A BAYESIAN APPROACH TO ANALYSIS OF LIMIT STANDARDS

Roy R. Creasey, Jr.

College of Business and Economics
201 High Street
Longwood University
Farmville, VA 23909, U.S.A.

K. Preston White, Jr.

Department of Systems and Information Engineering
151 Engineer's Way
University of Virginia
Charlottesville, VA 22904, U.S.A.

ABSTRACT

Limit standards are probabilistic requirements or benchmarks regarding the proportion of replications conforming or not conforming to a desired threshold. Sample proportions resulting from the analysis of replications are known to be beta distributed. As a result, standard constructs for defining a confidence interval on such a proportion, based on critical points from the normal or Student's t distribution, are increasingly inaccurate as the mean sample proportion approaches the limits of 0 or 1. We consider the Bayesian relationship between the beta and binomial distributions as the foundation for a sequential methodology in the analysis of limit standards. The benefits of using the beta distribution methodology are variance reduction, and smaller sample size (when compared to other analysis methodologies).

1   INTRODUCTION

In our continuing investigation into the sequential methodologies for analysis of limit standards, we were intrigued by the analysis and findings of Raatikainen (1995). He observed that, as the distribution of sample proportions moves closer to the limits of 0 or 1, the distribution becomes increasingly skewed. Therefore, the results obtained from use of the current confidence interval methodology, or

$$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

where $\hat{p}$ is the point estimator for the population (or system) proportion ρ and the sample size n is "sufficiently large" (Law and Kelton, 2000), becomes increasingly inaccurate as $\hat{p} \to 0$ or $\hat{p} \to 1$.

Raatikainen (1995) utilized an arcsin transformation methodology of the sample proportions, which provides a near Normal distribution at smaller or larger proportions

and has the added benefit of variance reduction, resulting in a tighter confidence interval of the population proportion. However, Raatikainen acknowledges this approach still results in skewed distributions at very small or very large proportions.

In the study of limit standards, we are particularly interested in the distribution of the sample proportions, because one objective is to determine if the population proportion meets the proportion standard. Furthermore, typical limit standards contain either very small or very large proportions. If any information can be obtained about the proportion distribution prior to taking an observation, fewer observations may be necessary to provide a statistically valid comparison of the system to the standard using the *posterior distribution*, or the resulting distribution after considering the observations.

Consider an M/M/1 queuing system with traffic intensity (utilization) of 0.8. Figure 1 depicts two distributions of sample proportions taken from 400 replications, where n = 24000 observations per replication. The distribution marked "A" depicts the proportion of observations with a queue length greater than 15 minutes. Next consider the distribution marked "B", which depicts, from the same 400 replications, the proportion of observations with a queue length greater than 20 minutes. Notice how the skewness of the distribution increases as the distribution approaches 0. The model and findings were consistent with those of Raatikainen (1995). Upon further review of Raatikainen's (1995) work, and through study of Bayesian statistics, the distribution of sample proportions follows a beta distribution when the output of the observations are considered random variables (r.v.) from a binomial distribution. (Lee, 1997)

This paper considers the relationship between the beta distribution and the binomial, and the application of Bayesian statistics in the analysis of limit standards.

Although the concept of Bayesian statistics in discrete event simulation (DES) is not old, interest has been renewed in recent years. Chick (2006) provides an excellent tutorial of its potential use. Several papers by Chick

(1997, 2001, 2006) and Chick and Inoue (1998) focused on the use of Bayesian statistics for input analysis and ranking and selection procedures.

Law and Kelton's (2000) methodology for DES output analysis of proportions centers on the confidence interval technique, with the assumption the distribution of sample proportions is Normal. Chen and Kelton (1999, 2004A, 2004B) have investigated the use of histograms and quantiles to determine confidence intervals and point estimators for proportions.
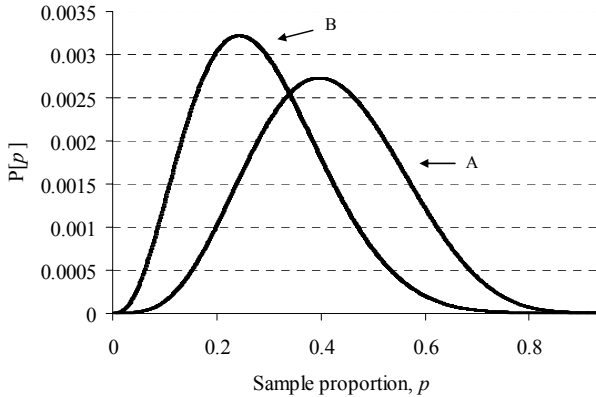

Figure 1. Distribution of proportions from M/M/1 system with traffic intensity of 0.8.

2 BACKGROUND ON LIMIT STANDARDS

To aid in model development and to provide a statistically valid analysis, customer-specified standards should be defined in probabilistic terms. One such type of standard is known as a *limit standard*, which is a maximum or minimum threshold the customer would not enjoy exceeding. (Creasey, *et al.* 2005; Creasey and White 2006, 2007) The typical limit standard is a proportional requirement, but differs from a confidence interval requirement in that the analyst is not interested in determining a point estimate and confidence interval for the population proportion, but simply if the population proportion meets some minimum or maximum proportion limit with a specified degree of confidence. In other words, confidence intervals are allowed to change as more observations are analyzed, but with limit standards the intervals are *fixed* and the degree of confidence changes as more observations are analyzed.

For example, a requirement from NASA's Constellation Program states "The Constellation Architecture shall deliver at least 20,000 kg (44,092 lbm) of cargo to the lunar surface for Lunar Outpost missions. The verification shall be considered successful when the analysis shows that the calculated Mass Delivered capability of the Constellation Architecture is equal to or greater than the Mass

Delivered requirement for 99.73% of the simulations with a 90% confidence." (White, *et al.* 2007)

Based on the definition of limit standards provided in Section 2, typical analysis of limit standards within DES models requires representation of the output observations as a set of Bernoulli random variables. Subsequently, the probability distribution of x successes out n observations with probability of success p exhibits a binomial density function.

To date, we have investigated and applied two methodologies used in quality control for acceptance sampling based on attributes. White, *et al.* (2007) and Creasey, *et al.* (2005) describe the use of binomial equations to predetermine the necessary sample size, n, and maximum number nonconforming, c, to meet customer-specified Type I (defined as α) and Type II (defined as β) risk. Creasey and White (2006) discuss the use of Wald's sequential probability ratio test (SPRT), which performs very well when the output result from a sample is either very good (very few observations exceed the standard) or very bad (many observations exceed the standard). In cases where the sample proportion exceeding is near the standard, the ability of the SPRT methodology to discern good or bad systems diminishes.

Define ω as a *system*, or a unique set of input, process and output distributions. Let ω_i denote the ith system of k systems under investigation, where $i = 0, 1, \ldots, k$. Let ω_0 denote the standard system. The other systems will be referred to as the *alternative* systems, or systems under investigation. Therefore, for a single system comparison to a standard, there are two systems: ω_0 and ω_1. Each alternative system contains a series of n observations of an output, which are defined as a set of random variables $X_i = \{X_{ij}, j = 1, 2, \ldots, n\}$. Each observation is assumed to be independent and identically distributed (i.i.d.). For the standard system, there is only one observation of interest, which is the limit standard. Let X_{max} and X_{min} be constants from system ω_0, defined as the upper and lower specification limits on X_{ij}, respectively, where $X_{max} \geq X_{min}$. It is, therefore, desired but not required that $X_{max} \geq X_{ij} \geq X_{min}$, for $\forall j$, and $\forall i > 0$.

Within each alternative system i, let $U_i \subseteq X_i$ be the set of all output observations that exceed the upper limit ($X_i > X_{max}$) and let $L_i \subseteq X_i$ be the set of all output observations that fall short of the lower limit ($X_i < X_{min}$). Let $|U_i|$ be the cardinality of U_i and $|L_i|$ be the cardinality of L_i, where $|U_i| + |L_i| \leq n$. Define the constants

$$\rho_{iU} = \lim_{n \to \infty} \frac{|U_i|}{n}$$

$$\rho_{iL} = \lim_{n \to \infty} \frac{|L_i|}{n}$$

as the limiting *proportions* from system ω_i (for $i > 0$) of the number of observations above and below the upper

and lower limits, respectively. Let the constants $\rho_{\max}$ and $\rho_{\min}$ be the desired upper and lower performance bounds from system ω_0 on ρ_{iU} and ρ_{iL}, respectively. We define a *limit standard* for the output as the septuple ($X_{\max}$, $X_{\min}$, $\rho_{\max}$, $\rho_{\min}$, ρ_2, α, β) from system ω_0, where ρ_2 is as defined below, α is the acceptable estimation-error probability for $\rho_{\max} + \rho_{\min}$ and β is the acceptable estimation-error probability for ρ_2.

From this definition, three types of limit standards are distinguished.

Type 1: Single-sided standard. This type can be a maximum standard, where $X_{\min} = -\infty$ (implying $\rho_{\min} = 0$), or a minimum standard, where $X_{\max} = \infty$ (implying $\rho_{\max} = 0$). A special case is where an absolute boundary has been established for $X_{\max}$ or $X_{\min}$. For example, in systems where elapsed time is the observation of interest, $X_{\min} = 0$ and $\rho_{\min} = 0$. The problem is to determine ρ_{iU} or ρ_{iL}, based on a sample of n observations and test the null hypothesis that

H_0: $\rho_{iU} \leq \rho_{\max}$ or $\rho_{iL} \leq \rho_{\min}$ with
P(accepting H_0 | H_0 true) $\geq 1 - \alpha$.

In addition, the customer or analyst may wish to impose a second hypothesis of
H_1: (ρ_{iU} or ρ_{iL}) $\geq \rho_2$ with
P(accepting H_1 | H_1 true) $\geq 1 - \beta$,

where ($\rho_{\max}$ or $\rho_{\min}$) $< \rho_2$.

Type 2: Double-sided standard with a single proportion limit. Define p_i as the total proportion of observations in system ω_i outside the limits, or $p_i = \rho_{iU} + \rho_{iL}$, for $i = 1, 2, \ldots, k$. Also, define ρ_0 as the total proportion of observations in system ω_0 outside the limits, or $\rho_0 = \rho_{\max} + \rho_{\min}$. This is the more common of the two types of double-sided standards, because the values of $\rho_{\max}$ and $\rho_{\min}$ are secondary to ρ_0. In other words, the customer need only decide upon ρ_0 and not the apportionment of ρ_0 to $\rho_{\max}$ and $\rho_{\min}$. The problem is to determine p_i, based on a sample of n observations and test the null hypothesis that

H_0: $p_i \leq \rho_0$ with P(accepting H_0 | H_0 true) $\geq 1 - \alpha$.

In addition, the customer or analyst may wish to impose a second hypothesis of
H_1: $p_i \geq \rho_2$ with P(accepting H_1 | H_1 true) $\geq 1 - \beta$,

where $\rho_0 < \rho_2$.

Type 3: Double-sided standard with two proportion limits. With this type of double-sided standard, the customer must define values for $\rho_{\max}$ and $\rho_{\min}$. For example,

in a bread production system, the customer may wish to limit the proportion of breads receiving less than the minimum necessary baking time and limit the proportion receiving more than the maximum necessary baking time. The problem is to determine ρ_{iU} and ρ_{iL}, based on a sample of n observations and test the null hypothesis that

H_0: $\rho_{iU} \leq \rho_{\max}$ and $\rho_{iL} \leq \rho_{\min}$ with
P(accepting H_0 | H_0 true) $\geq 1 - \alpha$.

This paper focuses on only one alternative system, ω_1, and only on Type I and II standards. A Type I or II limit standard can also be defined as the expected value of an indicator variable. In other words,

$E[I] = 1/j * \Sigma I_j$,
$E[I] \to p$ as $n \to \infty$, where
$$I_j = \begin{cases} 1, & X_j < X_{\min} \text{ or } X_{\max} < X_j \\ 0, & X_{\min} \leq X_j \leq X_{\max} \end{cases}.$$
$j = 1, 2, \ldots, n$,
H_0: $p \in [0, \rho_{\max} + \rho_{\min}]$ with
P(accepting H_0 | H_0 true) $\geq 1 - \alpha$, and
H_1: $p > \rho_2$ with P(accepting H_1 | H_1 true) $\geq 1 - \beta$ (if defined by the customer).

Accordingly, we can treat the 1/0 outcome of a series of observations as a set of Bernoulli trials. The P($X = x$) can be defined by the binomial distribution function, with probability density $bi(x; p, n)$, as

$$bi(x; p, n) = \binom{n}{x} p^x (1-p)^{n-x}. \qquad (1)$$

and cumulative distribution function $Bi(x; p, n)$ of

$$Bi(x; p, n) = \sum_{j=0}^{n} bi(j; p, n) = \sum_{j=0}^{x} \binom{n}{j} p^j (1-p)^{n-j} . \quad (2)$$

(Wasserman 2002)

The above definition of limit standards does not fit a Bayesian approach, because it considers both observations that were actually taken and evaluated and observations that could have been taken but were not. The Bayesian approach would simply find the two posterior probabilities

p_a = P($0 < p \leq \rho_{\max} + \rho_{\min}$| X)
p_b = P($p > \rho_2$| X)

and decide between the two hypotheses. Therefore, α is defined not as an error probability but simply $\alpha = 1 - p_a$. Similarly, β is defined as $\beta = 1 - p_b$. (Lee 1997)

3   BETA DISTRIBUTION

The Beta distribution, $b(\gamma, \xi)$ is generally defined as the distribution of $u/(u + w)$, where u is a r.v. from a χ^2_{v1} distribution and w is a r.v. from a χ^2_{v2} distribution, and u and w are independent. The ratio exhibits a Beta distribution with parameters $\gamma = v_1/2$ and $\xi = v_2/2$. (Johnson, *et al*, 1994) Specifically, the beta distribution has a density $be(t; \gamma, \xi)$ of

$$be(t; \gamma, \xi) = \frac{1}{b(\gamma, \xi)} t^{\gamma-1}(1-t)^{\xi-1}, \qquad (3)$$

where $be(t; \gamma, \xi) = 0$ for t not in the interval [0, 1],

$$b(\gamma, \xi) = \frac{\Gamma(\gamma)\Gamma(\xi)}{\Gamma(\gamma+\xi)} = \int_{t=0}^{1} t^{\gamma-1}(1-t)^{\xi-1} dt \qquad (4)$$

is the Beta function, $\Gamma(\bullet)$ is the Gamma function and γ, ξ > 0. The cumulative distribution function, $Be(t; \gamma, \xi)$ is defined as

$$Be(t; \gamma, \xi) = \int_{m=0}^{t} \frac{1}{b(\gamma, \xi)} m^{\gamma-1}(1-m)^{\xi-1}$$

$$Be(t; \gamma, \xi) = \frac{1}{b(\gamma, \xi)} \int_{m=0}^{t} m^{\gamma-1}(1-m)^{\xi-1} \qquad (5)$$

where the right hand integral $\int_{m=0}^{t} m^{\gamma-1}(1-m)^{\xi-1}$ is known as the *incomplete* or *partial* Beta distribution. (Lee, 1997)

The Beta distribution has the following properties. [Lee (1997) and Law and Kelton (2000)]

- $E(t) = \dfrac{\gamma}{\gamma + \xi}.$ \qquad (6)

- $Var(t) = \dfrac{\gamma\xi}{(\gamma + \xi)^2(\gamma + \xi + 1)}.$

- $be(t; 1, 1)$ is the Uniform distribution $U[0, 1]$.
- The distribution is limited to the interval [0, 1].
- If $\gamma = 1$ or $\xi = 1$, the distribution is strictly decreasing or increasing.
- If $\gamma > 1$ and $\xi > 1$, the distribution is unimodal.
- If $\gamma = \xi$, the distribution is symmetrical.
- When $\gamma + \xi = 1$, the distribution is an arcsin distribution.

4   RELATIONSHIP BETWEEN BETA AND BINOMIAL

One notices the similarity between Equations (1), (2), (3) and (4). A mathematical relationship exists between these equations. Specifically, Equation (4) is contained in the marginal distribution of Equation (1), as follows.

Assume the distribution of the proportion p is uniform between 0 and 1, or $f(p) = U[0, 1]$ The marginal distribution $f_x(x, p)$ of $bi(x; p, n)$ is

$$f_x(x, p) = \int_{p=0}^{1} f(x \mid p) dp$$

$$= \int_{p=0}^{1} \binom{n}{x} p^x (1-p)^{n-x} dp \qquad (7)$$

$$= \binom{n}{x} \int_{p=0}^{1} p^x (1-p)^{n-x} dp.$$

The integral part of Equation (7) constitutes the Beta function with parameters $b(x + 1, n - x + 1)$. In fact, any prior distribution of p, conditioned on the binomial density function, results in a posterior, or marginal, density function containing the Beta distribution. Consequently, if the conditioning function is a binomial density function, the prior density distribution of p is a Beta distribution. (Lee 1997 and Waterman, *et al*. 1976)

5   THE BAYESIAN RELATIONSHIP

In typical (or Frequentist) output analysis using the binomial distribution, the value of p is estimated or, in the case of limit standards, the maximum acceptable proportion p_0 is used. From this proportion, one can determine the minimum sample size necessary to make a statistically valid decision regarding the system (i.e., compliance to the standard). However, there can be considerable uncertainty in the sample proportion.

In Bayesian statistics analysis, the parameters of a conditioning function are treated as random variables. The analyst then attempts to specify the distribution of these parameters, known as the *prior distribution*. The prior distributions are then used in the conditioning function, resulting in the posterior distribution.

As Chick (2006) points out, one of the challenges with the application of Bayesian statistics to simulation experiments is in specifying these prior distributions. The elicitation methods used for determining these distributions can be challenging and subject to the desires of the customer. (Gavasakar, 1988)

Fortunately, for a binomial conditioning function the prior distribution is known to be a beta distribution. (Lee, 1997) Recall from Section 3 that one of the properties of the beta distribution is $be(t; 1, 1) = U[0, 1]$. If we consider the general $be(t; \gamma, \xi)$ distribution as the prior distribution of p, the resulting posterior distribution for p is

$$f(p \mid x) = \frac{1}{b(\gamma + x, \xi + n - x)} p^{\gamma+x-1}(1-p)^{\xi+n-x-1}, \qquad (8)$$

with expected value and variance of

$$E(p) = \frac{\gamma + x}{\gamma + \xi + n} \qquad (9)$$

$$Var(p) = \frac{(\gamma + x)(\xi + n - x)}{(\gamma + \xi + n)^2(\gamma + \xi + n + 1)}. \qquad (10)$$

(Lee, 1997)

Therefore, specifying the prior distribution reduces to selection of γ and ξ. Also (and this is generally true with Bayesian analysis), we enjoy a recursive nature, in that the posterior distribution becomes the prior distribution for consideration of the next observation. In the case where the observations exhibit a binomial distribution, the beta-distributed proportion is updated as we consider successive observations.

6 APPLICATION TO LIMIT STANDARD ANALYSIS

The most difficult task in Bayesian analysis is in development of the initial prior density distribution for p. If the initial prior does not adequately represent the preconceived assumptions of the customer, or is informative, then the use of the distribution could result in a bias of the final posterior. Correcting this bias usually requires reconsideration of the initial prior or evaluation of substantially more observations.

In general, the problem is one of input model uncertainty. Proper development of an input model requires knowledge of the distribution and the parameters associated with that distribution. (Henderson 2003) For the problem at hand, we know the distribution of the initial prior is Beta, but we do not know the initial values of γ and ξ, which we define as γ' and ξ'.

If we desire to use a Bayesian approach for limit standard analysis, determination of γ' and ξ' is not straightforward. Typical elicitation methods involving the Beta distribution requires the customer to define at least two points in the distribution, such as the expected value and first quartile or expected value and 95th percentile. However, the limit standard definition from above only provides one point of reference, or p_0 at the $1 - \alpha$ percentile.

Therefore, we must elicit from the customer additional information regarding the desired Beta distribution. To do so, we make two assumptions. First, we assume the customer desires the shape of the prior density distribution to resemble the shape of the final posterior distribution. (Wasserman 2002) Since the final posterior is typically unimodal, we assume the customer wants a unimodal distribution for the initial prior, which requires $\gamma' > 1$ and $\xi' > 1$.

Second, we believe the customer is interested in finding values of γ' and ξ' such that the initial prior dis-

tribution also meets the limit standard. In other words, we want to find γ' and ξ' such that, with a Beta distribution, $P[X \leq p_0] = 1 - \alpha.$, or

$$1 - \alpha = Be(p_0; \gamma', \xi') = \int_0^{p_0} \frac{\Gamma(\gamma' + \xi')}{\Gamma(\gamma')\Gamma(\xi')} p^{\gamma'-1}(1-p)^{\xi'-1} dp, (11)$$

where $\gamma' > 1$ and $\xi' > 1$. Since, for limit standards, a near infinite set of solutions exist for Equation (11), further elicitation of preferences from the customer is required. The remainder of this section explores an elicitation methodology based on p_0 and α.

Upon review of Equation (6), we observe that the expected value of the Beta distribution is a pseudo-proportion containing γ pseudo-failures out of $(\gamma + \xi)$ pseudo-observations. Now, consider Figure 2, which depicts three Beta distributions, each with $E(p) = 0.01$. The distribution with the shortest peak has $\gamma = 2$ (for $be(p; 2, 198)$), the next shortest peak $\gamma = 3$ (for $be(p; 3, 297)$), and the tallest peak has $\gamma = 4$ (for $be(p; 4, 396)$). Therefore, as the value of γ increases, the mode of the distribution gets closer to $E(p)$. In addition, small values of γ indicates less certainty about the distribution and large values of γ indicates greater certainty about the distribution. (Haigh 2002) Subsequently, the variance of the Beta distribution (Equation (7)) is a decreasing function of $(\gamma + \xi)$, with $0 < Var(t) < E(t)[1 - E(t)] = \gamma\xi/(\gamma + \xi)^2$. Therefore, as $(\gamma + \xi) \to \infty$, the mass becomes concentrated at $E(t)$. (Waterman, *et al.* 1976) Consequently, the larger the variance, the less informative the prior distribution will be on the posterior.
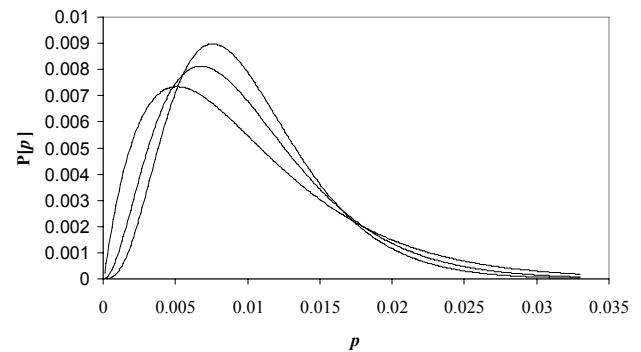
Figure 2. Three beta distributions of $b(2, 198)$ (leftmost), $b(3, 297)$ (center), $b(4, 396)$ (rightmost), all with $E(p) = 0.01$.

As stated in Section 2, the customer desires $p_i \leq p_0$. It is also desirable for the expected value of the initial prior, $E(p')$, to be less than p_0, or $E(p') = \gamma'/(\gamma' + \xi') < p_0$. We therefore must elicit from the customer how much less should $E(p')$ be compared to p_0. To assist in this elicitation, we define τ as the degree of separation between $E(p')$ and p_0, or $\tau = p_0/E(p')$. Since we are only concerned with small values of p_0 and $E(p')$, a separation of,

say, $\tau = 2$ means p_0 is twice as large as $E(p')$. Table 1 contains minimum values of γ' and ξ' for various values of $E(p')$, p_0 (with corresponding τ), and α. The values of γ' and ξ' were obtained by first selecting a value for ($\gamma' + \xi'$) such that the right side of Equation (10) was less than $1 - \alpha$, finding the corresponding values for γ' and ξ' using Equation (6), then increasing ($\gamma' + \xi'$) until Equation (11) was satisfied. This procedure can be expressed as an optimization problem:

Given: p_0, τ, α

Objective: $\min[\gamma' + \xi']$ (12)

Subject to: $\dfrac{p_0}{\tau} = \dfrac{\gamma'}{(\gamma' + \xi')}$,   (13)

$$1 - \alpha = \int_0^{p_0} \frac{\Gamma(\gamma' + \xi')}{\Gamma(\gamma')\Gamma(\xi')} p^{\gamma'-1}(1-p)^{\xi'-1} dp,$$

$\gamma', \xi' > 1.$

Upon review of Table 1, we observe two items of interest. First, as $\tau \to 1$ or $\tau \to \infty$, γ' and/or ξ' is significantly larger, with the later condition occurring because we desire $\gamma' > 1$, which forces $\xi' \gg 1$. Second, looking at the values of γ' for a specific τ and α, it appears that a linear relationship exists between $\min[\gamma']$ and p_0. Figure 3 depicts this relationship for specific values of α and for $\tau = 2$ within a range of p_0 of [0.005, 0.1].

Determination of a mathematical solution to the linear relationship using Equations (11), (12) and (13) is challenging. Not only is the goal to find the $\min[\gamma' + \xi']$ which satisfies the equations, we also require $\gamma', \xi' > 1$. As τ or α increases, values of γ' get closer to 1, and linearity is lost. For the three relationships in Figure 3, the approximate linear relationships are:

$\alpha = 0.01$: $\gamma' \approx 8.0004 - 6.0063 p_0$;
$\alpha = 0.02$: $\gamma' \approx 6.0473 - 4.4018 p_0$;
$\alpha = 0.05$: $\gamma' \approx 3.5618 - 2.3508 p_0$.

Therefore, in order to develop an initial prior distribution for a Bayesian sequential analysis of limit standards, the customer only needs to provide the desired value of τ. We recommend using $\tau = 2$, and do so for two reasons. First, since α should be small, the result is a slightly positively-skewed distribution, which mimics our findings and those of Raatikainen (1995) for small proportions. [A near-symmetrical distribution can be obtained by establishing a τ slightly less than 2.] Second, as observed in the above regression equations, for small α it is not possible for γ' to fall below 1 within the range $p_0 \in [0, 1]$.
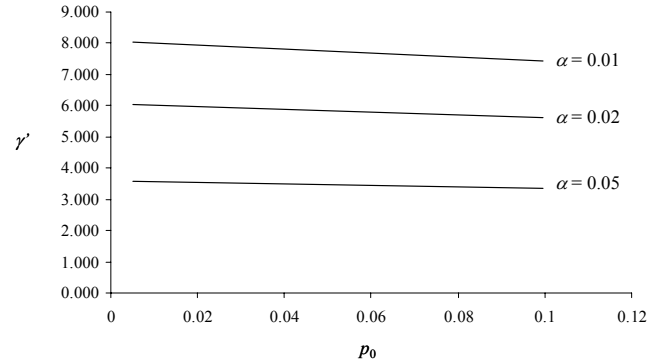

Figure 3. Relationship between γ' and p_0 at $\tau = 2$ for values of p_0 between [0.005, 0.1]

7   POSTERIOR DISTRIBUTION EVALUATION

As stated in Section 1, the posterior density distribution of p of a binomial conditioning function exhibits a beta distribution. The parameters of this distribution are $be(p; \gamma' + x_n, \xi' + (n - x_n))$, where x_n is defined as the number of failed observations after assessment (conditioning) of n observations, or $x_n = U_i + L_i$ (from Section 2). (Lee 1997)

In the sequential analysis of simulation output, the analyst must determine an appropriate sample size of observations. The goal is to determine this sample size prior to conducting experiments; in situations where conducting the experiment impacts cost or schedule, knowledge of the necessary sample size allows for proper planning and budgeting of the experiment. The sample size should be small enough to minimize cost and schedule impacts but large enough to minimize the need to re-conduct the experiment to obtain additional observations. Although we would normally develop a Bayesian informative stopping rule, for our initial investigation we determined the initial sample size n_0 using a desired half-width error, from Cochran (1963),

$$n_0 = \left\lceil \frac{z_{1-\alpha}^2 p_0(1 - p_0)}{\varepsilon^2} \right\rceil,   (14)$$

where $\lceil \bullet \rceil$ indicates rounding up to the next integer, z is the position on the standard normal distribution corresponding to a cumulative distribution of $1 - \alpha$, and ε is the desired upper half-width of the estimated proportion. We took the conservative approach of using p_0, which results in a larger sample size than using $E(p')$ while usually meeting the desired upper half-width.

After obtaining and analyzing n_0 observations for each replication, we determined the across-run average posterior values for γ and ξ, defined as γ'' and ξ'', using

Table 1. Minimum Initial Prior Parameter Values for Given $E(p')$, p_0, and α.

p_0	$E(p')$	τ	$\alpha = 0.05$		$\alpha = 0.02$		$\alpha = 0.01$	
			γ'	ξ'	γ'	ξ'	γ'	ξ'
0.10	0.09	1.111	207.495	2098.005	326.295	3299.205	420.075	4247.425
0.10	0.08	1.250	43.336	498.364	68.880	792.120	89.048	1024.052
0.10	0.07	1.429	15.687	208.413	25.277	335.823	32.837	436.263
0.10	0.06	1.667	6.942	108.758	11.394	178.506	14.904	233.496
0.10	0.05	2.000	3.330	63.270	5.610	106.590	7.400	140.600
0.08	0.07	1.143	129.815	1724.685	204.694	2719.506	263.816	3504.984
0.08	0.06	1.333	25.584	400.816	40.950	641.550	53.076	831.524
0.08	0.05	1.600	8.555	162.545	13.985	265.715	18.265	347.035
0.08	0.04	2.000	3.376	81.024	5.696	136.704	7.520	180.480
0.050	0.040	1.250	45.420	1090.080	72.280	1734.720	93.480	2243.520
0.050	0.030	1.667	7.218	233.382	11.880	384.120	15.552	502.848
0.050	0.025	2.000	3.445	134.355	5.828	227.273	7.703	300.398
0.050	0.010	5.000	1.001	99.099	1.001	99.099	1.001	99.099
0.020	0.015	1.333	26.9745	1771.3255	43.2525	2840.2475	56.0985	3683.8015
0.020	0.010	2.000	3.5150	347.9850	5.9600	590.0400	7.8810	780.2190
0.020	0.005	4.000	1.0005	199.0995	1.0005	199.0995	1.3170	262.0830
0.010	0.008	1.25	47.0888	5839.0112	75.0000	9300.0000	97.0320	12031.9680
0.010	0.005	2.00	3.5385	704.1615	6.0035	1194.6965	7.9400	1580.0600
0.010	0.002	5.00	1.0002	499.0998	1.0002	499.0998	1.0002	499.0998
0.010	0.001	10.00	1.0001	999.0999	1.0001	999.0999	1.0001	999.0999

$$\gamma'' = \gamma' + \frac{1}{k}\sum_k x,$$

$$\xi'' = \xi' + n_0 - \frac{1}{k}\sum_k x,$$

(15)

where x is the number of observations in a replication exceeding the standard and k is the number of replications. We then used the parameters from the posterior distribution to determine the posterior value of p_0, defined as p_{post}, satisfying Equation (5) with $P(p \leq p_{post}) = 1 - \alpha$.

If $p_{post} \leq p_0$, we would conclude the system (or model) from which the sample of observations was obtained meets the limit standard.

8   EMPIRICAL EXAMPLE

Although not the focus of this paper, we provide some initial results of our empirical experiment. We developed a model of an M/M/1 queuing system with utilization of 0.8. We chose this model because it is the same used by Raatikainen (1995), Kim and Nelson (2006) and others. Law and Kelton (2000) employed a similar M/M/1 system

but with utilization of 0.9. Each replication consisted of $n = 24000$ observations, with 400 replications made. We were interested in the proportion of observations exceeding a 15-minute delay time. We established the septuple limit standard requirement of (15, 0, 0.05, 0, 0, 0.05, 0), or $p_0 = 0.05$ with $\alpha = 0.05$. Using Equations (11), (12), and (13) with $\tau = 2$, we set the initial prior parameters to $\gamma' = 3.444$ and $\xi' = 134.326$. We evaluated the replications at a half-width of $\varepsilon = 0.005$, resulting in a sample size of $n_0 = 5141$ through the use of Equation (13). (As stated in Section 7, this is a non-Bayesian approach to determine sample size.) We also evaluated the replications at $n_0 = 24000$, with the shape of the distribution of proportions for this sample size shown in Figure 1. Using Equation (15), we found $\gamma'' = 202.999$ and $\xi'' = 5075.771$ for $n_0 = 5141$, resulting in a $p_{post} = 0.0429$. (If we use the normal confidence interval methodology, we find the $1 - \alpha$ confidence interval to be 0.0432.) We also found $\gamma'' = 915.035$ and $\xi'' = 23222.74$ for $n_0 = 24000$, resulting in a $p_{post} = 0.0399$. (If we use the normal confidence interval methodology, we find the $1 - \alpha$ confidence interval to be 0.0400.)

9   CONCLUSION

The beta distribution accurately represents the distribution of sample proportions of DES models. The Bayesian elicitation methodology defined above provides a rational process for developing an initial prior to evaluate limit standards.

This paper focused on an overview of the Bayesian approach to analysis of limit standards and development of the initial prior. Further research includes refinement of the process for determining the initial prior distribution, developing an informative stopping rule and further comparison to current (Frequentist) methodologies.

ACKNOWLEDGEMENTS

REFERENCES

Chen, E. and W. Kelton (1999). "Simulation-Based Estimation of Quantiles," *Proceedings of the 1999 Winter Simulation Conference*. P. Farrington, H. Nembhard, D. Sturrock, G. Evans, eds. Piscataway, NJ: IEEE, pp. 428—434.

Chen, E. and W. Kelton (2004). "Quantile and tolerance-interval estimation in simulation," accepted for *European Journal of Operational Research*.

Chen, E. and W. Kelton (2004). "Experimental Performance Evaluation of Histogram Approximation for Simulation Output Analysis," *Proceedings of the 2004 Winter Simulation Conference*. R. Ingalls, M. Rossetti, J. Smith, B. Peters, eds. Piscataway, NJ: IEEE, pp. 685—693.

Chick, S. (1997). "Selecting the Best System: A Decision-Theoretic Approach," *Proceedings of the 1997 Winter Simulation Conference*. S. Andradottir, K. J. Healy, D. H. Withers and B. L. Nelson, eds. Piscataway, NJ: IEEE, pp. 326—333.

Chick, S. (2001). "Input Distribution Selection for Simulation Experiments: Accounting for Uncertainty," *Operations Research*, 49, pp. 744—758.

Chick, S. (2006). "Bayesian Ideas and Discrete Event Simulation: Why, What and How," *Proceedings of the 2006 Winter Simulation Conference*. L. Perrone, F. Wieland, J. Liu, B. Lawson, D. Nicol, R. Fujimoto, eds. Piscataway, NJ: IEEE, pp. 96—106.

Chick, S. and K. Inoue (1998). "Sequential Allocations that Reduce Risk for Multiple Comparisons," *Proceedings of the 1998 Winter Simulation Conference*. D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds. Piscataway, NJ: IEEE, pp. 669—676.

Cochran, W. (1963). *Sampling Techniques*, 2nd ed. New York: Wiley.

Creasey, R., K. White, M. Marks, B. Waller (2005). "Determination of the 'Best' System that Meets a Limit Standard," *Proceedings of the 2005 Winter Simulation Conference*. M. Kuhl, N. Steiger, F. Armstrong, J. Joines, eds. Piscataway, NJ: IEEE, pp. 718—722.

Creasey, R. and K. White (2006). "Comparison of Limit Standards Using a Sequential Probability Ratio Test," *Proceedings of the 2006 Winter Simulation Conference*. L. Perrone, F. Wieland, J. Liu, B. Lawson, D. Nicol, R. Fujimoto, eds. Piscataway, NJ: IEEE, pp. 356—359.

Gavasakar, U. (1988). "A Comparison of Two Elicitation Methods for a Prior Distribution for a Binomial Parameter," *Management Science*, **34** (6), pp. 784-790.

Henderson, S. (2003). "Input Model Uncertainty: Why Do We Care and What Should We Do About It?" *Proceedings of the 2003 Winter Simulation Conference*. S. Chick, P. Sanchez, D. Ferrin, and D. Morrice, eds. Piscataway, NJ: IEEE, pp. 90—100.

Johnson, N., S. Kotz,. and N. Balakrishnan (1994). *Continuous Univariate Distributions*, 2nd ed., New York: John Wiley & Sons, Inc.

Johnson, N., A. Kemp, S. Kotz (2005). *Univariate Discrete Distributions*, 3rd. ed. Hoboken, NJ: John Wiley & Sons, Inc.

Kim, S. and B. Nelson (2006). "On the Asymptotic Validity of Fully Sequential Selection Procedures for Steady-State Simulation," *Operations Research*. 54 (3), 475—488.

Law, A.M. and W.D. Kelton (2000). *Simulation Modeling and Analysis*, 3rd ed. Boston, MA: McGraw-Hill.

Lee, P. (1997). *Bayesian Statistics: An Introduction*, 2nd ed., New York: John Wiley & Sons, Inc.

Raatikainen, K. (1995). "Simulation-based Estimation of Proportions," *Management Science*, **41** (7), pp. 1202—1223.

White, K., K. Johnson, T. Adams, C. Everline, and R. Creasey (2007). "Best practices and recommended statistical techniques for developing Monte Carlo sampling plans." *NASA Technical Paper*.

AUTHOR BIOGRAPHIES

ROY R. CREASEY, JR. is a Lecturer in Management at Longwood University, and a Ph.D. student in Systems and Information Engineering at the University of Virginia. He received the B.S. in Civil Engineering from the Virginia Military Institute and the M.E. in Systems Engineering from the University of Virginia. He has seventeen years prior experience as a systems and test engineer in the defense industry. He is a student member of INFORMS and IEEE/SMC. His e-mail is `<creaseyrr@longwood.edu>`.

K. PRESTON WHITE, JR. is a Professor of Systems and Information Engineering at the University of Virginia. He received the B.S.E., M.S., and Ph.D. degrees from Duke University. He has held faculty appointments at Polytechnic University and Carnegie-Mellon University and served as Distinguished Visiting Professor at Newport News Shipbuilding and at SEMATECH. He is U.S. Editor for *International Abstracts in Operations Research* and Associate Editor for *International Journal of Intelligent Automation* and *IEEE Transactions on Electronics Packaging Manufacturing Technology*. He is a member of INFORMS, SCS, and INCOSE and a senior member of IEEE and IIE. He is a member of the Council of the INFORMS Society on Simulation (ISIM) and represents ISIM on the INFORMS Subdivision Council. He sits on the Advisory Board of VMASC and represents IEEE/SMC on the Winter Simulation Conference Board. His e-mail is `<kpwhite@virginia.edu>`.