**AN INFORMATION THEORETIC APPROACH TO INCORPORATING
PRIOR INFORMATION IN BINOMIAL SAMPLING**

by

Danny Dyer and Paul Chiou

**Department of Mathematics
The University of Texas at Arlington
Arlington, Texas 76019**

Technical Report #198

May, 1983

AN INFORMATION THEORETIC APPROACH TO INCORPORATING
PRIOR INFORMATION IN BINOMIAL SAMPLING

Danny Dyer and Paul Chiou

Department of Mathematics
The University of Texas at Arlington
Arlington, Texas 76019

Key Words and Phrases: information theory; conservative prior distributions; likelihood dominated prior distributions; Kullback-Leibler distance; posterior robustness; reliability of series system.

ABSTRACT

The incorporation of prior information about θ , where θ is the success probability in a binomial sampling model, is an essential feature of Bayesian statistics. Methodology based on information theoretic concepts is introduced which (a) quantifies the amount of information provided by the sample data relative to that provided by the prior distribution and (b) allows for a ranking of prior distributions with respect to conservativeness, where conservatism refers to restraint of extraneous information about θ which is embedded in any prior distribution. In effect, the most conservative prior distribution from a specified class (each member of which carries the available prior information about θ) is that prior distribution within the class over which the likelihood function has the greatest average domination. The most conservative prior distributions from nine different families of prior distributions over the interval $(0,1)$ including the beta distribution are determined and compared for three situations: (1) no prior specification for θ is available, (2) a prior point estimate of θ is available, and (3) a prior

interval estimate of θ is available. The results of the comparisons not only advocate the use of the beta prior distribution in binomial sampling but also indicate which particular one to use in the three aforementioned situations.

1. INTRODUCTION

Suppose that sample data are obtained from a sequence of N independent Bernoulli trials each having probability of success θ , where N is fixed and θ is unknown. Such a situation is referred to as binomial sampling, and the sampling distribution of X , the number of successes, is a binomial distribution with pdf $f_{\theta}(x) = \binom{N}{x} \theta^x (1-\theta)^{N-x}$, $x = 0, 1, \dots, N$. Under the Bayesian approach, θ is assumed to be a value taken on by the random variable θ whose prior pdf is $g(\theta; \omega)$, where ω is a hyperparameter. A priori information as to the true value of θ is incorporated into the statistical procedure through $g(\theta; \omega)$. The sample data update this information with both sources of information about θ being then blended into the posterior pdf of θ given by

$$g(\theta|x;\omega) = \theta^x (1-\theta)^{N-x} g(\theta;\omega) / \int_0^1 \theta^x (1-\theta)^{N-x} g(\theta;\omega) d\theta, \quad 0 < \theta < 1. \quad (1.1)$$

Inferences about θ are ordinarily made through $g(\theta|x;\omega)$, and the merit of such inferences is dependent upon the credibility of $g(\theta;\omega)$. Unfortunately, a priori information about θ is usually limited and, in any event, insufficient to determine the precise form of $g(\theta;\omega)$. Indeed, a priori information about θ may be intangible, vague, or totally nonexistent; in which case some artificial form of $g(\theta;\omega)$ often surfaces albeit accompanied by theoretical justifications (e.g., Jeffreys 1961, Chap. 3; Jaynes 1968; Villegas 1977; Berger 1980, p. 74; Barnett 1973, p. 178).

When there is limited a priori information about θ , an

approach often used is (i) assume that $g(\theta; \omega)$ belongs to a specified family $G = \{g(\theta; \omega)\}$, usually the family of beta distributions, whose members are indexed by the hyperparameter ω , and (ii) choose the member which most closely matches prior beliefs. This is done by specifying prior moments and/or percentiles of the prior distribution, and then determining the value of ω , say ω_0 , such that $g(\theta; \omega_0)$ has these specifications (Weiler 1965; Martz and Waller 1982, pp. 236-237). However, such an approach presents troublesome situations such as (a) the unique determination of a k -dimensional hyperparameter ω requires k "pieces" of a priori information about θ , (b) slight changes in prior beliefs can often produce considerable changes in ω , thus causing significant changes in the decision procedure, and (c) not only is the a priori information about θ incorporated into the statistical procedure, but $g(\theta; \omega_0)$ also inadvertently carries extraneous information or noise about θ which may not be valid (Berger 1980, p. 66). In addition, it would seem that the most perplexing issue occurs when attempting to choose between $g(\theta; \omega_0)$ and another similarly determined prior pdf $g^*(\theta; \gamma_0)$ selected from a different specified family $G^* = \{g^*(\theta; \gamma)\}$.

Alternatively, we shall consider an information theoretic approach which will determine, within a specified family of prior pdfs $G = \{g(\theta; \omega)\}$, the particular prior distribution which not only incorporates prior specifications for θ , but also carries the *least* amount of extraneous information about θ . In addition, the unique determination of the hyperparameter ω can be made when the number of "pieces" of a priori information about θ is less than the dimension of ω . The procedure works equally well within a class of prior pdfs consisting of several specified families. Accordingly, we shall be able not only to determine which prior pdf should be chosen within a specified family but also to compare such chosen prior pdfs from several families.

2. MOST CONSERVATIVE PRIOR DISTRIBUTION

Information about θ is inherent in both the prior pdf of

θ , $g(\theta; \omega)$, and the likelihood function $L(\theta; x)$, which is the sampling distribution $f_\theta(x)$ regarded as a function of θ for fixed x . It is assumed that $g(\theta; \omega)$ belongs to a specified family $G = \{g(\theta; \omega)\}$ and encompasses the prior specifications, if any, for θ ; however, such specifications are not sufficient to uniquely determine $g(\theta; \omega)$. Joint consideration of the prior pdf and the likelihood function constitutes an *information base* for θ , say $I_x(\omega) = \{g(\theta; \omega), L(\theta; x)\}$, which carries not only a priori and data information about θ but also extraneous information about θ which is embedded in $g(\theta; \omega)$. It seems plausible that the amount of extraneous information should be less in an information base where the prior pdf is dominated by the likelihood function than in an information base where the likelihood function is dominated by the prior pdf. Therefore, as a means of reducing extraneous information about θ yet leaving intact the prior specifications, we consider domination of $g(\theta; \omega)$ by $L(\theta; x)$. Likelihood domination is usually appropriate since information supplied by the performance of the experiment should be much more precise than a priori information, otherwise, why perform the experiment (Box and Tiao 1973, pp. 21-22).

When $L(\theta; x)$ dominates $g(\theta; \omega)$, the standardized likelihood $L(\theta; x) / \int_0^1 L(\theta; x) d\theta$ is approximately numerically equal to the posterior pdf; on the other hand, it is the prior pdf that is approximately numerically equal to the posterior pdf when $g(\theta; \omega)$ dominates $L(\theta; x)$ (e.g., Box and Tiao 1973, p. 23). Accordingly, we associate with the information base $I_x(\omega)$ a measure of likelihood domination as quantified by the Kullback-Leibler distance measure between the prior and the posterior pdfs of θ , namely,

$$I[g(\theta|x; \omega), g(\theta; \omega)] = \int_0^1 g(\theta|x; \omega) \ln[g(\theta|x; \omega)/g(\theta; \omega)] d\theta. \quad (2.1)$$

The quantity $I[g(\theta|x; \omega), g(\theta; \omega)]$ may also be thought of as (a) the amount of information about θ in the observed number of successes, x , relative to the prior pdf $g(\theta; \omega)$

(e.g., Lindley 1956) and (b) the gain in information about θ when the prior pdf $g(\theta; \omega)$ is replaced by the posterior pdf $g(\theta|x; \omega)$ (e.g., Rényi 1970, p. 562). As noted, $I[g(\theta|x; \omega), g(\theta; \omega)]$ measures the domination of the likelihood function over $g(\theta; \omega)$ after observing x successes in N independent Bernoulli trials. Prior to the performance of the experiment, we may consider the possible information bases $I_0(\omega), I_1(\omega), \dots, I_N(\omega)$ and their corresponding measures of likelihood domination over $g(\theta; \omega)$. The *expected likelihood domination* over $g(\theta; \omega) \in G$ or, the term we prefer, the *average gain in information about θ provided by N independent Bernoulli trials* relative to the prior pdf $g(\theta; \omega) \in G$ is given by

$$\begin{aligned} E_X\{I[g(\theta|X; \omega), g(\theta; \omega)]\} &= \sum_{x=0}^N f(x) I[g(\theta|x; \omega), g(\theta; \omega)] \\ &= \int_0^1 \sum_{x=0}^N f_\theta(x) g(\theta; \omega) \ln[g(\theta|x; \omega)/g(\theta; \omega)] d\theta \\ &= \int_0^1 g(\theta; \omega) \sum_{x=0}^N f_\theta(x) \ln[g(\theta|x; \omega)/g(\theta; \omega)] d\theta \\ &= N \int_0^1 \theta \ln \theta g(\theta; \omega) d\theta + N \int_0^1 (1-\theta) \ln(1-\theta) g(\theta; \omega) d\theta \\ &\quad - \sum_{x=0}^N f(x) \ln[f(x)/\binom{N}{x}], \end{aligned} \quad (2.2)$$

where $f(x) = \int_0^1 \binom{N}{x} \theta^x (1-\theta)^{N-x} g(\theta; \omega) d\theta$ is the marginal distribution of X . Note that the right-hand side of (2.2) depends only on ω and N ; henceforth, for the sake of brevity, we write $\bar{I}_G(\omega; N)$ instead of $E_X\{I[g(\theta|X; \omega), g(\theta; \omega)]\}$. The following result will be needed.

Theorem 1. Suppose $g^*(\theta; \gamma)$ is the mirror image of $g(\theta; \omega) \in G$, where $g^*(\theta; \gamma)$ belongs to the specified family $G^* = \{g^*(\theta; \gamma)\}$ which is not necessarily distinct from G . Then $\bar{I}_{G^*}(\gamma; N) = \bar{I}_G(\omega; N)$.

Proof. Since $g^*(\theta; \gamma)$ is the mirror image of $g(\theta; \omega)$, then $g^*(\theta; \gamma) = g(1-\theta; \omega)$. By letting $y = N - x$ and $\eta = 1 - \theta$ in

(2.2), we have

$$\begin{aligned}
\bar{I}_{G^*}(\underline{\gamma}; N) &= N \int_0^1 \theta \ln \theta g^*(\theta; \underline{\gamma}) d\theta + N \int_0^1 (1-\theta) \ln(1-\theta) g^*(\theta; \underline{\gamma}) d\theta \\
&\quad - \sum_{x=0}^N \left\{ \int_0^1 \binom{N}{x} \theta^x (1-\theta)^{N-x} g^*(\theta; \underline{\gamma}) d\theta \right\} \\
&\quad \left\{ \ln \left[\int_0^1 \theta^x (1-\theta)^{N-x} g^*(\theta; \underline{\gamma}) d\theta \right] \right\} \\
&= N \int_0^1 (1-\eta) \ln(1-\eta) g^*(1-\eta; \underline{\gamma}) d\eta + N \int_0^1 \eta \ln \eta g^*(1-\eta; \underline{\gamma}) d\eta \\
&\quad - \sum_{y=0}^N \left\{ \int_0^1 \binom{N}{N-y} \eta^y (1-\eta)^{N-y} g^*(1-\eta; \underline{\gamma}) d\eta \right\} \\
&\quad \left\{ \ln \left[\int_0^1 \eta^y (1-\eta)^{N-y} g^*(1-\eta; \underline{\gamma}) d\eta \right] \right\} \\
&= N \int_0^1 \eta \ln \eta g(\eta; \underline{\omega}) d\eta + N \int_0^1 (1-\eta) \ln(1-\eta) g(\eta; \underline{\omega}) d\eta \\
&\quad - \sum_{y=0}^N \left\{ \int_0^1 \binom{N}{y} \eta^y (1-\eta)^{N-y} g(\eta; \underline{\omega}) d\eta \right\} \\
&\quad \left\{ \ln \left[\int_0^1 \eta^y (1-\eta)^{N-y} g(\eta; \underline{\omega}) d\eta \right] \right\} \\
&= \bar{I}_G(\underline{\omega}; N).
\end{aligned}$$

Bernardo (1979) has used $\bar{I}_G(\underline{\omega}; N)$ under rather general sampling conditions and derived the asymptotic form of $g(\theta; \underline{\omega})$ relative to the class of admissible prior distributions. The asymptotic form of $g(\theta; \underline{\omega})$, which is called the reference prior pdf, is thought to approximate the true prior pdf. Its derivation is based on the concept of maximizing the contribution

of the data and minimizing the contribution of the prior distribution. By restricting this underlying idea to specified families of prior distributions, we now lay the cornerstone of our procedure.

Definition 1. Let $g(\theta; \omega_i)$, $i = 1, 2$, be two prior pdfs which belong to the specified family $G = \{g(\theta; \omega)\}$. For fixed N , $g(\theta; \omega_1)$ is said to be a *more conservative* prior pdf than $g(\theta; \omega_2)$ if $\bar{I}_G(\omega_1; N) > \bar{I}_G(\omega_2; N)$.

Thus, we are provided with a means of ranking prior pdfs in G with respect to conservativeness, where conservatism refers to restraint of extraneous information about θ which is embedded in any prior pdf of θ . The degree of such restraint is based on the extent of the likelihood function's expected domination over the prior pdf. This ranking ability leads to the following.

Definition 2. For fixed N , the *most conservative* prior pdf in the specified family $G = \{g(\theta; \omega)\}$ is that $g(\theta; \omega^*) \in G$ which maximizes $\bar{I}_G(\omega; N)$.

The existence of the most conservative prior pdf in G is assured since $\bar{I}_G(\omega; N)$ is a concave functional of $g(\theta; \omega)$ for fixed N (Lindley 1956). Its dependency on N is a consequence of imposing maximum expected likelihood domination as the means of achieving minimum embedded extraneous information about θ yet leaving intact the prior specifications. In fact, since the prior specifications for θ are not sufficient to uniquely determine $g(\theta; \omega)$, we do not see that the determination of $g(\theta; \omega)$ can be made independently of the information bases for θ . Our approach, while in violation of the so-called "likelihood principle" (Berger 1980, pp. 23-28), is consistent with an increasingly popular position adopted by Box and Tiao (1973, pp. 44-46), Zellner (1977), Bernardo (1979), and Akaike (1980) that the sampling distribution should be taken into account when specifying the prior distribution in the event there is little or no available a priori information about θ . The emphasis here both first

and foremost is to incorporate prior specifications, if any, for θ into the statistical procedure. To achieve this goal with a minimal amount of noise requires methodology which yields the most conservative prior distribution as a by-product.

In the following sections, we shall determine the most conservative prior distributions within various specified families and under certain prior specifications for θ . Ultimately, such prior distributions may be compared by extending Definition 1 to a class of prior distributions consisting of several specified families.

3. FAMILIES OF PRIOR DISTRIBUTIONS

We now discuss several families of prior distributions for θ . In addition, for each family the computational form of $\bar{I}_G(\omega; N)$ will be given.

3.1. Beta Distribution: $B(\alpha, \beta)$

The most widely used prior distribution for θ is the beta distribution, denoted by $B(\alpha, \beta)$, the pdf of which is

$$g_1(\theta; \alpha, \beta) = [1/B(\alpha, \beta)] \theta^{\alpha-1} (1-\theta)^{\beta-1}, \quad 0 < \theta < 1 \quad (3.1)$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ with parameters $\alpha, \beta > 0$. The uniform prior distribution is $B(1, 1)$; Jeffreys' noninformative prior distribution is $B(.5, .5)$ which is sometimes called the arcsine distribution. Under a $B(\alpha, \beta)$ prior distribution, the posterior distribution is $B(x+\alpha, N-x+\beta)$; consequently, $B(\alpha, \beta)$ is referred to as the conjugate prior distribution for binomial sampling. There is a sample-oriented interpretation of the hyperparameters. Specifically, the a priori information about θ in $B(\alpha, \beta)$ is equivalent to observing α "pseudo" successes in $\alpha + \beta$ "pseudo" Bernoulli trials (Barnett 1973, p. 188). It is often mentioned that an advantage in using the $B(\alpha, \beta)$ prior distribution is the richness in pdf shapes, thereby providing a reasonable approximation to the true prior distribution. The different shapes of $B(\alpha, \beta)$ are summarized in Table 1. In

addition the graph of the pdf of $B(\alpha, \beta)$ is (i) symmetric about $\theta = 1/2$ when $\alpha = \beta$ and (ii) the mirror image of the graph of the pdf of $B(\beta, \alpha)$ (see Johnson and Kotz 1970, pp. 42-43).

TABLE 1.
Shape Properties of $B(\alpha, \beta)$

$\alpha \backslash \beta$	$0 < \beta < 1$	$\beta = 1$	$\beta > 1$
$0 < \alpha < 1$	U-shaped with antimode at $(\alpha-1)/(\alpha+\beta-2)$	reverse J-shaped	reverse J-shaped
$\alpha = 1$	J-shaped	uniform	reverse J-shaped
$\alpha > 1$	J-shaped	J-shaped	unimodal with mode at $(\alpha-1)/(\alpha+\beta-2)$

The marginal distribution of X is the beta-binomial distribution with pdf

$$f(x) = \int_0^1 \binom{N}{x} \theta^x (1-\theta)^{N-x} g_1(\theta; \alpha, \beta) d\theta$$

$$= \binom{N}{x} B(x+\alpha, N-x+\beta) / B(\alpha, \beta), \quad x = 0, 1, \dots, N.$$

In addition,

$$\int_0^1 \theta \ln \theta g(\theta; \alpha, \beta) d\theta = [1/B(\alpha, \beta)] \int_0^1 \theta^\alpha (1-\theta)^{\beta-1} \ln \theta d\theta$$

$$= [\alpha/(\alpha+\beta)] [\psi(\alpha+1) - \psi(\alpha+\beta+1)],$$

and

$$\int_0^1 (1-\theta) \ln(1-\theta) g(\theta; \alpha, \beta) d\theta = [1/B(\alpha, \beta)] \int_0^1 \theta^{\alpha-1} (1-\theta)^\beta \ln(1-\theta) d\theta$$

$$= [\beta/(\alpha+\beta)] [\psi(\beta+1) - \psi(\alpha+\beta+1)],$$

where $\psi(\cdot)$ is the digamma function (Gradshteyn and Ryzhik 1965, p. 538). It follows from (2.2) that the average gain in information about θ provided by N independent Bernoulli trials, relative to the $B(\alpha, \beta)$ prior distribution, is

$$\bar{I}_B(\alpha, \beta; N) = N[\alpha\psi(\alpha+1) + \beta\psi(\beta+1)]/(\alpha+\beta) - N\psi(\alpha+\beta+1)$$

$$- \sum_{x=0}^N \binom{N}{x} [B(x+\alpha, N-x+\beta)/B(\alpha, \beta)] \{\ln[B(x+\alpha, N-x+\beta)/B(\alpha, \beta)]\}. \quad (3.2)$$

By Theorem 1, $\bar{I}_B(\alpha, \beta; N) = \bar{I}_B(\beta, \alpha; N)$; thus, $\bar{I}_B(\alpha, \beta; N)$ is a symmetric function in α and β for fixed N .

3.2. Negative Log-Gamma Distribution: $NLG(\alpha, \beta)$

If $-\ln \theta$ has a gamma prior distribution with scale parameter $1/\alpha$ and shape parameter β , then θ has a negative log-gamma Type 1 prior distribution, denoted by $NLG_1(\alpha, \beta)$, with pdf

$$g_2(\theta; \alpha, \beta) = [\alpha^\beta / \Gamma(\beta)] \theta^{\alpha-1} (-\ln \theta)^{\beta-1}, \quad 0 < \theta < 1 \quad (3.3)$$

with parameters $\alpha, \beta > 0$. If instead we initially begin with $-\ln(1-\theta)$, then θ has a negative log-gamma Type 2 prior distribution, denoted by $NLG_2(\alpha, \beta)$, with pdf

$$g_3(\theta; \alpha, \beta) = [\alpha^\beta / \Gamma(\beta)] (1-\theta)^{\alpha-1} [-\ln(1-\theta)]^{\beta-1}, \quad 0 < \theta < 1. \quad (3.4)$$

Except for the uniform prior distribution, $NLG_1(1,1)$ or $NLG_2(1,1)$, which is common to both families, these two families of prior distributions are distinct. Martz and Waller (1982, pp. 242-246 and p. 306) have discussed the use and selection of a $NLG_1(\alpha, \beta)$ prior distribution in Bayesian reliability analysis. Griffiths and Schafer (1981) found the shapes of the beta and both types of negative log-gamma distributions to be quite similar; thus, they conclude that any of the three distributions could be used in most applications. The graph of the pdf of $NLG_1(\alpha, \beta)$ is the mirror image of the graph of the pdf of $NLG_2(\alpha, \beta)$. The different shapes of $NLG_1(\alpha, \beta)$ are discussed by Grassia (1977) and summarized in Table 2.

TABLE 2.

Shape Properties of $NLG_1(\alpha, \beta)$

$\alpha \backslash \beta$	$0 < \beta < 1$	$\beta = 1$	$\beta > 1$
$0 < \alpha < 1$	U-shaped with antimode at $\exp[-(\beta-1)/(\alpha-1)]$	reverse J-shaped	reverse J-shaped
$\alpha = 1$	J-shaped	uniform	reverse J-shaped
$\alpha > 1$	J-shaped	J-shaped	unimodal with mode at $\exp[-(\beta-1)/(\alpha-1)]$

By the change of variable $t = -\ln \theta$, we find that

$$\int_0^1 \theta^\alpha (-\ln \theta)^\beta d\theta = \Gamma(\beta+1)/(\alpha+1)^{\beta+1}.$$

Furthermore, by using the series $\ln(1-\theta) = -\sum_{j=1}^{\infty} \theta^j/j$, $0 \leq \theta < 1$, we obtain

$$\begin{aligned} & \int_0^1 \theta^{\alpha-1} (1-\theta) \ln(1-\theta) (-\ln \theta)^{\beta-1} d\theta \\ &= \sum_{j=1}^{\infty} \sum_{k=1}^2 (-1)^k \Gamma(\beta)/j(j+\alpha+k-1)^\beta \\ &= \Gamma(\beta) \{-1/(\alpha+1)^\beta + \sum_{j=2}^{\infty} [1/j(j-1)(j+\alpha)^\beta]\}. \end{aligned}$$

It follows from (2.2) that the average gain in information about θ provided by N independent Bernoulli trials, relative to the $NLG_1(\alpha, \beta)$ prior distribution, is

$$\begin{aligned} \bar{I}_{NLG_1}(\alpha, \beta; N) &= -N(\alpha+\beta+1)\alpha^\beta/(\alpha+1)^{\beta+1} \\ &+ N \sum_{j=2}^{\infty} [1/j(j-1)][\alpha/(\alpha+j)]^\beta - \sum_{x=0}^N f(x) \ln[f(x)/\binom{N}{x}], \end{aligned} \quad (3.5)$$

where the marginal distribution of X is

$$f(x) = \int_0^1 \binom{N}{x} \theta^x (1-\theta)^{N-x} g_2(\theta; \alpha, \beta) d\theta.$$

The marginal distribution of X can be numerically evaluated using a Gauss-Legendre quadrature formula (Stroud and Secrest 1966, pp. 99-156) following the transformation $t = 2\theta - 1$. A possible singularity of the integrand at the lower limit of integration when $x = 0$ can be removed by first splitting the integral into two integrals, i.e., $\int_0^1 h(\theta) d\theta = \int_0^{.5} h(\theta) d\theta + \int_{.5}^1 h(\theta) d\theta$. By Theorem 1, $\bar{I}_{NLG_1}(\alpha, \beta; N) = \bar{I}_{NLG_2}(\alpha, \beta; N)$.

3.3. Normal Generated Distribution: $NG(\alpha, \beta)$

Chiu (1974) derived a prior distribution for the fraction defective, say θ , of a batch of items each of which is assumed to have a normally distributed quality characteristic whose mean has a normal prior distribution with mean α and standard deviation β . Any item whose quality measurement is less than a certain specified value, taken to be zero, is classified as a defective. The resulting prior distribution for θ is called a normal generated prior distribution, denoted by $NG(\alpha, \beta)$, with pdf given by

$$g_4(\theta; \alpha, \beta) = (1/\beta) \exp\{[(\beta^2 - 1)\mu^2 + 2\mu\alpha - \alpha^2]/2\beta^2\}, \quad 0 < \theta < 1 \quad (3.6)$$

where $\theta = \Phi(-\mu)$ and $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. The parameters are $-\infty < \alpha < \infty$, $\beta > 0$. If the fraction nondefective is represented by θ , then θ has a $NG(-\alpha, \beta)$ prior distribution. The uniform prior distribution is $NG(0, 1)$. The different shapes of $NG(\alpha, \beta)$ are discussed by Chiu (1974) and summarized in Table 3. The graph of the pdf of $NG(\alpha, \beta)$ is the mirror image of the graph of the pdf of $NG(-\alpha, \beta)$.

TABLE 3.

Shape Properties of $NG(\alpha, \beta)$

$\alpha \backslash \beta$	$0 < \beta < 1$	$\beta = 1$	$\beta > 1$
$\alpha > 0$	unimodal with mode at $\phi[\alpha/(\beta^2-1)]$	reverse J-shaped	U-shaped with antimode at $\phi[\alpha/(\beta^2-1)]$
$\alpha = 0$	unimodal and symmetric about $\theta = 1/2$	uniform	U-shaped and symmetric about $\theta = 1/2$

By (2.2), the average gain in information about θ provided by N independent Bernoulli trials, relative to the $NG(\alpha, \beta)$ prior distribution, is

$$\begin{aligned} \bar{I}_{NG}(\alpha, \beta; N) = & N \int_0^1 \theta \ln \theta g_4(\theta; \alpha, \beta) d\theta + N \int_0^1 (1-\theta) \ln(1-\theta) g_4(\theta; \alpha, \beta) d\theta \\ & - \sum_{x=0}^N f(x) \ln[f(x)/\binom{N}{x}], \end{aligned} \quad (3.7)$$

where the marginal distribution of X is

$$f(x) = \int_0^1 \binom{N}{x} \theta^x (1-\theta)^{N-x} g_4(\theta; \alpha, \beta) d\theta.$$

The first two integrals in (3.7) as well as the marginal distribution of X except when $x = 0$ and $x = N$ have no singularities and can be numerically evaluated using a Gauss-Legendre quadrature formula following the transformation $t = 2\theta - 1$.

The marginal distribution of X has possible singularities of the integrand at the lower limit of integration when $x = 0$ and at the upper limit of integration when $x = N$. These possible singularities may be removed by making the transformation

$t = -[\phi^{-1}(\theta) + \alpha]/\sqrt{2\beta}$ and then numerically evaluating the integrals by using a Gauss-Hermite quadrature formula (Stroud and Secrest 1966, pp. 217-252). By Theorem 1, $\bar{I}_{NG}(\alpha, \beta; N)$

$= \bar{I}_{NG}(-\alpha, \beta; N)$; thus, $\bar{I}_{NG}(\alpha, \beta; N)$ is symmetric about $\alpha = 0$ for

fixed N .

3.4. Johnson S_B Distribution: $JSB(\alpha, \beta)$

By assuming that $\ln[\theta/(1-\theta)]$ has a normal prior distribution with mean α and standard deviation β , it follows that θ has a Johnson S_B prior distribution, denoted by $JSB(\alpha, \beta)$, with pdf given by

$$g_5(\theta; \alpha, \beta) = [1/\sqrt{2\pi}\beta \theta(1-\theta)] \exp\{-[\ln[\theta/(1-\theta)] - \alpha]^2 / 2\beta^2\}, \quad 0 < \theta < 1 \quad (3.8)$$

with parameters $-\infty < \alpha < \infty$, $\beta > 0$. The pdf given by (3.8) is a subfamily of one of the three general families of distributions proposed by Johnson (1949) for fitting a set of data. Martz and Waller (1982, p. 275) have mentioned the possible use of $JSB(\alpha, \beta)$ as the prior distribution for θ . The different shapes of $JSB(\alpha, \beta)$ are discussed by Johnson (1949) and Hahn and Shapiro (1967, pp. 201-204). The graph of the pdf of $JSB(\alpha, \beta)$ is (i) either unimodal or M-shaped (bimodal with an antimode in between) and (ii) the mirror image of the graph of the pdf of $JSB(-\alpha, \beta)$. Because of the bimodality cases as well as the transition cases when one mode and the antimode coalesce plus the fact that $g_5(\theta; \alpha, \beta) \rightarrow 0$ as $\theta \rightarrow 0$ and $\theta \rightarrow 1$, some rather unusual shapes are possible. However, if we restrict our attention to the subinterval $(\epsilon, 1-\epsilon)$, where ϵ is a small positive number, the shapes of $JSB(\alpha, \beta)$ are comparable to those of the other families previously considered and are summarized in Table 4.

TABLE 4.
Shape Properties of $JSB(\alpha, \beta)^*$

$\alpha \backslash \beta$	$0 < \beta < \sqrt{2}$	$\beta = \sqrt{2}$	$\beta > \sqrt{2}$
$\alpha > 0$	unimodal	unimodal	U-shaped or J-shaped
$\alpha = 0$	unimodal and symmetric about $\theta = 1/2$	uniform	U-shaped and symmetric about $\theta = 1/2$

*relative to the interval $(\epsilon, 1-\epsilon)$

By (2.2), the average gain in information about θ provided by N independent Bernoulli trials, relative to the $\text{JSB}(\alpha, \beta)$ prior distribution, is

$$\begin{aligned} \bar{I}_{\text{JSB}}(\alpha, \beta; N) = & N \int_0^1 \theta \ln \theta g_5(\theta; \alpha, \beta) d\theta + N \int_0^1 (1-\theta) \ln(1-\theta) g_5(\theta; \alpha, \beta) d\theta \\ & - \sum_{x=0}^N f(x) \ln[f(x) / \binom{N}{x}], \end{aligned} \quad (3.9)$$

where the marginal distribution of X is

$$f(x) = \int_0^1 \binom{N}{x} \theta^x (1-\theta)^{N-x} g_5(\theta; \alpha, \beta) d\theta.$$

There are no singularities in the integrals in (3.9). Numerical evaluation can be carried out by a Gauss-Legendre quadrature formula following the transformation $t = 2\theta - 1$. By Theorem 1, $\bar{I}_{\text{JSB}}(\alpha, \beta; N) = \bar{I}_{\text{JSB}}(-\alpha, \beta; N)$; thus, $\bar{I}_{\text{JSB}}(\alpha, \beta; N)$ is symmetric about $\alpha = 0$ for fixed N .

3.5. Negative Log-Lognormal Distribution: $\text{NLL}(\alpha, \beta)$

If $-\ln \theta$ has a lognormal prior distribution or, equivalently, $\ln(-\ln \theta)$ has a normal prior distribution with mean α and standard deviation β , then θ has a negative log-lognormal Type 1 prior distribution, denoted by $\text{NLL}_1(\alpha, \beta)$, with pdf given by

$$g_6(\theta; \alpha, \beta) = [1/\sqrt{2\pi} \beta \theta (-\ln \theta)] \exp\{-[\ln(-\ln \theta) - \alpha]^2 / 2\beta^2\}, \quad 0 < \theta < 1 \quad (3.10)$$

with parameters $-\infty < \alpha < \infty$, $\beta > 0$. If θ has a negative log-lognormal Type 2 prior distribution, denoted by $\text{NLL}_2(\alpha, \beta)$, the pdf is

$$g_7(\theta; \alpha, \beta) = \{1/\sqrt{2\pi} \beta (1-\theta) [-\ln(1-\theta)]\} \exp\{-[\ln[-\ln(1-\theta)] - \alpha]^2 / 2\beta^2\}, \quad 0 < \theta < 1. \quad (3.11)$$

Martz and Waller (1982, p. 275) have mentioned the possible use

of $NLL_2(\alpha, \beta)$ as the prior distribution of θ . The graph of the pdf of $NLL_1(\alpha, \beta)$ is (i) either reverse J-shaped or ω -shaped with a pole at $\theta = 0$ and (ii) the mirror image of the graph of the pdf of $NLL_2(\alpha, \beta)$. In addition, $g_6(\theta; \alpha, \beta) \rightarrow \infty$ as $\theta \rightarrow 0$ and $g_6(\theta; \alpha, \beta) \rightarrow 0$ as $\theta \rightarrow 1$. As in the case of $JSB(\alpha, \beta)$, the shapes of $NLL_1(\alpha, \beta)$ over $(\epsilon, 1-\epsilon)$ are comparable to those of the other families previously considered and are summarized in Table 5.

TABLE 5.

Shape Properties of $NLL_1(\alpha, \beta)^*$

$\alpha \backslash \beta$	$0 < \beta < 1$	$\beta = 1$	$\beta > 1$
$\alpha < 0$	unimodal	unimodal or J-shaped	U-shaped or J-shaped
$\alpha = 0$	unimodal	reverse J-shaped	U-shaped
$\alpha > 0$	reverse J-shaped	reverse J-shaped	U-shaped or reverse J-shaped

*relative to the interval $(\epsilon, 1-\epsilon)$

By (2.2), the average gain in information about θ provided by N independent Bernoulli trials, relative to the $NLL_1(\alpha, \beta)$ prior distribution, is

$$\begin{aligned}
 \bar{I}_{NLL_1}(\alpha, \beta; N) &= N \int_0^1 \theta \ln \theta g_6(\theta; \alpha, \beta) d\theta \\
 &\quad + N \int_0^1 (1-\theta) \ln(1-\theta) g_6(\theta; \alpha, \beta) d\theta \\
 &\quad - \sum_{x=0}^N f(x) \ln[f(x) / \binom{N}{x}], \quad (3.12)
 \end{aligned}$$

where the marginal distribution of X is

$$f(x) = \int_0^1 \binom{N}{x} \theta^x (1-\theta)^{N-x} g_6(\theta; \alpha, \beta) d\theta.$$

The first two integrals in (3.12) as well as the marginal distribution of X except when $x = 0$ have no singularities and can be numerically evaluated using a Gauss-Legendre quadrature formula following the transformation $t = 2\theta - 1$. When $x = 0$ the marginal distribution of X has a singularity of the integrand at the lower limit of integration which may be removed by making the transformation $t = [\ln(-\ln \theta) - \alpha] / \sqrt{2} \beta$. The resulting integral can be numerically evaluated using a Gauss-Hermite quadrature formula. By Theorem 1, $\bar{I}_{NLL_1}(\alpha, \beta; N) = \bar{I}_{NLL_2}(\alpha, \beta; N)$.

3.6. Kummer Distribution: $K(\alpha, \beta, \gamma)$

As previously noted, the conjugate prior distribution for binomial sampling is $B(\alpha, \beta)$. It is possible, however, to generalize this family to a conjugate three-parameter family of prior distributions by using the integral representation of Kummer's confluent hypergeometric function (see Erdélyi 1953, p. 255). Specifically, the Kummer distribution, denoted by $K(\alpha, \beta, \gamma)$, has pdf given by

$$g_\theta(\theta; \alpha, \beta, \gamma) = [1/B(\alpha, \beta) {}_1F_1(\alpha; \alpha + \beta; \gamma)] \theta^{\alpha-1} (1-\theta)^{\beta-1} \exp(\gamma\theta), \quad 0 < \theta < 1 \quad (3.13)$$

with parameters $\alpha, \beta > 0$ and $-\infty < \gamma < \infty$. The confluent hypergeometric function has series representation

$${}_1F_1(a; b; x) = 1 + \frac{a}{b} x + \frac{a(a+1)}{b(b+1)} \frac{x^2}{2!} + \frac{a(a+1)(a+2)}{b(b+1)(b+2)} \frac{x^3}{3!} + \dots$$

For specified α and β , the graph of the pdf of $K(\alpha, \beta, \gamma)$ is either identical to or a distortion of the graph of the pdf of $B(\alpha, \beta)$ depending on whether or not $\gamma = 0$. Thus, the $K(\alpha, \beta, \gamma)$ family is richer in shapes than is the $B(\alpha, \beta)$ family. By using Kummer's transformation (Erdélyi 1953, p. 253), ${}_1F_1(a; c; x) = e^x {}_1F_1(c-a; c; -x)$, it follows that $g_\theta(\theta; \alpha, \beta, \gamma) = g_\theta(1-\theta; \beta, \alpha, -\gamma)$. Therefore, the graph of the pdf of $K(\alpha, \beta, \gamma)$ is the mirror image of the graph of the pdf of $K(\beta, \alpha, -\gamma)$.

By (2.2), the average gain in information about θ provided by N independent Bernoulli trials, relative to the $K(\alpha, \beta, \gamma)$ prior distribution, is

$$\begin{aligned} \bar{I}_K(\alpha, \beta, \gamma; N) = & N \int_0^1 \theta \ln \theta g_\theta(\theta; \alpha, \beta, \gamma) d\theta \\ & + N \int_0^1 (1-\theta) \ln(1-\theta) g_\theta(\theta; \alpha, \beta, \gamma) d\theta \\ & - \sum_{x=0}^N f(x) \ln[f(x) / \binom{N}{x}], \end{aligned} \quad (3.14)$$

where the marginal distribution of X is

$$\begin{aligned} f(x) = & \int_0^1 \frac{\binom{N}{x} \theta^{x+\alpha-1} (1-\theta)^{N-x+\beta-1} \exp(\gamma\theta)}{B(\alpha, \beta) {}_1F_1(\alpha; \alpha+\beta; \gamma)} d\theta \\ = & \binom{N}{x} \frac{B(x+\alpha, N-x+\beta) {}_1F_1(x+\alpha; N+\alpha+\beta; \gamma)}{B(\alpha, \beta) {}_1F_1(\alpha; \alpha+\beta; \gamma)}. \end{aligned}$$

There are no singularities in the integrals in (3.14), and each can be numerically evaluated using a Gauss-Legendre quadrature formula following the transformation $t = 2\theta - 1$. By Theorem 1,

$$\bar{I}_K(\alpha, \beta, \gamma; N) = \bar{I}_K(\beta, \alpha, -\gamma; N).$$

3.7. Euler Distribution: $E(\alpha, \beta, \gamma, \delta)$

The Euler distribution, denoted by $E(\alpha, \beta, \gamma, \delta)$, is a conjugate four-parameter family of prior distributions for binomial sampling and has pdf given by

$$g_\theta(\theta; \alpha, \beta, \gamma, \delta) = [1/B(\alpha, \beta) {}_2F_1(\delta, \alpha; \alpha+\beta; \gamma)] \theta^{\alpha-1} (1-\theta)^{\beta-1} (1-\gamma\theta)^{-\delta},$$

$$0 < \theta < 1 \quad (3.15)$$

with parameter $\alpha, \beta > 0$, $|\gamma| < 1$, and $-\infty < \delta < \infty$. The hypergeometric function has series representation

$$\begin{aligned} {}_2F_1(a, b; c; x) = & 1 + \frac{ab}{c} x + \frac{a(a+1)b(b+1)}{c(c+1)} \frac{x^2}{2!} \\ & + \frac{a(a+1)(a+2)b(b+1)(b+2)}{c(c+1)(c+2)} \frac{x^3}{3!} + \dots \end{aligned}$$

which converges for $|x| < 1$. The pdf in (3.15) is called an Euler distribution due to Euler's integral representation of the hypergeometric function (Erdélyi 1953, p. 59). Such a distribution is used in a Bayesian analysis of component availability assuming exponentially distributed failure and repair times (Martz and Waller 1982, pp. 552-553; Brender 1968). For specified α and β , the graph of the pdf of $E(\alpha, \beta, \lambda, \delta)$ is either identical to or a distortion of the graph of the pdf of $B(\alpha, \beta)$ depending on whether or not $\gamma = 0$ or $\delta = 0$. Like the $K(\alpha, \beta, \gamma)$ family, the $E(\alpha, \beta, \gamma, \delta)$ family is richer in shapes than is the $B(\alpha, \beta)$ family. By using the transformation (Erdélyi 1953, p. 64), ${}_2F_1(a, b; c; x) = (1-x)^{-a} {}_2F_1[a, c-b; c; x/(x-1)]$, it follows that $g_\theta(\theta; \alpha, \beta, \gamma, \delta) = g_\theta[1-\theta; \beta, \alpha, \gamma/(\gamma-1), \delta]$. Therefore, the graph of the pdf of $E(\alpha, \beta, \gamma, \delta)$ is the mirror image of the graph of the pdf of $E[\beta, \alpha, \gamma/(\gamma-1), \delta]$.

By (2.2), the average gain in information about θ provided by N independent Bernoulli trials, relative to the $E(\alpha, \beta, \gamma, \delta)$ prior distribution, is

$$\begin{aligned} \bar{I}_E(\alpha, \beta, \gamma, \delta; N) = N \int_0^1 \theta \ln \theta g_\theta(\theta; \alpha, \beta, \gamma, \delta) d\theta \\ + N \int_0^1 (1-\theta) \ln(1-\theta) g_\theta(\theta; \alpha, \beta, \gamma, \delta) d\theta \\ - \sum_{x=0}^N f(x) \ln[f(x)/\binom{N}{x}], \end{aligned} \quad (3.16)$$

where the marginal distribution of X is

$$\begin{aligned} f(x) &= \int_0^1 \frac{\binom{N}{x} \theta^{x+\alpha-1} (1-\theta)^{N-x+\beta-1} (1-\gamma\theta)^{-\delta}}{B(\alpha, \beta) {}_2F_1(\delta, \alpha; \alpha+\beta; \gamma)} d\theta \\ &= \binom{N}{x} \frac{B(x+\alpha, N-x+\beta) {}_2F_1(\delta, x+\alpha; N+\alpha+\beta; \gamma)}{B(\alpha, \beta) {}_2F_1(\delta, \alpha; \alpha+\beta; \gamma)}. \end{aligned}$$

There are no singularities in the integrals in (3.16), and each may be numerically evaluated using a Gauss-Legendre quadrature formula following the transformation $t = 2\theta - 1$. By Theorem 1,

$$\bar{I}_E(\alpha, \beta, \gamma, \delta; N) = \bar{I}_E[\beta, \alpha, \gamma/(\gamma-1), \delta; N].$$

4. NO AVAILABLE PRIOR SPECIFICATION FOR θ

In the event of no available a priori information about θ , the most conservative prior distribution in a specified family $G = \{g(\theta; \omega)\}$ is determined by maximizing $\bar{I}_G(\omega; N)$ for fixed N . In this situation, most conservative does not reflect "total ignorance" but rather that the amount of extraneous information about θ embedded in the most conservative prior distribution is minimal relative to that which N independent Bernoulli trials can be expected to provide.

4.1. Computational Procedures

A double-precision computer program was written in FORTRAN to evaluate each expression for $\bar{I}_G(\omega; N)$ given in the previous section. Numerical integration based on a Gauss-Legendre quadrature formula used 160 points, and a Gauss-Hermite quadrature formula used 60 points. In addition, the program written to evaluate $\bar{I}_B(\alpha, \beta; N)$ made use of a 100,000 term digamma function subroutine accurate to the tenth significant digit to the right of the decimal.

Because of the concavity and symmetry of $\bar{I}_B(\alpha, \beta; N)$ for fixed N , the maximum value of $\bar{I}_B(\alpha, \beta; N)$ occurs on the line $\alpha = \beta$. When no a priori information about θ is available, the most conservative beta prior distribution is $B(\alpha_0, \alpha_0)$, where α_0 is the value of α which maximizes $\bar{I}_B(\alpha, \alpha; N)$. By utilizing the IMSL routine ZXGSN in conjunction with the computer program that evaluates $\bar{I}_B(\alpha, \beta; N)$, the values of α_0 for sample sizes $N = 2(1)20(10)50, 100, 200$ are given in Table 6. The corresponding values of $\bar{I}_B(\alpha_0, \alpha_0; N)$ are also tabulated and will be used to compare $B(\alpha_0, \alpha_0)$ with the most conservative prior distributions from other families. Table 6 also gives $\bar{I}_B(\alpha, \beta; N)$ for three commonly used noninformative prior distributions for θ : (a) Jeffreys' prior distribution, $B(.5, .5)$, (b) the uniform prior distribution, $B(1, 1)$, and (c) Villegas' improper prior

TABLE 6. (no available prior specification for θ)

The most conservative beta prior distribution, $B(\alpha_0, \alpha_0)$, and the average gain in information about θ provided by N independent Bernoulli trials, relative to noninformative $B(\alpha, \beta)$ prior distributions.

N	α_0	$\bar{I}_B(\alpha, \beta; N)$			
		$B(\alpha_0, \alpha_0)$	noninformative		
			$B(.5, .5)$	$B(1, 1)$	$B(0, 0)$
2	.041	.73	.48	.33	.69
3	.086	.80	.61	.44	.69
4	.118	.87	.71	.52	.69
5	.142	.93	.79	.60	.69
6	.161	.99	.85	.66	.69
7	.177	1.03	.91	.72	.69
8	.190	1.08	.96	.77	.69
9	.201	1.12	1.01	.81	.69
10	.211	1.15	1.05	.85	.69
11	.220	1.18	1.09	.89	.69
12	.228	1.22	1.13	.93	.69
13	.235	1.24	1.16	.96	.69
14	.242	1.27	1.19	.99	.69
15	.248	1.30	1.22	1.02	.69
16	.254	1.32	1.25	1.05	.69
17	.259	1.35	1.28	1.08	.69
18	.263	1.37	1.30	1.10	.69
19	.268	1.39	1.32	1.13	.69
20	.272	1.41	1.35	1.15	.69
30	.303	1.57	1.53	1.33	.69
40	.323	1.69	1.66	1.46	.69
50	.338	1.79	1.76	1.57	.69
100	.377	2.10	2.08	1.90	.69
200	.407	2.42	2.41	2.24	.69

distribution, $B(0, 0)$. For fixed N , $\bar{I}_B(.5, .5; N) > \bar{I}_B(1, 1; N)$; thus, Jeffreys' noninformative prior distribution is more conservative than the uniform prior distribution. For small to moderate sample sizes, under $B(.5, .5)$ *two or three additional* independent Bernoulli trials are required to gain, on the average, as much information about θ as provided by the original sample under $B(\alpha_0, \alpha_0)$. For small to moderate sample sizes, $B(1, 1)$ generally requires about *twice as many* independent Bernoulli trials as required by $B(\alpha_0, \alpha_0)$ to gain, on the average, an equal amount of information about θ . For sample sizes in excess

of 100, Jeffreys' noninformative prior distribution could be used as an approximation to the most conservative beta prior distribution. In fact, Bernardo (1979) has shown that $\lim_{N \rightarrow \infty} B(\alpha_0, \alpha_0) = B(.5, .5)$. Villegas' improper prior distribution, $B(0, 0)$, performs poorly since it produces a constant ($\ln 2 = .693$) average gain in information about θ from the sample regardless of the sample size. This is due to the marginal distribution of X , in this case, being $\text{Prob}(X = 0) = \text{Prob}(X = N) = .5$. If X represents the number of heads obtained when tossing a coin which is known to be either two-tailed or two-headed, with either being equally likely, a single toss of the coin reveals its nature, and all subsequent tosses produce no additional information about the coin.

Because of the concavity and symmetry about $\alpha = 0$ of $\bar{I}_{NG}(\alpha, \beta; N)$ for fixed N , the maximum value of $\bar{I}_{NG}(\alpha, \beta; N)$ is found by maximizing $\bar{I}_{NG}(0, \beta; N)$. By utilizing the IMSL routine ZXGSN in conjunction with the computer program that evaluates $\bar{I}_{NG}(\alpha, \beta; N)$, the maximum values of $\bar{I}_{NG}(\alpha, \beta; N)$ are given in Table 7. An identical statement to the above can also be made for $\bar{I}_{JSB}(\alpha, \beta; N)$. The maximum values of $\bar{I}_{NLG_1}(\alpha, \beta; N)$ and $\bar{I}_{NLL_1}(\alpha, \beta; N)$ are also given in Table 7 having been determined by using the IMSL routine ZXMIN in conjunction with the computer programs that evaluate these functions.

Interestingly enough, although both the $K(\alpha, \beta, \gamma)$ and $E(\alpha, \beta, \gamma, \delta)$ families are richer in shapes than the $B(\alpha, \beta)$ family, neither family can produce a prior distribution which is more conservative than $B(\alpha_0, \alpha_0)$. To see that this is true, we make use of the concavity of $\bar{I}_K(\alpha, \beta, \gamma; N)$ for fixed N and the fact that $\bar{I}_K(\alpha, \beta, \gamma; N) = \bar{I}_K(\beta, \alpha, -\gamma; N)$. It follows that the maximum value of $\bar{I}_K(\alpha, \beta, \gamma; N)$ occurs when $\alpha = \beta$ and $\gamma = 0$. But $K(\alpha, \alpha, 0) \equiv B(\alpha, \alpha)$; thus, the most conservative $K(\alpha, \beta, \gamma)$ prior distribution is $B(\alpha_0, \alpha_0)$. A similar result is obtained for the $E(\alpha, \beta, \gamma, \delta)$ family since $\bar{I}_E(\alpha, \beta, \gamma, \delta; N) = \bar{I}_E[\beta, \alpha, \gamma/(\gamma-1), \delta; N]$ implies the maximum value of $\bar{I}_E(\alpha, \beta, \gamma, \delta; N)$ occurs when $\alpha = \beta$ and $\gamma = 0$. However, $E(\alpha, \alpha, 0, \delta) \equiv B(\alpha, \alpha)$; thus, the most

TABLE 7. (no available prior specification for θ)

The average gain in information about θ provided by N independent Bernoulli trials, relative to the most conservative prior distributions from specified families.

N	B	NLG ₁ /NLG ₂	NG	JSB	NLL ₁ /NLL ₂
2	.73	.73	.73	.72	.73
3	.80	.80	.80	.80	.80
4	.87	.87	.87	.86	.87
5	.93	.93	.94	.92	.92
6	.99	.99	1.00	.97	.98
7	1.03	1.03	1.04	1.01	1.02
8	1.08	1.08	1.09	1.05	1.06
9	1.12	1.12	1.13	1.09	1.10
10	1.15	1.15	1.16	1.12	1.13
11	1.18	1.18	1.19	1.15	1.16
12	1.22	1.22	1.23	1.19	1.20
13	1.24	1.24	1.25	1.21	1.22
14	1.27	1.27	1.28	1.24	1.25
15	1.30	1.30	1.31	1.27	1.28
16	1.32	1.32	1.33	1.29	1.30
17	1.35	1.35	1.36	1.32	1.32
18	1.37	1.37	1.38	1.34	1.34
19	1.39	1.39	1.40	1.36	1.36
20	1.41	1.41	1.42	1.38	1.38
30	1.57	1.57	1.58	1.54	1.54
40	1.69	1.69	1.70	1.66	1.65
50	1.79	1.79	1.80	1.76	1.75
100	2.10	2.10	2.10	2.07	2.05
200	2.42	2.42	2.42	2.39	2.36

conservative $E(\alpha, \beta, \gamma, \delta)$ prior distribution is $B(\alpha_0, \alpha_0)$. We anticipate that $B(\alpha_0, \alpha_0)$ is the most conservative prior distribution from the conjugate multi-parameter family $\{g(\theta; \alpha, \beta, \omega) = c \theta^{\alpha-1} (1-\theta)^{\beta-1} h(\theta; \omega), \text{ where } h(\theta; \omega) \equiv 1 \text{ for some } \omega\}$.

4.2. Discussion

From Table 7 we see that, with one exception, the average gain in information about θ provided by N independent Bernoulli trials is essentially the same regardless of which of the families of prior distributions is selected *provided the most conservative member of that family is chosen to be the prior*

distribution for θ . The exception to the above occurs when $N \geq 12$. Here, the most conservative prior distributions from the $JSB(\alpha, \beta)$ family and both of the $NLL(\alpha, \beta)$ families require at least one additional independent Bernoulli trial to gain, on the average, as much information about θ as is gained from the original sample relative to the most conservative prior distributions from the other families. It is likely that this is due to the fact that, for all N , the most conservative prior distributions from the other families are true U-shaped [i.e., U-shaped over the interval $(0,1)$], whereas the most conservative prior distributions from the $JSB(\alpha, \beta)$ and $NLL(\alpha, \beta)$ families are only approximately U-shaped [i.e., U-shaped over the interval $(\epsilon, 1-\epsilon)$].

The U-shapedness of the most conservative prior distribution is essentially opposite to the unimodal shape of the corresponding posterior distribution. This oppositeness in shapes is a consequence of maximizing the Kullback-Leibler distance, as measured by (2.1), between the prior and posterior distributions of θ . We have depicted these shapes in Figure 1 when $N = 10$ and $x = 4$ for the $B(\alpha, \beta)$, $NLG_1(\alpha, \beta)$, $NG(\alpha, \beta)$, $JSB(\alpha, \beta)$, and $NLL_1(\alpha, \beta)$ families. Figure 1 shows an example of the similarity between the most conservative prior distributions which, in turn, produce almost identical posterior distributions. Since inferences about θ are made from the posterior distribution, one would expect these inferences to be in close agreement regardless of which family of prior distributions is assumed from those considered. Table 8 shows this to be the case. The means of the various posterior distributions are given when $N = 10, 20$, and 50 with selected values of x . The mean is a point estimate of θ under a squared-error loss function. When $N = 20$ and $x = 8$, the 5th and 95th percentiles of the various posterior distributions are also given. These percentiles correspond to lower and upper limits for a symmetric 90% two-sided Bayesian probability interval for θ .

FIGURE 1.

The most conservative prior and posterior distributions from selected families when $N = 10$ and $x = 4$.

TABLE 8. (no available prior specification for θ)

Point and interval estimates of θ based on the posterior distributions corresponding to the most conservative prior distributions from specified families.

		Point estimate of θ (mean of the posterior distribution)						
N	x	B	NLG ₁	NLG ₂	NG	JSB	NLL ₁	NLL ₂
10	2	.212	.211	.210	.216	.206	.232	.199
	4	.404	.412	.399	.406	.402	.425	.384
	6	.596	.601	.588	.594	.598	.615	.574
	8	.788	.790	.779	.784	.794	.800	.768
	10	.980	.977	.973	.975	.977	.968	.970
20	4	.208	.213	.207	.210	.204	.217	.201
	8	.403	.407	.400	.403	.401	.414	.392
	12	.597	.600	.593	.597	.599	.607	.587
	16	.792	.793	.787	.790	.795	.799	.783
	20	.987	.983	.984	.986	.983	.976	.985
50	10	.204	.205	.204	.205	.202	.208	.202
	20	.401	.402	.401	.402	.401	.406	.397
	30	.599	.599	.598	.598	.599	.603	.594
	40	.796	.796	.795	.795	.798	.799	.792
	50	.993	.992	.993	.993	.991	.985	.993

Symmetric 90% two-sided probability interval for θ
(5th and 95th percentiles of the posterior distribution)
N = 20, x = 8

limits	B	NLG ₁	NLG ₂	NG	JSB	NLL ₁	NLL ₂
lower	.234	.238	.232	.235	.232	.243	.226
upper	.582	.586	.579	.582	.582	.593	.571

4.3. Conclusions

The close similarities between the most conservative prior distributions from each of the families considered suggest a general form of *robustness with respect to the choice of the prior distribution* in the following sense. If we define the subclass, Γ , of prior distributions for θ to consist of the most conservative prior distributions from families which have a wide

variety of true U-shaped members, the Bayesian analysis of θ is basically unaffected regardless of which member of Γ is chosen (see Berger 1980, pp. 129-133). Incidentally, we also determined the most conservative prior distributions from two other families: the two-parameter gamma and Weibull distributions truncated to the right of one. Neither family has a member with any resemblance to U-shapedness. Under the most conservative prior distribution from either family, about twice as many independent Bernoulli trials are required to gain, on the average, as much information about θ as provided by the original sample under, say, $B(\alpha_0, \alpha_0)$.

With regard to minimizing the extraneous information about θ embedded in a prior distribution, each of the most conservative prior distributions which we considered performs, more or less, no better nor no worse than the others. However, inferences about θ are made from the posterior distribution, and it is here that we feel that *there are certain advantages in using the most conservative $B(\alpha, \beta)$ prior distribution.*

- (1) Either the mean or the mode of the posterior distribution is typically taken to be a point estimate of θ . Under the $B(\alpha_0, \alpha_0)$ prior distribution, the posterior mean is $(x + \alpha_0) / (N + 2\alpha_0)$ and the posterior mode is $(x + \alpha_0 - 1) / (N + 2\alpha_0 - 1)$. The mean or mode of the posterior distribution corresponding to the most conservative prior distribution from each of the other families must be numerically evaluated by using Gaussian quadrature formulas.
- (2) Percentiles of the posterior distribution correspond to lower and/or upper limits for a Bayesian probability interval for θ . Such percentiles can be easily determined under the $B(\alpha_0, \alpha_0)$ prior distribution by using available computer programs that evaluate the incomplete beta function ratio (e.g., IMSL routine MDBETAI).

- (3) The beta distribution has been widely studied over the years, and a great many results have been obtained. For example, recent studies have been made in which the distribution of the product of beta random variables has been derived (Springer 1979, p. 104). Such products occur when doing a Bayesian analysis of the reliability of certain types of systems (see Section 7).

5. PRIOR POINT ESTIMATE OF θ IS AVAILABLE

Suppose that the mean of the prior distribution of θ is specified to be a fixed number M . The most conservative prior distribution with specified mean M from a specified family $G = \{g(\theta; \omega)\}$ is that prior distribution in G which not only has mean M but also leaves the most additional information about the true value of θ to be gained, on the average, from N independent Bernoulli trials. Such a prior distribution is determined by maximizing $\bar{I}_G(\omega; N)$ for fixed N subject to the parametric constraint $\int_0^1 \theta g(\theta; \omega) d\theta = M$.

5.1. Computational Procedures

Since the mean of the $B(\alpha, \beta)$ prior distribution is $\alpha/(\alpha+\beta)$, for fixed N the most conservative $B(\alpha, \beta)$ prior distribution with specified mean M is $B(\alpha_M, \beta_M)$, where α_M is the value of α which maximizes $\bar{I}_B[\alpha, (1-M)\alpha/M; N]$ and $\beta_M = (1-M)\alpha_M/M$. By utilizing the IMSL routine ZXGSN in conjunction with the computer program that evaluates $\bar{I}_B(\alpha, \beta; N)$, the values of α_M along with the corresponding values of $\bar{I}_B(\alpha_M, \beta_M; N)$ are given in Table 9 for sample sizes $N = 2(1)20(10)50, 100, 200$ and $M = .5(.1).9, .99, .999$. Table 9 can also be used when $M < .5$. By Theorem 1 it follows that $B(\beta_M, \alpha_M)$ is the most conservative $B(\alpha, \beta)$ prior distribution with specified mean $1 - M$, and $\bar{I}_B(\beta_M, \alpha_M; N) = \bar{I}_B(\alpha_M, \beta_M; N)$.

Grassia (1977) has shown that the mean of the $NLG_1(\alpha, \beta)$ prior distribution is $[\alpha/(\alpha+1)]^\beta$. For fixed N , the most

TABLE 9. (prior point estimate of θ is available)

The most conservative beta prior distribution, $B(\alpha_M, \beta_M)$, with specified mean M and the corresponding average gain in information about θ , $\bar{I}_B(\alpha_M, \beta_M; N)$, provided by N independent Bernoulli trials. Note that $\beta_M = (1-M)\alpha_M/M$.

N	M	α_M	$\bar{I}_B$	N	M	α_M	$\bar{I}_B$	N	M	α_M	$\bar{I}_B$
2	.5	.041	.73	3	.5	.086	.80	4	.5	.118	.87
	.6	.050	.71		.6	.106	.78		.6	.145	.85
	.7	.063	.65		.7	.134	.71		.7	.185	.78
	.8	.083	.53		.8	.180	.59		.8	.252	.65
	.9	.128	.35		.9	.288	.39		.9	.416	.44
	.99	.466	.06		.99	1.124	.08		.99	1.677	.09
	.999	1.189	.01		.999	2.601	.01		.999	3.816	.02
5	.5	.142	.93	6	.5	.161	.99	7	.5	.177	1.03
	.6	.175	.91		.6	.199	.96		.6	.218	1.01
	.7	.223	.83		.7	.255	.88		.7	.281	.92
	.8	.308	.69		.8	.355	.74		.8	.394	.78
	.9	.522	.47		.9	.613	.51		.9	.693	.54
	.99	2.155	.10		.99	2.579	.12		.99	2.965	.13
	.999	4.913	.02		.999	5.926	.02		.999	6.876	.02
8	.5	.190	1.08	9	.5	.201	1.12	10	.5	.211	1.15
	.6	.235	1.05		.6	.249	1.09		.6	.262	1.12
	.7	.303	.96		.7	.323	1.00		.7	.340	1.04
	.8	.429	.81		.8	.459	.85		.8	.487	.88
	.9	.764	.57		.9	.829	.60		.9	.888	.62
	.99	3.319	.14		.99	3.647	.15		.99	3.955	.16
	.999	7.776	.03		.999	8.632	.03		.999	9.453	.03
11	.5	.220	1.18	12	.5	.228	1.22	13	.5	.235	1.24
	.6	.273	1.16		.6	.283	1.19		.6	.292	1.21
	.7	.356	1.07		.7	.370	1.10		.7	.382	1.12
	.8	.512	.91		.8	.534	.93		.8	.555	.96
	.9	.943	.64		.9	.993	.67		.9	1.040	.69
	.99	4.245	.17		.99	4.519	.17		.99	4.779	.18
	.999	10.242	.03		.999	11.003	.03		.999	11.740	.04
14	.5	.242	1.27	15	.5	.248	1.30	16	.5	.254	1.32
	.6	.301	1.24		.6	.308	1.27		.6	.315	1.29
	.7	.394	1.15		.7	.405	1.17		.7	.415	1.20
	.8	.575	.98		.8	.593	1.00		.8	.609	1.02
	.9	1.084	.71		.9	1.125	.73		.9	1.164	.74
	.99	5.028	.19		.99	5.266	.20		.99	5.494	.21
	.999	12.455	.04		.999	13.149	.04		.999	13.824	.04

(continued)

17	.5	.259	1.35	18	.5	.263	1.37	19	.5	.268	1.39
	.6	.322	1.31		.6	.328	1.34		.6	.334	1.36
	.7	.424	1.22		.7	.433	1.24		.7	.441	1.26
	.8	.625	1.05		.8	.640	1.06		.8	.654	1.08
	.9	1.201	.76		.9	1.236	.78		.9	1.269	.79
	.99	5.713	.21		.99	5.925	.22		.99	6.129	.23
	.999	14.483	.04		.999	15.126	.05		.999	15.754	.05
20	.5	.272	1.41	30	.5	.303	1.57	40	.5	.323	1.69
	.6	.339	1.38		.6	.379	1.54		.6	.405	1.66
	.7	.449	1.28		.7	.507	1.43		.7	.546	1.55
	.8	.668	1.10		.8	.772	1.25		.8	.844	1.36
	.9	1.301	.81		.9	1.554	.94		.9	1.735	1.03
	.99	6.326	.23		.99	8.014	.29		.99	9.351	.34
	.999	16.368	.05		.999	21.917	.07		.999	26.686	.08
50	.5	.338	1.79	100	.5	.377	2.10	200	.5	.407	2.42
	.6	.424	1.75		.6	.475	2.06		.6	.517	2.38
	.7	.575	1.64		.7	.656	1.94		.7	.724	2.26
	.8	.899	1.44		.8	1.058	1.73		.8	1.198	2.04
	.9	1.874	1.11		.9	2.293	1.38		.9	2.672	1.67
	.99	10.468	.39		.99	14.348	.54		.99	18.721	.74
	.999	30.926	.09		.999	47.575	.15		.999	70.362	.23

conservative $NLG_1(\alpha, \beta)$ prior distribution with specified mean M is $NLG_1(\alpha_M, \beta_M)$, where α_M is the value of α which maximizes $\bar{I}_{NLG_1}(\alpha, \ln M / \ln[\alpha/(\alpha+1)]; N)$ and $\beta_M = \ln M / \ln[\alpha_M/(\alpha_M+1)]$. By utilizing the IMSL routine ZXGSN in conjunction with the computer program that evaluates $\bar{I}_{NLG_1}(\alpha, \beta; N)$, the values of $\bar{I}_{NLG_1}(\alpha_M, \beta_M; N)$ are given in Table 10 for sample sizes $N = 2(1)10(10)50, 100, 200$ and $M = .01, .2(.2).8, .99$. The tabulations in Table 10 reflect the fact that the most conservative $NLG_2(\alpha, \beta)$ prior distribution with specified mean $1-M$ is the mirror image of the most conservative $NLG_1(\alpha, \beta)$ prior distribution with specified mean M .

Chiu (1974) has shown that the mean of the $NG(\alpha, \beta)$ prior distribution is $\phi(-\alpha/\sqrt{1+\beta^2})$. For fixed N , the most conservative $NG(\alpha, \beta)$ prior distribution with specified mean M is $NG(\alpha_M, \beta_M)$, where β_M is the value of β which maximizes $\bar{I}_{NG}[-\sqrt{1+\beta^2}\phi^{-1}(M), \beta; N]$ and $\alpha_M = -\sqrt{1+\beta_M^2}\phi^{-1}(M)$. By utilizing the IMSL routine ZXGSN in conjunction with the computer program that evaluates $\bar{I}_{NG}(\alpha, \beta; N)$, the values of $\bar{I}_{NG}(\alpha_M, \beta_M; N)$ are given in Table 10.

TABLE 10. (prior point estimate of θ is available)

The average gain in information about θ provided by N independent Bernoulli trials, relative to the most conservative prior distributions with specified mean M from specified families.

M	N	B	NLG ₁	NLG ₂	NG	JSB	NLL ₁	NLL ₂
.60	2	.71	.71	.71	.71	.71	.71	.71
	3	.78	.78	.78	.78	.78	.78	.78
	4	.85	.85	.85	.85	.85	.85	.85
	5	.91	.91	.91	.92	.90	.90	.90
	6	.96	.96	.96	.97	.95	.95	.95
	7	1.01	1.01	1.01	1.02	1.00	1.00	1.00
	8	1.05	1.05	1.05	1.06	1.03	1.03	1.04
	9	1.09	1.09	1.09	1.10	1.07	1.07	1.08
	10	1.12	1.12	1.12	1.13	1.10	1.10	1.11
	20	1.38	1.38	1.38	1.39	1.35	1.35	1.35
	30	1.54	1.54	1.54	1.55	1.51	1.51	1.51
	40	1.66	1.66	1.66	1.67	1.63	1.62	1.62
	50	1.75	1.75	1.75	1.76	1.72	1.71	1.71
.80	100	2.06	2.06	2.06	2.06	2.03	2.01	2.01
	200	2.38	2.38	2.38	2.38	2.35	2.33	2.33
	2	.53	.53	.53	.53	.53	.53	.53
	3	.59	.59	.59	.59	.58	.58	.59
	4	.65	.65	.65	.65	.64	.63	.65
	5	.70	.70	.70	.70	.69	.68	.70
	6	.74	.74	.74	.74	.73	.72	.74
	7	.78	.78	.78	.78	.77	.76	.78
	8	.81	.81	.81	.81	.80	.79	.81
	9	.85	.85	.85	.85	.83	.83	.85
	10	.88	.88	.88	.88	.86	.86	.88
	20	1.10	1.10	1.10	1.10	1.07	1.07	1.10
	30	1.25	1.25	1.25	1.25	1.21	1.21	1.24
	40	1.36	1.36	1.36	1.36	1.32	1.32	1.35
	50	1.44	1.44	1.44	1.44	1.40	1.40	1.43
	100	1.73	1.73	1.73	1.73	1.69	1.68	1.72
	200	2.04	2.04	2.04	2.04	2.00	1.99	2.02

(continued)

.99	2	.06	.06	.06	.06	.06	.06	.06
	3	.08	.08	.08	.08	.07	.07	.08
	4	.09	.09	.09	.09	.08	.08	.09
	5	.10	.10	.10	.10	.09	.09	.10
	6	.12	.12	.12	.11	.10	.10	.12
	7	.13	.13	.13	.12	.11	.11	.13
	8	.14	.14	.14	.13	.12	.12	.14
	9	.15	.15	.15	.14	.13	.13	.15
	10	.16	.16	.16	.15	.14	.14	.16
	20	.23	.23	.22	.22	.20	.19	.23
	30	.29	.29	.28	.27	.25	.24	.29
	40	.34	.34	.32	.31	.29	.28	.33
	50	.39	.39	.36	.36	.32	.31	.38
	100	.54	.54	.51	.50	.45	.44	.52
	200	.74	.73	.69	.68	.62	.61	.72

By Theorem 1, $NG(-\alpha_M, \beta_M)$ is the most conservative $NG(\alpha, \beta)$ prior distribution with specified mean $1-M$, and $\bar{I}_{NG}(-\alpha_M, \beta_M; N) = \bar{I}_{NG}(\alpha_M, \beta_M; N)$.

The mean of the $JSB(\alpha, \beta)$ prior distribution must be determined by numerically integrating $\int_0^1 \theta g_5(\theta; \alpha, \beta) d\theta$. The integrand has no singularities. By making the transformation $t = \{\ln[\theta/(1-\theta)] - \alpha\}/\sqrt{2}\beta$, the resulting integral is $(1/\sqrt{\pi}) \times \int_{-\infty}^{\infty} [1 + \exp(-\sqrt{2}\beta t + \alpha)]^{-1} \exp(-t^2) dt$ which may be numerically evaluated by a 60-point Gauss-Hermite quadrature formula. For fixed N , the most conservative $JSB(\alpha, \beta)$ prior distribution with specified mean M is $JSB(\alpha_M, \beta_M)$, where α_M is the value of α which maximizes $\bar{I}_{JSB}[\alpha, \beta_M(\alpha); N]$ and $\beta_M(\alpha)$ is the value of β which satisfies the constraint equation $\int_{-\infty}^{\infty} [1 + \exp(-\sqrt{2}\beta t + \alpha)]^{-1} \exp(-t^2) dt - \sqrt{\pi} M = 0$. By utilizing the IMSL routine ZXGSN in conjunction with a computer program that evaluates $\bar{I}_{JSB}(\alpha, \beta; N)$ subject to the constraint equation (which was solved by the IMSL routine ZBRENT), the values of $\bar{I}_{JSB}(\alpha_M, \beta_M; N)$ are given in Table 10. By Theorem 1, $JSB(-\alpha_M, \beta_M)$ is the most conservative $JSB(\alpha, \beta)$ prior distribution with specified mean $1-M$, and $\bar{I}_{JSB}(-\alpha_M, \beta_M; N) = \bar{I}_{JSB}(\alpha_M, \beta_M; N)$.

The mean of the $NLL_1(\alpha, \beta)$ prior distribution must be determined by numerically integrating $\int_0^1 \theta g_6(\theta; \alpha, \beta) d\theta$. The integrand has no singularities. By making the transformation

$t = [\ln(-\ln \theta) - \alpha] / \sqrt{2\beta}$, the resulting integral is $(1/\sqrt{\pi}) \times \int_{-\infty}^{\infty} \exp[-\exp(\sqrt{2\beta}t + \alpha)] \exp(-t^2) dt$ which may be numerically evaluated by a 60-point Gauss-Hermite quadrature formula. For fixed N , the most conservative $NLL_1(\alpha, \beta)$ prior distribution with specified mean M is $NLL_1(\alpha_M, \beta_M)$, where α_M is the value of α which maximizes $\bar{I}_{NLL_1}[\alpha, \beta_M(\alpha); N]$ and $\beta_M(\alpha)$ is the value of β which satisfies the constraint equation $\int_{-\infty}^{\infty} \exp[-\exp(\sqrt{2\beta}t + \alpha)] \exp(-t^2) dt - \sqrt{\pi} M = 0$. By utilizing the IMSL routine ZXGSN in conjunction with a computer program that evaluates $\bar{I}_{NLL_1}(\alpha, \beta; N)$ subject to the constraint equation (which was solved by the IMSL routine ZBRENT), the values of $\bar{I}_{NLL_1}(\alpha_M, \beta_M; N)$ are given in Table 10 taking into account the mirror image property within the $NLL(\alpha, \beta)$ families.

As was the case in the previous section, neither of the $K(\alpha, \beta, \gamma)$ and $E(\alpha, \beta, \gamma, \delta)$ families can produce a prior distribution with specified mean M which is more conservative than $B(\alpha_M, \beta_M)$. For fixed N , the most conservative $K(\alpha, \beta, \gamma)$ prior distribution with specified mean M is determined by maximizing $\bar{I}_K[\alpha, \beta_M(\alpha, \gamma), \gamma; N]$, where $\beta_M(\alpha, \gamma)$ is the value of β which satisfies the constraint equation $\int_0^1 \theta g_\beta(\theta; \alpha, \beta, \gamma) d\theta - M = 0$. Fix N and M . It can be shown that for every pair (a, c) , $c > 0$, there is at least one pair $(a^*, -c)$ which satisfies the equation

$$\bar{I}_K[a^*, \beta_M(a^*, -c), -c; N] = \bar{I}_K[a, \beta_M(a, c), c; N].$$

Since $\bar{I}_K[\alpha, \beta_M(\alpha, \gamma), \gamma; N]$ has a unique maximum, that maximum must occur when $\gamma = 0$. Thus, we maximize $\bar{I}_K[\alpha, \beta_M(\alpha, 0), 0; N] = \bar{I}_B[\alpha, (1-M)\alpha/M; N]$; hence, the most conservative $K(\alpha, \beta, \gamma)$ prior distribution with specified mean M is $B(\alpha_M, \beta_M)$. A similar result holds for the $E(\alpha, \beta, \gamma, \delta)$ family.

5.2. Discussion

When $.2 \leq M \leq .8$, we see from Table 10 the same type of robustness with respect to the choice of the prior distribution

that occurred in the previous section when there was no available prior specification for θ . There is no essential difference with respect to the average gain in information provided by N independent Bernoulli trials, relative to the most conservative prior distributions with specified mean M , where $.2 \leq M \leq .8$, from each of the $B(\alpha, \beta)$, $NLG_1(\alpha, \beta)$, $NLG_2(\alpha, \beta)$, and $NG(\alpha, \beta)$ families. And, as before, additional Bernoulli trials are required under the most conservative prior distributions from the $JSB(\alpha, \beta)$ and $NLL(\alpha, \beta)$ families.

This type of robustness does not, however, extend to extreme values of M . For example, when considering prior distributions with specified mean $M = .99$, the greatest average gain in information about θ provided by the original sample occurs relative to the most conservative prior distributions from only the $B(\alpha, \beta)$ and $NLG_1(\alpha, \beta)$ families. When $N \geq 20$, additional Bernoulli trials are required under the most conservative prior distributions from the $JSB(\alpha, \beta)$, $NLL(\alpha, \beta)$, $NG(\alpha, \beta)$, and $NLG_2(\alpha, \beta)$ families to gain, on the average, as much information about θ as is gained from the original sample relative to, say, $B(\alpha_{.99}, \beta_{.99})$. The reason for this is due to very slight differences in the shapes of the prior distributions. As can be seen from Tables 1 and 9, $B(\alpha_{.99}, \beta_{.99})$ is J-shaped. However, the $JSB(\alpha, \beta)$ and $NLL(\alpha, \beta)$ families only have approximate J-shaped members. The $NG(\alpha, \beta)$ family has J-shaped members, but only when $\alpha < 0$ and $\beta = 1$. Thus, the $NG(\alpha, \beta)$ family does not appear to provide as much of a variety of J-shapes as is provided by the $B(\alpha, \beta)$ family (see Table 1). Because of this J-shaped limitation, $NG(\alpha_{.99}, \beta_{.99})$ actually turns out to be (a) U-shaped with antimode very near zero when $N \leq 50$ and (b) unimodal with mode very near one when $N \geq 100$; therefore, $NG(\alpha_{.99}, \beta_{.99})$ is only approximately J-shaped. Finally, the J-shape of the most conservative $NLG_2(\alpha, \beta)$ prior distribution with specified mean $M = .99$ is slightly flatter than the J-shape of $B(\alpha_{.99}, \beta_{.99})$. Incidentally, since $NLG_1(\alpha, \beta)$ is the mirror image of $NLG_2(\alpha, \beta)$, it follows that when $M = .01$, it is now the reverse J-shaped most conservative prior

distributions from the $B(\alpha, \beta)$ and $NLG_2(\alpha, \beta)$ families which are preferred, with additional Bernoulli trials being required under the slightly flatter reverse J-shaped most conservative $NLG_1(\alpha, \beta)$ prior distribution.

5.3. Conclusions

From among the families of prior distributions considered, only the $B(\alpha, \beta)$ family gives a most conservative prior distribution with specified mean M which, for all M , consistently allows the original sample to provide the greatest average gain in information about θ . This dominating feature plus the three additional reasons discussed in the previous section make $B(\alpha_M, \beta_M)$ the clear choice when the only available a prior information is the specification that M is the mean of θ .

For fixed N , $\bar{I}_{B(\alpha_M, \beta_M; N)}$ is an increasing function of M when $M \leq .5$, then decreases when $M > .5$. Increases (decreases) in the specification of M when $M \geq .5$ ($M < .5$) must be accompanied by additional Bernoulli trials in order to maintain a stable average gain in sample information about θ . The most significant feature of $B(\alpha_M, \beta_M)$ can be seen from Table 9. Over a rather wide range of middle values of M , there are only slight changes in the magnitudes of both α_M and β_M relative to the magnitude of N when N is fixed. In other words, $B(\alpha_M, \beta_M)$ is *robust with respect to moderate misspecifications of M other than extreme values*. To illustrate, under a squared-error loss function, the Bayes point estimate of θ is given by $\hat{\theta} = (x + \alpha_M) / (N + \alpha_M + \beta_M)$. Suppose $N = 10$ and consider Table 11 where $\hat{\theta}$ is the entry. We see that the estimate of θ based on a misspecified value of M which is within .2 of the true value of M is generally within .015 of the estimate of θ based on the true value of M . For larger values of N the overall agreement between estimates is even better. Finally, we note the interesting result that the Bayes and the usual classical estimate, x/N , of θ are equal whenever the classical estimate of θ is numerically equal to the specification for M .

TABLE 11.

Point estimates of θ based on the mean of the posterior distribution corresponding to the $B(\alpha_M, \beta_M)$ prior distribution with $N = 10$.

$x \backslash M$	.3	.4	.5	.6	.7	.8
2	.205	.208	.212	.217	.223	.234
4	.395	.400	.404	.408	.413	.423
6	.586	.592	.596	.600	.605	.611
8	.777	.783	.788	.792	.795	.800
10	.968	.975	.980	.983	.986	.989

6. PRIOR INTERVAL ESTIMATE OF θ IS AVAILABLE

Suppose an a priori interval estimate of θ is available which is of the form $\text{Prob}(\theta_L < \theta < \theta_U) = \gamma$, where θ_L, θ_U , and γ are specified. Such an interval estimate could be viewed as a prior specification that, with probability γ , the true value of θ is $\theta_0 \pm d$. For fixed N , the most conservative $B(\alpha, \beta)$ prior distribution for which $\text{Prob}(\theta_L < \theta < \theta_U) = \gamma$ holds is $B(\alpha_F, \beta_F)$ where α_F is the value of α which maximizes $\bar{I}_B[\alpha, \beta(\alpha); N]$, $\beta(\alpha)$ is the value(s) of β which satisfies the constraint equation $\int_{\theta_L}^{\theta_U} [1/B(\alpha, \beta)] t^{\alpha-1} (1-t)^{\beta-1} dt - \gamma = 0$ for fixed α , and $\beta_F \equiv \beta(\alpha_F)$. The constraint equation may be solved using the IMSL routines MDBETA and ZBRENT. By utilizing the IMSL routine ZXGSN in conjunction with the computer program that evaluates $\bar{I}_B(\alpha, \beta; N)$, the values of α_F and β_F along with the corresponding values of $\bar{I}_B(\alpha_F, \beta_F; N)$ are given in Table 12 for choices of (θ_L, θ_U) which correspond to all possible 1/4, 1/2, and 3/4 subdivisions of the interval $(0, 1)$. The value of γ was chosen to be .95, and the sample sizes are $N = 2(1)20(10)50, 100, 200$. Since $B(\alpha_F, \beta_F)$ is the most conservative $B(\alpha, \beta)$ prior distribution when $\text{Prob}(\theta_L < \theta < \theta_U) = .95$, then $B(\beta_F, \alpha_F)$ is the most conservative $B(\alpha, \beta)$ prior distribution when $\text{Prob}(1-\theta_U < \theta < 1-\theta_L) = .95$. Except in one extreme situation, there are no U-shaped prior distributions. Excluding this case, $B(\alpha_F, \beta_F)$ is J-shaped when $(\theta_L, \theta_U) = (.25, 1), (.50, 1), (.75, 1)$;

TABLE 12. (prior interval estimate of θ is available)

The most conservative beta prior distribution, $B(\alpha_F, \beta_F)$, for which $\text{Prob}(\theta_L < \theta < \theta_U) = .95$ and the corresponding average gain in information about θ , $\bar{I}_B(\alpha_F, \beta_F; N)$, provided by N independent Bernoulli trials.

N	(θ_L, θ_U)	α_F	β_F	$\bar{I}_B$	N	(θ_L, θ_U)	α_F	β_F	$\bar{I}_B$
2	(.25, 1)	.973	.168	.33	3	(.25, 1)	1.080	.210	.42
	(.50, 1)	1.048	.079	.23		(.50, 1)	1.509	.142	.29
	(.75, 1)	1.072	.040	.15		(.75, 1)	1.555	.058	.18
	(.25, .75)	6.916	6.916	.07		(.25, .75)	6.916	6.916	.10
	(.50, .75)	35.100	20.740	.02		(.50, .75)	35.100	20.740	.03
4	(.25, 1)	1.147	.240	.49	5	(.25, 1)	1.197	.264	.56
	(.50, 1)	1.719	.177	.35		(.50, 1)	1.857	.204	.40
	(.75, 1)	1.978	.077	.21		(.75, 1)	2.371	.097	.25
	(.25, .75)	6.916	6.916	.13		(.25, .75)	6.916	6.916	.15
	(.50, .75)	35.100	20.740	.04		(.50, .75)	35.100	20.740	.04
6	(.25, 1)	1.234	.282	.61	7	(.25, 1)	1.263	.298	.66
	(.50, 1)	1.960	.225	.45		(.50, 1)	2.043	.242	.49
	(.75, 1)	2.740	.117	.27		(.75, 1)	3.052	.136	.30
	(.25, .75)	6.916	6.916	.18		(.25, .75)	6.916	6.916	.20
	(.50, .75)	35.100	20.740	.05		(.50, .75)	35.100	20.740	.06
8	(.25, 1)	1.288	.311	.71	9	(.25, 1)	1.310	.323	.75
	(.50, 1)	2.110	.257	.52		(.50, 1)	2.167	.270	.56
	(.75, 1)	3.313	.153	.33		(.75, 1)	3.541	.169	.35
	(.25, .75)	6.916	6.916	.23		(.25, .75)	6.916	6.916	.25
	(.50, .75)	35.100	20.740	.07		(.50, .75)	35.100	20.740	.08
10	(.25, 1)	1.328	.333	.78	11	(.25, 1)	1.344	.342	.82
	(.50, 1)	2.217	.282	.59		(.50, 1)	2.261	.293	.62
	(.75, 1)	3.734	.183	.37		(.75, 1)	3.900	.195	.39
	(.25, .75)	6.916	6.916	.27		(.25, .75)	6.916	6.916	.29
	(.50, .75)	35.100	20.740	.08		(.50, .75)	35.100	20.740	.09
12	(.25, 1)	1.358	.350	.85	13	(.25, 1)	1.371	.358	.88
	(.50, 1)	2.299	.302	.64		(.50, 1)	2.334	.311	.67
	(.75, 1)	4.042	.206	.41		(.75, 1)	4.167	.216	.43
	(.25, .75)	6.916	6.916	.31		(.25, .75)	6.916	6.916	.33
	(.50, .75)	35.100	20.740	.10		(.50, .75)	35.100	20.740	.11
14	(.25, 1)	1.383	.365	.91	15	(.25, 1)	1.393	.371	.93
	(.50, 1)	2.367	.319	.70		(.50, 1)	2.395	.326	.72
	(.75, 1)	4.277	.225	.45		(.75, 1)	4.379	.234	.47
	(.25, .75)	6.916	6.916	.35		(.25, .75)	6.916	6.916	.37
	(.50, .75)	35.100	20.740	.11		(.50, .75)	35.100	20.740	.12

(continued)

16	(.25,1)	1.403	.377	.96	17	(.25,1)	1.412	.383	.98
	(.50,1)	2.423	.333	.74		(.50,1)	2.446	.339	.76
	(.75,1)	4.467	.242	.49		(.75,1)	4.552	.249	.50
	(.25,.75)	6.916	6.916	.38		(.25,.75)	6.916	6.916	.40
	(.50,.75)	35.100	20.740	.13		(.50,.75)	35.100	20.740	.13
18	(.25,1)	1.422	.389	1.01	19	(.25,1)	1.429	.394	1.03
	(.50,1)	2.470	.346	.78		(.50,1)	2.489	.351	.80
	(.75,1)	4.632	.256	.52		(.75,1)	4.705	.263	.54
	(.25,.75)	6.916	6.916	.42		(.25,.75)	6.916	6.916	.43
	(.50,.75)	35.100	20.740	.14		(.50,.75)	35.100	20.740	.15
20	(.25,1)	1.436	.398	1.05	30	(.25,1)	1.487	.431	1.22
	(.50,1)	2.509	.356	.82		(.50,1)	2.655	.397	.97
	(.75,1)	4.768	.268	.55		(.75,1)	5.268	.316	.68
	(.25,.75)	6.916	6.916	.45		(.25,.75)	6.916	6.916	.58
	(.50,.75)	35.100	20.740	.15		(.50,.75)	35.100	20.740	.22
40	(.25,1)	1.520	.453	1.34	50	(.25,1)	1.544	.469	1.44
	(.50,1)	2.750	.424	1.09		(.50,1)	2.817	.444	1.18
	(.75,1)	5.586	.348	.77		(.75,1)	5.816	.373	.86
	(.25,.75)	6.916	6.916	.68		(.25,.75)	6.916	6.916	.76
	(.50,.75)	35.100	20.740	.27		(.50,.75)	35.100	20.740	.32
100	(.25,1)	1.603	.511	1.76	200	(.25,1)	1.645	.542	2.09
	(.50,1)	2.990	.498	1.49		(.50,1)	3.120	.541	1.81
	(.75,1)	6.436	.442	1.13		(.75,1)	6.920	.501	1.43
	(.25,.75)	6.916	6.916	1.05		(.25,.75)	6.916	6.916	1.37
	(.50,.75)	35.100	20.740	.51		(.50,.75)	35.100	20.740	.76

reverse J-shaped when $(\theta_L, \theta_U) = (0, .25), (0, .50), (0, .75)$; and unimodal when $(\theta_L, \theta_U) = (.25, .75), (.25, .50), (.50, .75)$.

For the choices of θ_L, θ_U , and γ considered in Table 12, a comparison was made between $B(\alpha_F, \beta_F)$ and the most conservative prior distributions from the other families discussed in Section 3. The results are quite similar to those obtained in the previous section and will not be given. Generally speaking, when $N \geq 20$, additional Bernoulli trials are required under the most conservative prior distributions from the other families to gain, on the average, a comparable amount of information about θ as is provided by the original sample relative to $B(\alpha_F, \beta_F)$.

7. NUMERICAL EXAMPLE

Consider a series system consisting of three different

independently operating pass/fail components, where a pass means surviving for a fixed number of cycles or to a fixed time t . System reliability is given by $R = \prod_{i=1}^3 R_i$, where R_i is the reliability of the i^{th} component, $i = 1, 2, 3$. A life test of the i^{th} component consists of N_i identical units tested with x_i successes (passes) recorded. Assuming a beta prior distribution, $B(\alpha_i, \beta_i)$, for the reliability of the i^{th} component, the posterior distribution of R_i is $B(x_i + \alpha_i, N_i - x_i + \beta_i)$. By using inverse Mellin transforms, Springer (1979, p. 104) has derived an expression for the pdf of the product of beta random variables. By using this result, the posterior pdf of system reliability is

$$g(R|x_1, x_2, x_3) = \left[\prod_{i=1}^3 \frac{\Gamma(N_i + \alpha_i + \beta_i)}{\Gamma(x_i + \alpha_i)} \right] G_{3,3}^{3,0} \left(R \middle| \begin{matrix} N_1 + \alpha_1 + \beta_1 - 1, N_2 + \alpha_2 + \beta_2 - 1, N_3 + \alpha_3 + \beta_3 - 1 \\ x_1 + \alpha_1 - 1, x_2 + \alpha_2 - 1, x_3 + \alpha_3 - 1 \end{matrix} \right),$$

$$0 < R < 1 \quad (7.1)$$

where $G_{p,q}^{m,n}(\cdot)$ is the Meijer G-function (see Erdélyi 1953, pp. 206-222). When the parameters in the $B(\alpha_i, \beta_i)$ prior distributions are positive integers (as is the case for uniform prior distributions), a finite series representation for $g(R|x_1, x_2, x_3)$ may be obtained (Springer 1979, p. 105; Springer and Thompson 1966). Under the most conservative beta prior distributions given in the previous sections, it is usually the case that no two of $x_i + \alpha_i - 1$, $i = 1, 2, 3$, differ by an integer. When this condition holds, the expression for $g(R|x_1, x_2, x_3)$ may be written in terms of the generalized hypergeometric function

$${}_3F_2(\alpha, \beta, \gamma; a, b; x) = \sum_{r=0}^{\infty} \frac{(\alpha)_r (\beta)_r (\gamma)_r}{(a)_r (b)_r r!} x^r, \quad (7.2)$$

where $(\delta)_0 = 1$, $(\delta)_r = \delta(\delta+1)\dots(\delta+r-1)$ when $r \geq 1$. The series in (7.2) is known to be convergent if $|x| < 1$. Specifically, we

find that upon simplification

$$g(R|x_1, x_2, x_3) = \sum_{i=1}^3 \frac{\prod_{j=1}^3 \Gamma(x_j - x_i + \alpha_j - \alpha_i)}{\prod_{j=1}^3 \Gamma(N_j - x_i + \beta_j + \alpha_j - \alpha_i)} R^{x_i + \alpha_i - 1} {}_3F_2[-N_1 + x_1 + 1, -\alpha_1 - \alpha_1 - \beta_1, -N_2 + x_1 + 1 + \alpha_1 - \alpha_2 - \beta_2, -N_3 + x_1 + 1 + \alpha_1 - \alpha_3 - \beta_3; (-x_1 + x_1 + 1 + \alpha_1 - \alpha_1)^*, (-x_2 + x_1 + 1 + \alpha_1 - \alpha_2)^*, (-x_3 + x_1 + 1 + \alpha_1 - \alpha_3)^*; R], \quad 0 < R < 1 \quad (7.3)$$

where the asterisk indicates the omission of the i^{th} denominator-parameter, that is, $(-x_1 + x_1 + 1 + \alpha_1 - \alpha_1)$. The posterior cdf of system reliability is

$$G(R|x_1, x_2, x_3) = \int_0^R g(t|x_1, x_2, x_3) dt, \quad (7.4)$$

which may be written in series form by using the result

$$\begin{aligned} \int_0^R t^u {}_3F_2(\alpha, \beta, \gamma; a, b; t) dt \\ = \frac{1}{u+1} R^{u+1} + \frac{\alpha\beta\gamma}{ab(u+2)} R^{u+2} + \frac{\alpha(\alpha+1)\beta(\beta+1)\gamma(\gamma+1)}{a(a+1)b(b+1)(u+3)2!} R^{u+3} + \dots \\ + \frac{\alpha(\alpha+1)\dots(\alpha+n)\beta(\beta+1)\dots(\beta+n)\gamma(\gamma+1)\dots(\gamma+n)}{a(a+1)\dots(a+n)b(b+1)\dots(b+n)(u+n+2)(n+1)!} R^{u+n+2} + \dots \end{aligned} \quad (7.5)$$

A double precision computer program was written in FORTRAN to evaluate the expression for $G(R|x_1, x_2, x_3)$ given by (7.4)-(7.5). The gamma functions were computed by using the IMSL routine DGAMMA. The series in (7.5) was truncated when both $|y_i - y_{i+1}| < 10^{-10}$ and $|y_{i+1}| < 10^{-10}$, where y_i is the i^{th} term. By utilizing the program to evaluate $G(R|x_1, x_2, x_3)$ in conjunction with the IMSL routine ZBRENT, the solution for R in the equation

$G(R|x_1, x_2, x_3) = n$, where n is specified, may be determined. Consequently, by solving the equation $G(R_L|x_1, x_2, x_3) = (1-n)/2$ for R_L and the equation $G(R_U|x_1, x_2, x_3) = (1+n)/2$ for R_U , we may obtain a symmetric 100n percent two-sided Bayesian probability interval (TBPI) for system reliability, i.e., $R_L < R < R_U$.

Suppose life tests on the three components yielded the following results (the data are from Springer and Thompson 1966):

$$N_1 = 10, x_1 = 8; N_2 = 9, x_2 = 7; \text{ and } N_3 = 4, x_3 = 3.$$

When there are no available prior estimates of component reliabilities, we find from Table 6 the parameters of the most conservative beta prior distributions to be $(\alpha_1, \beta_1) = (.211, .211)$, $(\alpha_2, \beta_2) = (.201, .201)$, and $(\alpha_3, \beta_3) = (.118, .118)$. By solving the equations $G(R_L|8, 7, 3) = .05$ and $G(R_U|8, 7, 3) = .95$ for R_L and R_U , respectively, we get a symmetric 90 percent TBPI for system reliability to be $.187 < R < .707$. By comparison, under uniform prior distributions on component reliabilities we get $.157 < R < .594$. The shorter length of the latter interval is due to more a priori information (but, in this case, extraneous information) about the R_i 's being introduced through the uniform prior distributions than through the most conservative prior distributions.

Now suppose there are available the following prior estimates of component reliabilities: $E(R_1) = .9$, $E(R_2) = .7$, and $\text{Prob}(.5 < R_3 < 1) = .95$. From Tables 9 and 12 we find the parameters of the most conservative beta prior distributions to be $(\alpha_1, \beta_1) = (.888, .099)$, $(\alpha_2, \beta_2) = (.323, .138)$, and $(\alpha_3, \beta_3) = (1.719, .177)$. A symmetric 90 percent TBPI for system reliability is $.257 < R < .742$.

ACKNOWLEDGEMENTS

A portion of the first author's research was financially supported by Los Alamos National Laboratory through its project R665, "Bayesian Techniques in Risk Analysis," from the Nuclear

Regulatory Commission. In addition, several discussions with Drs. Harry F. Martz, Maurice C. Bryson, and Richard J. Beckman in the Statistics Group at Los Alamos National Laboratory were very helpful as well as encouraging and are greatly appreciated. All computer programming was done on an IBM 4341 at The University of Texas at Arlington.

BIBLIOGRAPHY

- Akaike, Hirotugu (1980). Likelihood and the Bayes procedure. In *Bayesian Statistics, Proceedings of the First International Meeting held in Valencia (Spain)*, (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds.), 143-203. Valencia, Spain: University Press.
- Barnett, Vic (1973). *Comparative Statistical Inference*. New York: John Wiley.
- Berger, James O. (1980). *Statistical Decision Theory*. New York: Springer-Verlag.
- Bernardo, Jose M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society, Series B*, 41, 113-147.
- Box, George E. P., and Tiao, George C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, Mass.: Addison-Wesley.
- Brender, David M. (1968). The prediction and measurement of system availability: a Bayesian treatment. *IEEE Transactions on Reliability*, 17, 127-138.
- Chiu, W. K. (1974). A new prior distribution for attributes sampling. *Technometrics*, 16, 93-102.
- Erdélyi, A. (1953). *Higher Transcendental Functions (Volume I)*. New York: McGraw-Hill.
- Gradshteyn, I. S., and Ryzhik, I. M. (1965). *Tables of Integrals, Series, and Products*. New York: Academic Press.
- Grassia, A. (1977). On a family of distributions with argument between 0 and 1 obtained by transformation of the gamma and derived compound distributions. *Australian Journal of Statistics*, 19, 108-114.
- Griffiths, David, and Schafer, Christine (1981). Closeness of Grassia's transformed gammas and the beta distribution. *Australian Journal of Statistics*, 23, 240-246.

- Hahn, Gerald J., and Shapiro, Samuel S. (1967). *Statistical Models in Engineering*. New York: John Wiley.
- Jaynes, Edwin T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, 4, 227-241.
- Jeffreys, Harold (1961). *Theory of Probability* (3rd edition). Oxford: Clarendon Press.
- Johnson, Norman L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36, 149-176.
- Johnson, Norman L., and Kotz, Samuel (1970). *Distributions in Statistics: Continuous Univariate Distributions-2*. New York: John Wiley.
- Lindley, Dennis V. (1956). On a measure of the information provided by an experiment. *Annals of Mathematical Statistics*, 27, 986-1005.
- Martz, Harry F., and Waller, Ray A. (1982). *Bayesian Reliability Analysis*. New York: John Wiley.
- Rényi, Alfred (1976). *Probability Theory*. Amsterdam: North-Holland.
- Springer, M. D. (1979). *The Algebra of Random Variables*. New York: John Wiley.
- Springer, M. D., and Thompson, W. E. (1966). Bayesian confidence limits for the product of N binomial parameters. *Biometrika*, 53, 611-613.
- Stroud, A. H., and Secrest, D. H. (1966). *Gaussian Quadrature Formulas*. Englewood Cliffs, N. J.: Prentice-Hall.
- Villegas, C. (1977). On the representation of ignorance. *Journal of the American Statistical Association*, 72, 651-654.
- Weiler, H. (1965). The use of incomplete beta functions for prior distributions in binomial sampling. *Technometrics*, 7, 335-347.
- Zellner, Arnold (1977). Maximal data information prior distributions. In *New Developments in the Applications of Bayesian Methods*, (A. Aykac and C. Brumat, eds.), 211-232. Amsterdam: North-Holland.

REFERENCES

- Abramowitz, M. and Stegun, I. A. (1970), Handbook of Mathematical Functions. Dover, New York.
- Johnson, N. L. and Kotz, S. (1970), Continuous Univariate Distributions-2. Houghton Mifflin, Boston.
- Saxena, K. M. L. and Alam, K. (1982) "Estimation of the noncentrality parameter of a chi-squared distribution," Ann. Statist. 10, pp. 1012-1016.