Proceedings of the 2008 Winter Simulation Conference
S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds.

COMPARISON OF BAYESIAN PRIORS FOR HIGHLY RELIABLE LIMIT MODELS

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

K. Preston White, Jr.
Department of Systems and Information Engineering
University of Virginia
151 Engineers' Way
Charlottesville, VA, U.S.A.

Linda B. Wright
Cheryl F. Davis
College of Business and Economics
Longwood University
201 High Street
Farmville, VA 23909, U.S.A.

ABSTRACT

Limit standards are probability interval requirements for proportions. Simulation literature has focused on finding the confidence interval of the population proportion, which is inappropriate for limit standards. Further, some Frequentist approaches cannot be utilized for highly reliable models, or models which produce no or few nonconforming trials. Bayesian methods provide approaches that can be utilized for all limit standard models. We consider a methodology developed for Bayesian reliability analysis, where historical data is used to define the *a priori* distribution of proportions p, and the customer desired *a posteriori* maximum probability is utilized to determine sample size for a replication.

1 INTRODUCTION

All methods and theories regarding statistical inference – such as sampling (or Frequentist) theories, Bayesian methods, and likelihood approaches – require some degree of subjectivity in their use. For example, sampling theory requires assumptions in order to utilize confidence intervals, to determine the proper estimator to use, or to declare that observations are independent and identically distributed (i.i.d.).

Bayesian methods, on the other hand, provide a way of introducing and incorporating assumptions about prior knowledge or ignorance. Through the use of Bayes' Theorem, these assumptions are incorporated into posterior descriptions and inferences; these are descriptions and inferences about some parameter of interest after incorporating observations.

Consider proportional output analysis for discrete event simulation (D.E.S.). D.E.S. literature focuses on determining the confidence interval for a proportion, also known as the *Wald confidence interval*, defined as

$$\Pr(\bar{p} - z_{1-\alpha/2}\sigma_{\bar{p}} < \rho < \bar{p} + z_{1-\alpha/2}\sigma_{\bar{p}}) \geq 1-\alpha \text{, where} \quad (1)$$

ρ = population proportion,

$$I_j = \begin{cases} 1 & \text{the desired performance is achieved for trial } j \\ 0 & \text{otherwise,} \end{cases}$$

$$\bar{p} = E[I] = \frac{1}{n}\sum_{j=1}^{n} I_j,$$

$$\sigma_{\bar{p}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n}},$$

α = level of significance.

for j = 1, 2, …, n trials (Law 2007, Agresti and Coull 1998). Use of the Wald interval assumes each observation I_j is i.i.d. and the probability density function (p.d.f.) of sample proportions is asymptotic normal. However, for any given n, as the mass of the p.d.f. approaches 0 or 1, the distribution becomes increasingly non-normal; correcting this skewness requires substantially larger sample sizes. (Raatikainen 1995) Chen and Kelton's (1999, 2004A, 2004B) use of quantile estimation relaxed the normality requirement, but their methodology also requires a very large sample size.

Many standards containing proportional requirements are not interested in determining a confidence interval for ρ, but in determining the *probability interval* of p, or the probability that the random variable (r.v.) P is within a specific range of proportions. (Mandel 1964) In D.E.S., this form of proportional requirements is known as a proportional *limit standard*.

Limit standards are divided into two general categories – those measuring proportion conforming and those measuring proportion nonconforming. Discussion in this paper is limited to conforming limit standards, which for a single replication is defined as the triple $\{I, p^*, \beta\}$, where I was previously defined, p^* is the minimum desired proportion conforming and

$$\Pr(P < p^*) \le \beta, \quad (2)$$

or

$$\Pr(p^* < P < 1) \ge \beta. \quad (3)$$

(Creasey and White 2007) The procedures defined in this paper also apply (with changes in the definition of p^* and inequality directions) to non-conforming limit standards.

2 FREQUENTIST APPROACH

Since a limit standard is based on a Bernoulli r.v., the binomial cumulative distribution function (c.d.f.) can be utilized to determine *a priori* the maximum number of failures, c, that can occur in n_c trials and continue to meet the standard, or

$$\beta \ge \sum_{k=0}^{c} \frac{n_c!}{k!(n_c - k)!}(1-p^*)^k (p^*)^{n_c - k}. \quad (4)$$

[Creasey, *et al*. (2005), Creasey and White (2007)]

Consider a conforming limit standard of $\{I, 0.9973, 0.1\}$. Using Equation (4), several values of n_c and c meeting this standard are depicted in Table 1.

Table 1: Combinations of (c, n_c) for a binomial c.d.f. meeting an $\{I, 0.9973, 0.1\}$ limit standard.

c	n_c
0	852
1	1440
2	1972
3	2471
4	2958
5	3434
6	3902
7	4357
8	4812
9	5259
20	10014

If Equation (4) is generalized to

$$\Pr_a = F(p) = \sum_{k=0}^{c} \frac{n_c!}{k!(n_c - k)!}(1-p)^k (p)^{n_c - k}, \quad (5)$$

where $\Pr_a$ is the probability of accepting the system, then for any (c, n_c) combination, a c.d.f. for p, defined as $F(p)$, can be determined. Also, for any (c, n_c) combination, a p.d.f. for p, defined as $f(p \mid x)$, can be determined using

$$f(p \mid x) = \binom{n_c}{c}(1-p)^c (p)^{n_c - c}. \quad (6)$$

Equation (6) is also known as the likelihood function of p, or $L(p \mid c, n_c)$.

Figures 1 and 2 depict the c.d.f. and p.d.f., respectively, of p for several (c, n_c) combinations listed in Table 1. [Quality control analysts call the curves depicted in Figure 1 *operational characteristic* (OC) curves. (Montgomery 2001)] Upon further review of Figure 2, for small n, the p.d.f. of p is negatively skewed, confirming that the Wald interval (Equation (1)) is not always the correct methodology to use for limit standards.

The (c, n_c) distribution functions developed using Equations (5) and (6) are considered "pre-posterior" distributions of p; in other words, the distributions are the resulting (or *posterior*) distributions of p after taking a sample of n_c trials and observing c failures. For example, after taking n_c = 1440 observations and observing only one failure, the p.d.f. of p is as depicted in curve B of Figure 2. (Martz and Waller 1982)

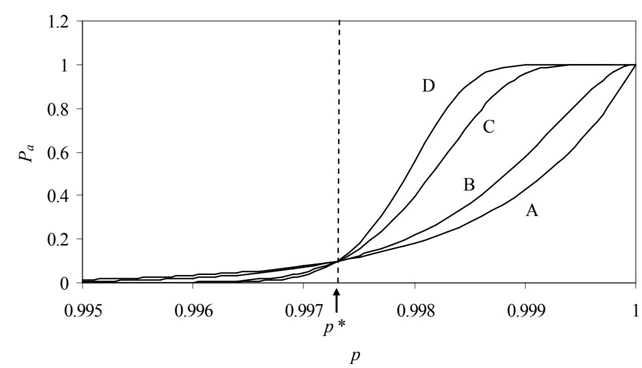

Figure 1: Chart of p versus $\Pr_a$ depicting four combinations of (c, n_c) for an $\{I, 0.9973, 0.1\}$ limit standard: A-(0, 852); B-(1, 1440); C-(9, 5259); D-(20, 10014).

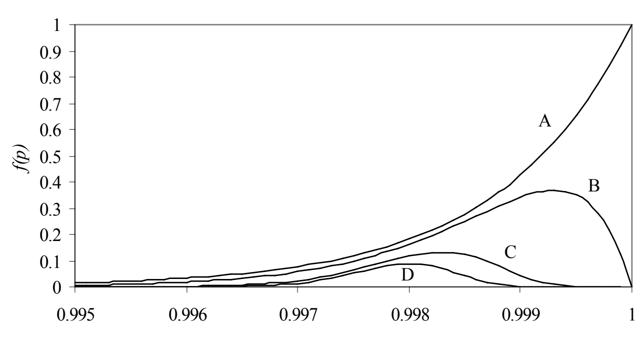

Figure 2: Chart of p versus $f(p)$ depicting four combinations of (c, n) for an $\{I, 0.9973, 0.1\}$ limit standard: A-(0, 852); B-(1, 1440); C-(9, 5259); D-(20, 10014).

The posterior distributions of p are, in fact, Beta distributed. (Creasey and White 2007) Since the binomial p.d.f. provides an exact probability interval, the Clopper-Pearson (1934) interval can be utilized, resulting in the Bayesian (or Jeffrey's) interval (Brown et al. 2001, Krishnamoorthy and Peng 2007). For a conforming limit standard, changing Equation (2) to a Bayesian interval yields

$$p' = 1 - \text{Be}^{-1}(1 - \beta, x, n - x + 1), \tag{7}$$

where p' is the resulting (or posterior) proportion conforming at $\Pr_a = \beta$, $\text{Be}^{-1}(d, \gamma, \xi)$ is the d-th quantile of the Beta c.d.f. with parameters γ and ξ, x is the number of failed trials in a sample of n trials, and $0 < x \le n$. Using the above limit standard example, after taking $n = 5259$ observations and observing $x = 9$ failures, $p' = 0.99753$. For conforming standards, as long as $p^* \le p'$, the model continues to meet the standard.

3 AGRESTI-COULL METHODOLOGY

There are several issues with the Frequentist approaches outlined above. First, as observed in Figures 1 and 2, a unique situation occurs when $c = x = 0$. For this condition, Equation (6) reduces to $f(p) = p^n$, which results in a strictly increasing p.d.f. and c.d.f. When $x > 0$, the resulting p.d.f. is always unimodal. Further, if the analyst believes *a priori* to execution the model is very stable and should result in no failed trials, use of Equation (7) may not be appropriate, since it requires $x > 0$.

Second, the analyst makes an *a priori* assumption the model replication begins in an empty and idle state. As a result, in a sequential analysis of trials no information regarding the interim proportion is available until the first failed trial is observed.

Third, as stated above, use of the Wald interval is not appropriate for limit standards. Further, use of the Wald interval requires significantly larger sample sizes near the extremes of 0 or 1. The impact of not increasing sample size results in an overestimation of the mean proportion and an underestimation of the standard deviation (for proportion conforming) (Agresti and Coull 1998, White 2007).

To reduce the impact of the estimation errors, Agresti and Coull (1998) developed a simple adjustment to the Wald interval. For proportion conforming, they defined

$$\tilde{n} = n + 4 \text{ and } \tilde{p} = \frac{n - x + 2}{\tilde{n}},$$

then substituted $\tilde{n}$ for n and $\tilde{p}$ for $\bar{p}$ in Equation (1). If the upper limit of the confidence interval exceeds 1, the upper limit is adjusted to 1. This methodology injects an initial pseudo-sample of size four containing two initial pseudo-successes, for an initial $\tilde{p} = 0.5$. Use of this methodology improves the estimation of confidence intervals, but provides minimal improvement in skewness.

For example, if a population has a proportion $\rho = 0.9973$, then on average 27 failures should occur in every 10,000 observations. However, for samples of size $n = 100$, only one in every four samples will observe a failure. For samples where no failures are observed, $\bar{p} = 1$ and $\sigma_{\bar{p}} = 0$. Using the Agresti-Coull methodology, $\tilde{p} = 0.9808$ and $\sigma_{\bar{p}} = 0.01346$ for samples with no failures, resulting in an adjusted 95% confidence interval of [0.9544, 1]. (White 2007)

4 BAYESIAN APPROACH

While the Agresti-Coull methodology is utilized for confidence intervals, it does show the benefit of establishing an *a priori* pseudo-proportion.

We therefore consider a Bayesian approach, where the final, or posterior distribution of p is the result of an *a priori* distribution conditioned by actual data, or

$$g(p \mid x') = \frac{f(x' \mid p) g(p)}{\sum_p f(x' \mid p) g(p)} \tag{8}$$

for discrete distributions or

$$g(p \mid x') = \frac{f(x' \mid p) g(p)}{\int_0^1 f(x' \mid p) g(p) dp} \tag{9}$$

for continuous distributions, where $x' = n - x$ (indicating the number of successes), $g(p)$ is the prior distribution of the r.v. ρ, $f(x' \mid p)$ is the conditioning function or sampling distribution, and $g(p \mid x')$ is the posterior distribution of the r.v. ρ, for $0 < p < 1$. (Martz and Waller 1982) [Note that in Bayesian analysis the parameter ρ is considered to be an r.v.] If the conditioning function $f(x|p)$ is defined as the binomial p.d.f. found in Equation (6), and substituting x' for c, Equation (9) becomes

$$g(p \mid x') = \frac{p^{x'} (1-p)^{n-x'} g(p)}{\int_0^1 p^{x'} (1-p)^{n-x'} g(p) dp}. \tag{10}$$

(Martz and Waller, 1982)

5 DEVELOPING AN INITIAL PRIOR DISTRIBUTION

Utilizing Equation (10), we wish to find an initial prior distribution for P. In Bayesian statistics, the prior distribution is typically developed from the analyst's or customer's initial belief or from empirical (historical) data. Chick (2001) provides a discussion of these approaches and how they should be utilized in simulation input analysis. Using the customer's initial beliefs typically results in subjective probabilities which may not substantially benefit the simulation analyst. Chick (2001) also suggests that, if historical data is utilized, it must be unique to each replication. Other examples of defining initial prior distributions from customer's initial belief are in Martz and Waller (1982) and Pham (1990).

We consider two types of distributions as candidates for prior distributions: a binomial $Bi(n, p)$ distribution and a Beta $Be(\gamma_0, \xi_0)$ distribution. Although the uniform $U(0, 1)$ distribution is utilized because it represents a complete lack of knowledge regarding the prior distribution, the uniform distribution is in fact one form of the Beta distribution (Martz and Waller 1982, Lee 1997, Creasey and White 2007).

Binomial Bi(n, p) Prior Distribution

The distribution $g(p)$ is defined as

$$g(p) = \begin{cases} 1 & 0 < p < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Using Equation (10), we have

$$g(p \mid x') = \frac{p^{x'}(1-p)^{n-x'}}{\int_0^1 p^{x'}(1-p)^{n-x'}\,dp},$$

for $0 < p < 1$.

Thus, $g(p \mid x')$ is a $Be(x'+1, n+2)$ distribution, with a mean $E(P \mid x')$ and variance $\mathrm{Var}(P \mid x')$ of

$$E(P \mid x') = \frac{x'+1}{n+2}, \tag{11}$$

$$Var(P \mid x') = \frac{(x'+1)(n-x'+1)}{(n+2)^2(n+3)}. \tag{12}$$

Beta Be(γ_0, ξ_0) Prior Distribution

In addition to being a natural conjugate (i.e.; the posterior distribution is also a Beta distribution), the Beta distribution is very robust, assuming many different symmetrical or asymmetrical shapes, based on the values of γ_0 pseudo-successes and ξ_0 pseudo-trials. For example, when $\xi_0 - \gamma_0 > 1$ and $\gamma_0 > 1$ the distribution has a single mode, or when $\gamma_0 = 1$ and $\xi_0 = 2$ the distribution is uniform, and the distribution is J-shaped (strictly increasing) when $\xi_0 - \gamma_0 < 1$.

The distribution $g(p)$ is defined as

$$g(p) = \frac{1}{be(\gamma_0, \xi_0)} p^{\gamma_0 - 1}(1-p)^{\xi_0 - 1},$$

where $g(p) = 0$ for p not in the interval [0, 1],

$$b(\gamma_0, \xi_0) = \frac{\Gamma(\gamma_0)\Gamma(\xi_0)}{\Gamma(\gamma_0 + \xi_0)} = \int_{p=0}^{1} p^{\gamma_0 - 1}(1-p)^{\xi_0 - 1}\,dp$$

is the Beta function, $\Gamma(\bullet)$ is the Gamma function and γ_0, $\xi_0 > 0$.

Using Equation (10), we have

$$g(p \mid x') = \frac{1}{be(x'+\gamma_0, n+\xi_0)} p^{(x'+\gamma_0)-1}(1-p)^{(n+\xi_0-x'-\gamma_0)-1}$$

which is a $Be(x' + \gamma_0, n + \xi_0)$ distribution, for $0 < p < 1$. The mean $E(P \mid x')$ and variance $\mathrm{Var}(P \mid x')$ are

$$E(P \mid x') = \frac{x'+\gamma_0}{n+\xi_0}, \tag{13}$$

$$Var(P \mid x') = \frac{(x'+\gamma_0)(n+\xi_0-x'-\gamma_0)}{(n+\xi_0)^2(n+\xi_0+1)}. \tag{14}$$

We now consider the application of three Beta-distributed priors, $Be(1,2)$, which is equivalent to the $U(0,1)$ distribution, $Be(2,4)$, which is similar to the Agresti-Coull method, and $Be(6,8)$. (Note in both priors the pseudo prior proportion of conforming trials is 0.5) Using c as defined above, where $c = n - x'$, Table 2 shows the minimum values of n for a specific value of c necessary to meet the standard of $\beta = 0.1$.

Table 2: Minimum sample size necessary for three prior distributions and different maximum number of observed failed trials.

c	Be(1,2)	Be(2,4)	Be(6,8)
0	852	1435	1431
1	1437	1964	1960
2	1966	2466	2462
3	2468	2951	2947
4	2953	3424	3420
5	3426	3889	3885
6	3891	4347	4345
7	4349	4799	4795
8	4801	5246	5242
9	5248	5690	5686
20	9992	10411	10407

6 CONCLUSION

Considering Table 2, we observe the necessary sample size when using the Be(1,2) prior is slightly less than that required in the Frequentist method (Table 1). In fact, when the number of nonconforming historical trials is ≤ 1 (or $\xi_0 - \gamma_0 \leq 1$) and the number of historical trials is $\gamma_0 \geq 2$, the Bayesian method provides a slight benefit in minimum sample size. However, when $\xi_0 - \gamma_0 \geq 1$, the necessary minimum sample size increases.

7 FUTURE RESEARCH

We plan to expand on this research, by considering an approach by Sheng and Fan (1992), where a maximum probability of any proportion, ε, is defined *a priori* by the customer, or

$$\max \Pr[P = p] \leq \varepsilon. \tag{15}$$

Sheng and Fan then applied this approach to the results of historical data to determine the minimum sample size n and number of successes $n - x$ necessary to satisfy Equation (15), using a binomial p.d.f. for the prior and conditioning distributions. However, the authors limit their research to conditions where no failures occur in the historical and test data. While the authors focused on using a step function and binomial distribution, they state a more exact conclusion would be realized by using the Beta conjugate prior distribution.

REFERENCES

Agresti, A. and B. Coull. 1998. Approximate is better than 'exact' for interval estimation of binomial proportions, *The American Statistician* **52** (2):119-126

Brown, L. D., Cai, T., Das Gupta, A. 2001. Interval estimation for a binomial proportion, *Statist. Sci.* **16**:101–133,

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. 2001. Input Distribution Selection for Simulation Experiments: Accounting for Uncertainty, *Operations Research*, 49, pp. 744-758.

Clopper. C. J., and Pearson, E. S. 1934. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial, *Biometrika.* **26**: 404-413.

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.

Creasey, R. and K. White. 2007. A Bayesian Approach to Analysis of Limit Standards, *Proceedings of the 2007 Winter Simulation Conference*. S. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. Tew, and R. Barton, eds. Piscataway, NJ: IEEE, pp.

Krishnamoorthy, K. and J. Peng. 2007. Some Properties of the Exact and Score Methods for Binomial Proportion and Sample Size Calculation, *Comm. in Stat.—Sim. and Comp.* **36**: 1171-1186.

Law, A.M. 2007. *Simulation Modeling and Analysis*, 4th ed. Boston, MA: McGraw-Hill.

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

Mandel, J. 1964 *The Statistical Analysis of Experimental Data*. New York: John Wiley & Sons, Inc.

Martz, H. and R. Waller. 1982. *Bayesian Reliability Analysis*. New York: John Wiley & Sons, Inc.

Montgomery, D. 2001. *Introduction to Statistical Quality Control*, 4th ed. New York: John Wiley & Sons, Inc.

Pham, T.G. 1990. Comment on: A Bayes Sensitivity Analysis When Using the Beta Distribution as a Prior, *IEEE Transactions on Reliability*, **39** (1), pp. 6-9.

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

Sheng, Z. and D. Fan. 1992. Bayes Attribute Acceptance-Sampling Plan, *IEEE Transactions on Reliability*, **41** (2), pp. 307-309.

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 an Assistant Professor in Management at Longwood University. He received the B.S. in Civil Engineering from the Virginia Military Institute and the M.E. and Ph.D. in Systems Engineering from the University of Virginia. He is U.S. Co-Editor for *International Abstracts in Operations Research*. He has seventeen years prior experience as a systems and test engineer in the defense industry. He is a member of

INFORMS, INFORMS Simulation Society (ISIM) and IEEE. 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. Co-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`>.

LINDA B. WRIGHT is an Associate Professor of Marketing at Longwood University. She received her MBA and BGS degrees from the University of Michigan, her JD degree from Wayne State University, and her Ph.D. degree from the University of Tennessee. She is on the Editorial Board for the *Journal of Digital Forensic, Security, and Law*. Her e-mail is <`wrightlb@longwood.edu`>.

CHERYL F. DAVIS is a Senior Lecturer of Computer Information Management Systems at Longwood University. She is currently a graduate student at Lynchburg College, and previously received a B.A. from Virginia Commonwealth University and a M.Ed. from Longwood University. Her e-mail is <`daviscf@longwood.edu`>.