribution. Atwood (1996) presents constrained noninformative priors and their application to PRA. Constrained noninformative prior distributions are seeing use in PRA, although not as much as Jeffreys' priors.

Other ways of defining noninformative prior distributions exist. See Martz and Waller (1991) and Berger (1985) for more information.

REFERENCES FOR APPENDIX B

Atwood, C. L., 1996, Constrained Noninformative Priors in Risk Assessment, in *Reliability Engineering and System Safety*, vol. 53, pp. 37-46.

Berger, J. O., 1985, *Statistical Decision Theory and Bayesian Analysis, Second Edition*, Springer-Verlag, New York.

Bernardo, J. M., and A. F. M. Smith, 1994, *Bayesian Theory*, John Wiley & Sons, New York.

Box, G. E. P., and G. C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, MA.

Carlin, B. P. and T. A. Louis, 1996, *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.

Conover, W. J., 1999, *Practical Nonparametric Statistics*, 3rd ed. John Wiley & Sons, NY.

Encyclopedia of Statistical Sciences, 1982, vol. 1. John Wiley & Sons, NY.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 1995, *Bayesian Data Analysis*. Chapman & Hall, London.

Hogg, R. V. and A. T. Craig, 1995, *Introduction to Mathematical Statistics*. Macmillan, NY.

Jeffreys, H., 1961, *Theory of Probability, Third Ed.*, Clarendon Press, Oxford, England.

Kendall, M. G. and A. Stuart, 1973, *The Advanced Theory of Statistics*, Vol. 2. Hafner, NY.

Lindley, D. V., 1958, Fiducial Distributions and Bayes' Theorem, in *Journal of the Royal Statistical Society, Series B*, Vol. 20, pp. 102-107.

Lindley, D. V., 1965, *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge University Press, Cambridge, UK.

Martz, H. F. and R. A. Waller, 1991, *Bayesian Reliability Analysis*. R. E. Krieger Publishing Co., Malabar, FL.

NRC, 1994, *A Review of NRC Staff uses of Probabilistic Risk Assessment*, NUREG-1489. U. S. Nuclear Regulatory Commission, Washington, D.C.

Raiffa, H., and R Schlaifer, 1961, *Applied Statistical Decision Theory*, The M.I.T. Press, Cambridge, MA.

SAS, 1988, *SAS Procedures Guide, Release 6.03 Edition*, SAS Institute, Inc., Cary, NC.

Siu, N., and D. Kelly, 1998, Bayesian Parameter Estimation in Probabilistic Risk Assessment, in *Reliability Engineering and System Safety*, Vol. 62, pp. 89-116.

Snedecor, G. W., and W. G. Cochran, 1989, *Statistical Methods*, Iowa State University Press, Ames, IA.