COMPARISON OF BAYESIAN AND FREQUENTIST ASSESSMENTS OF UNCERTAINTY FOR SELECTING THE BEST SYSTEM

Koichiro Inoue
Stephen E. Chick

Department of Industrial and Operations Engineering
The University of Michigan
1205 Beal Avenue
Ann Arbor, MI 48109-2117, U.S.A.

ABSTRACT

An important problem in discrete-event stochastic simulation is the selection of the best system from a finite set of alternatives. There are many techniques for ranking and selection and multiple comparisons discussed in the literature. Most procedures employ classical frequentist approaches, although there has been recent attention to Bayesian methods. In this paper, we compare Bayesian and frequentist assessments of unknown means of simulation output. First, we present a Bayesian formulation for describing the probability that a system is the best, given prior information and simulation output. This formulation provides a measure of evidence that a given system is best when there are two or more systems, with either independent or common random numbers, with known or unknown variance and covariance for the simulation output, given a Gaussian output assumption. Many, but not all, frequentist assessments are shown to be derivable from assumptions of normality of simulation output when certain limits are taken. So we compare Bayesian probability of correct selection (P(CS)) with frequentist P-value as a measure of evidence that the best system is selected under normality assumptions.

1 INTRODUCTION

In this paper, we explore similarities and differences of standard frequentist approaches (Law and Kelton, 1991) and a Bayesian approach (deGroot, 1970 and Chick, 1997) to measure the evidence that a given simulated system is best. Best here is taken to be the system with the highest mean value of performance measure for simulation output.

Goldsman and Nelson (1994) present a state-of-art review of multiple comparison, ranking and selection procedures and related topics in computer simulation. A large class of these techniques relies on frequentist assessments and the 'indifference zone' notion, which attempts to identify differences of at least a certain magnitude between systems.

On the other hand, there is a growing literature for the Bayesian decision-theoretic simulation output analysis, especially for computer budget and sample allocation procedure (Gupta and Miescke, 1994; Gupta and Miescke, 1996). Formally, decision theory talks about expected utilities. To simplify, we here talk more generally about the mean value of output.

We discuss Bayesian assessments of unknown mean for selecting the best system with normally distributed simulation output, independent or dependent (e.g. common random numbers) simulation replications, and known or unknown covariance structures, for $K = 2$ or $K > 2$ systems. In Section 4, we compare frequentist P-values and Bayesian posterior P(CS) as evidence for correct selection. We consider five simulated (s, S) inventory policies to illustrate ideas.

2 A BAYESIAN ASSESSMENT OF UNKNOWN MEAN FOR A SINGLE SYSTEM

2.1 Normality Assumptions

We first look at a Bayesian single system model ($K = 1$) in order to provide a basis for $K \geq 2$ systems. Suppose stochastic simulation provides independent output, $o_1, ..., o_R$ that is normally distributed with unknown mean μ and precision $\ddot{e}$.

$$p\left(o_r \mid \mu, \ddot{e}\right) = \left(\frac{\ddot{e}}{2\partial}\right)^{1/2} \exp\left[-\frac{1}{2} \ddot{e}\left(o_r - \mu\right)^2\right]$$

We would like to know whether the system mean μ is larger than some specific value μ^*.

2.2 Bayesian Approach: Known Precision

Assume precision λ is known, and that the unknown mean μ has the conjugate normal prior distribution with mean μ_0 and precision λ_0. The conjugate prior allows for closed form solutions. In particular, the posterior distribution of μ is then:

$$p(\mu \mid D) \sim Normal\left(\frac{\lambda_0 \mu_0 + R\lambda\overline{O}}{\lambda_0 + R\lambda}, \lambda_0 + R\lambda\right)$$

where $D = \{o_1, ..., o_R\}$ and $\overline{O} = \sum_{r=1}^{R} \frac{o_r}{R}$

So the probability that the mean exceeds μ^* is

$$P(\mu > \mu^* \mid D) = \int_{\mu > \mu^*} p(\mu \mid D) = \Phi\left(\sqrt{\lambda_0 + R\lambda}\left(\frac{\lambda_0 \mu_0 + R\lambda\overline{O}}{\lambda_0 + R\lambda} - \mu^*\right)\right)$$

where Φ is the cumulative distribution function of the standard normal distribution.

Now suppose that there is little information with the normal prior density. deGroot (1970) suggests to use a reference prior distribution which is uniform over the whole real line such that $p(\mu) = c$ given by taking limit $\lambda_0 \to 0$, where c is constant. This reference prior provides the following posterior distribution:

$$p(\mu \mid D) \sim Normal(\overline{O}, R\lambda)$$

Then the probability that whether the system mean μ is larger than some specific value μ^* is

$$P(\mu > \mu^* \mid D) = \int_{\mu > \mu^*} p(\mu \mid O) = \Phi\left(\sqrt{R\lambda}(\overline{O} - \mu^*)\right)$$

2.3 Bayesian Approach: Unknown Precision

Next we consider the case where precision λ is unknown. We use the normal-gamma conjugate prior for mean μ and precision λ: the conditional distribution of μ given λ is normal with mean μ_0 and precision $n_0\lambda_0$ and the marginal distribution of λ is $Gamma(\alpha, \beta)$:

$$p(\mu, \lambda) = \left(\frac{n_0\lambda}{2\pi}\right)^{1/2} \exp\left[-\frac{1}{2} n_0\lambda(\mu - \mu_0)^2\right] \bullet \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda)$$

The posterior marginal distribution of μ is then:

$$p(\mu \mid D) \sim Student\left(\mu_R, \frac{(n_0 + R)\alpha_R}{\beta_R}, 2\alpha_R\right)$$

where $\quad \mu_R = \frac{n_0\mu_0 + R\overline{O}}{n_0 + R}$

$$\alpha_R = \alpha + \frac{R}{2} \qquad \beta_R = \beta + \frac{s}{2} + \frac{n_0 R(\mu_0 - \overline{O})^2}{2(n_0 + R)}$$

$$s = \sum_{r=1}^{R}(o_r - \overline{O})^2$$

and the density function of $Student(\mu, \hat{\epsilon}, \nu)$ is

$$p(x \mid \mu, \hat{\epsilon}, \nu) = c\left[1 + \frac{\hat{\epsilon}}{\nu}(x - \mu)^2\right]^{-(\nu+1)/2}$$

where $\quad c = |\hat{\epsilon}|^{1/2} \Gamma((\nu+1)/2) / \left(\Gamma(\nu/2)(\nu\pi)^{1/2}\right)$

So the probability that the system has mean greater than μ^* is

$$P(\mu > \mu^* \mid O) = \int_{\mu > \mu^*} p(\mu \mid O) = \Phi_{2\alpha_R}\left(\sqrt{\frac{(n_0 + R)\alpha_R}{\beta_R}}(\mu_R - \mu^*)\right)$$

where Φ_α is the cumulative function of the standardized Student distribution with α degrees of freedom.

Vague knowledge of λ is often represented by the reference prior $p(\mu, \lambda) \propto 1/\lambda$ ($n_0 \to 0$, $\alpha \to 0$, $\beta \to 0$ for conjugate prior), which is improper. The posterior distribution becomes:

$$p(\mu \mid D) \sim Student\left(\overline{O}, \frac{R(R-1)}{s}, R-1\right)$$

The mean is greater than μ^* with probability

$$P(\mu > \mu^* \mid O) = \Phi_{R-1}\left(\sqrt{\frac{R(R-1)}{s}}(\overline{O} - \mu^*)\right)$$

3 BAYESIAN MODEL FOR MULTIPLE SYSTEMS

Suppose there are K different systems, and the objective is to select the system with the maximum mean. The probability that a given system is best, P(CS), will be calculated from the posterior distribution of the mean.

3.1 Independent Replications

This section presents assumptions, analysis, and approximation for the determining the P(CS) that a given simulated system has the highest mean, given the

assumption of independent simulation replications for known or unknown precision.

We make the following assumptions:

1. Let $o_{k,1},...,o_{k,R_k}$ be the output for the k-th system. The number of replications for each system is not necessarily equal.
2. The O_k are independent and normally distributed with (potentially unknown) mean μ_k and precision τ_k.

Since the output for each system is independent, we can apply the single system model to define the posterior distribution for μ_k.

Fact 3.1 *Assume the above, except that τ_k is known. Further assume that the unknown mean μ_k has the conjugate normal prior distribution with mean μ_{k0} and precision τ_{k0}. The posterior distribution of the conditional) mean is a normal distribution:*

$$p(\mu_k|D_k) \sim Normal\left(\frac{\tau_{k0}\mu_{k0}+R_k\tau\overline{O}_k}{\tau_{k0}+R_k\tau_k},\tau_{k0}+R_k\tau_k\right) \quad (1)$$

where $D_k = \{o_{k,1},...,o_{k,R_k}\}$ and $\overline{O}_k = \sum_{r_k=1}^{R_k}\frac{o_{r_k}}{R_k}$

Further if a reference prior is used,

$$p(\mu_k|D_k) \sim Normal(\overline{O}_k,R_k\tau_k) \quad (2)$$

Likewise, we have the following fact for the case of unknown precision.

Fact 3.2 *Given the normality assumptions with unknown τ_k. Using the normal-gamma conjugate prior about mean μ_k and precision τ_k: the conditional distribution of μ_k given τ_k is normal with mean μ_{k0} and precision $n_{k0}\tau_{k0}$ and the marginal distribution of τ_k is $Gamma(\acute{a}_k,\hat{a}_k)$. The posterior distribution of μ_k is then:*

$$p(\mu_k|D_k) \sim Student\left(\mu_{kR},\frac{(n_{k0}+R_k)\acute{a}_{kR}}{\hat{a}_{kR}},2\acute{a}_{kR}\right)$$
(3)

where
$$\mu_{kR} = \frac{n_{k0}\mu_{k0}+R_k\overline{O}_k}{n_{k0}+R_k}$$

$$\acute{a}_{kR} = \acute{a}_k + \frac{R_k}{2}$$

$$\hat{a}_{kR} = \hat{a}_k + \frac{s_k}{2} + \frac{n_{k0}R_k(\mu_{k0}-\overline{O}_k)^2}{2(n_{k0}+R_k)}$$

$$s_k = \sum_{r_k=1}^{R_k}\left(o_{k,r_k}-\overline{O}_k\right)^2$$

Further if the reference prior is used,

$$p(\mu_k|D_k) \sim Student\left(\overline{O}_k,\frac{R_k(R_k-1)}{s_k},R_k-1\right) \quad (4)$$

Proofs of fact 3.1 and 3.2 The proofs follow directly from the definition of conjugate distributions (see, e.g. deGroot, 1970 and Inoue, 1998).

Applying these facts, it is possible to find the probability that system i is best, P(CS) for either known or unknown precision case as follows:

$$P(CS) = P(\text{system } i \text{ is best}|D) = P(\mu_i \geq \mu_j \text{ for all } j \neq i|D)$$

where D is the output from all simulation replications.

When there are only two systems considered ($K = 2$), the methodology is equivalent to single system model based on the difference of two means (Inoue, 1998). However Monte Carlo simulation from the posterior marginal distribution of $(\mu_1,\mu_2,...,\mu_K)$ might be required for $K > 2$ to estimate P(CS).

3.2 Dependent Replications

Because of significant savings of computational effort for system analysis, it is often desirable to use the dependent output from system to system such as common random numbers (CRN). In this section we explore a Bayesian approach to account for such dependencies.

We make the following assumptions:

1. Assume that common random numbers are used and the number of replications, R, for each system is the same.
2. Let $o_r = (o_{1,r},o_{2,r},...,o_{K,r})$ be the vector of outputs from r-th replication. The vectors $o_r (r=1,...,R)$ are independent, but the components are not necessarily independent.
3. $o_r = (o_{1,r},o_{2,r},...,o_{K,r})$ is K-dimensional multivariate normal with unknown mean vector μ and precision matrix τ.

$$p(o_r|\mu,\tau) = \left(\frac{\tau}{2\delta}\right)^{K/2}\exp\left[-\frac{1}{2}(o_r-\mu)^t\tau(o_r-\mu)\right]$$

Sample statistics from R simulation replications, the sample K-variant mean and a sample $K \times K$ covariance are denoted:

$$\overline{O} = \sum_{r=1}^{R}\frac{o_r}{R} \qquad S = \sum_{r=1}^{R}(o_r-\overline{O})(o_r-\overline{O})^t$$

When the precision matrix is known, we have the following posterior distribution for the unknown mean vector.

Fact 3.3 *In addition to the above assumptions 1- 3, assume that the precision matrix is known τ. Second, assume that the unknown mean vector μ has conjugate normal distribution with mean vector μ_0 and precision matrix $\tau_0 \in \Re^{K \times K}$. Then the posterior distribution for the mean vector is*

$$p(\mu|D) \sim Normal_K(\mu_R, R\tau + \tau_0) \qquad (5)$$

where $\quad \mu_R = (R\tau + \tau_0)^{-1}(\tau_0\mu_0 + R\tau\bar{O})$

With the reference prior ($\tau_0 \to$ zero matrix, $p(\mu) = c$ for all $\mu \in \Re^K$), it follows that

$$p(\mu|D) \sim Normal_K(\bar{O}, R\tau) \qquad (6)$$

For the unknown precision matrix case, the posterior distribution of the unknown mean performance is determined as follows.

Fact 3.4 *In addition to the above assumptions 1 - 3, assume that the precision matrix is unknown τ. Second assume the appropriate multivariate generalization, a conjugate normal-Wishart distribution: the conditional distribution of unknown mean vector μ is normal distribution with mean vector μ_0 and precision matrix $n_0\tau$ and the marginal distribution of τ is a K dimensional Wishart distribution with parameters $\acute{a}$ and $\hat{a}$ (with $2\acute{a} > K - 1$ and $\hat{a}$ is $K \times K$ a symmetric, nonsingular matrix).*

$$p(\mu, \tau) = \left(\frac{n_o\tau}{2\delta}\right)^{K/2}\exp\left[-\tfrac{1}{2}(\mu - \mu_0)^t n_0\tau(\mu - \mu_0)\right]$$
$$\bullet\, c_w|\hat{a}|^{\alpha}|\tau|^{\acute{a}-(K+1)/2}\exp[-tr(\hat{a}\tau)]$$

where $\quad c_w = \left[\delta^{K(K-1)/4}\prod_{k=1}^{K}\Gamma\left(\frac{2\acute{a}+1-k}{2}\right)\right]^{-1}$

Then the posterior distribution for the mean vector is a multivariate Student distribution,

$$p(\mu|D) \sim Student_K\left(\mu_R, \frac{(n_0 + R)(2\acute{a} + R - K + 1)}{2}\hat{a}_R^{-1}, 2\acute{a} + R - K + 1\right) \quad (7)$$

where $\quad \mu_R = \frac{(n_0\mu_0 + R\bar{O})}{n_0 + R}$

$$\hat{a}_R = \hat{a} + \frac{S}{2} + \frac{n_0 R}{2(n_0 + R)}(\mu_0 - \bar{O})(\mu_0 - \bar{O})^t$$

and the density function of $Student_K(\mu, \hat{e}, \nu)$ is

$$p(x|\mu, \hat{e}, \nu) = c_s\left[1 + \frac{(x-\mu)^t \hat{e}(x-\mu)}{\nu}\right]^{-(\nu+1)/2}$$

where $\quad c_S = |\hat{e}|^{1/2}\tilde{A}((\nu + K)/2)/\left(\Gamma(\nu/2)(\nu\delta)^{K/2}\right)$

With the reference prior, $p(\mu, \tau) = \frac{1}{|\tau|^{(K+1)/2}}$ ($n_0 \to 0$, $\acute{a} \to 0$, $\hat{a} \to$ zero matrix), it follows that

$$p(\mu|D) \sim Student_K\left(\bar{O}, R(R - K)S^{-1}, R - K\right) \qquad (8)$$

Proofs of fact 3.3 and 3.4 The proofs follow directly from the definition of conjugate distributions (see, e.g. deGroot, 1970, Chick, 1997 and Inoue, 1998).

The probability that system i is best, for either known or unknown precision matrix is as follows:

$$P(CS) = P(\mu_i \geq \mu_j \text{ for all } j \neq i | D) = \int_{\mu_i \geq \mu_j \text{ for all } j \neq i} p_{\mu|D}(\mu)$$

where D is the output from all simulation replications.

4 COMPARISON OF BAYESIAN AND FREQUENTIST APPROACHES

In what way, if any, does a P-value for multiple selection compare to the posterior probability correct selection? This is an important question for evaluating whether or not a frequentist approach can be justified from a Bayesian approach. We explore the answer to this question for a collection of such tests for the unknown mean.

4.1 A Single ($K = 1$) System with A Standard

Before considering which is the best of multiple systems, it is helpful to look at the simplest case, $K = 1$. In particular, we want to see if the unknown mean μ exceeds some standard Λ.

In frequentist approach, we could analyze this problem using the one-sided hypothesis test of $H_0 : \mu \geq \Lambda$ versus $H_1 : \mu < \Lambda$. When the precision τ is unknown, a widely used test statistic T is

$$T = \sqrt{\frac{R(R-1)}{s}}(\bar{O} - \Lambda)$$

In this case, P-value is

$$P-value = \Phi_{R-1}\left(\sqrt{\tfrac{R(R-1)}{s}}\left(\overline{O}-\grave{I}\right) \right)$$

This and Section 2.1.3 lead to the following result.

Result 4.1 *Under the above assumptions and approximations with single system model, the Bayesian posterior P(CS) with improper prior is exactly same as the frequentist one-sided P-value.*

This result shows that for one-sided hypothesis testing, it is possible to reconcile Bayesian evidence against null hypothesis using a reference prior with frequentist evidence (also see Casella and Berger, 1987).

When the number of replications is large we have the following results.

Result 4.2 *Under the normality assumption Section 2.1 with single system model, the sizes of Bayesian posterior central credible set and frequentist confidence interval are the same order of magnitude.*

Proof For the unknown precision case, the frequentist approach has the confidence interval whose size is like

$$t_{R-1,\bullet}\left(\sqrt{\tfrac{R(R-1)}{s}} \right)$$

For the analogous Bayesian central credible set with a reference prior (Bernardo and Smith, 1994), the size is also

$$t_{R-1,\bullet}\left(\sqrt{\tfrac{R(R-1)}{s}} \right)$$

On the other hand, Bayesian approach with proper conjugate prior has a central credible set whose width is like (see Equation (3)).

$$t_{2\acute{a}+R,\bullet}\left(\sqrt{\tfrac{(n_0+R)(2\acute{a}+R)}{2\hat{a}+s+\tfrac{n_0 R(\grave{\imath}_0-\overline{O})^2}{n_0+R}}} \right)$$

As $R \to \infty$, this is of order

$$t_{2\acute{a}+R,\bullet}\left(\sqrt{\tfrac{R^2}{s}} \right)$$

As R increase, $t_{2\acute{a}+R,\bullet} \approx t_{R,\bullet} \approx z_\bullet$ and $\tfrac{R(R-1)}{s} \approx \tfrac{R^2}{s}$ to first order, so the result holds.

So if a reference prior is used or the number of replications is very large, they all have confidence/credible sets whose size is the same order of magnitude.

4.2 Two ($K = 2$) Systems

4.2.1 Independent Case

When there are two systems, we would like to know whether system 1 has a larger mean than system 2. Assume that the precisions $\ddot{e}_k$ are unknown and unequal $(\ddot{e}_1 \neq \ddot{e}_2)$, but we have the same number of replications for each system $(R_1 = R_2 = R)$. One frequentist approach considers this problem as the single system model using the difference of two means. Thus we set the hypothesis,

$$H_0 : \grave{\imath}_1 - \grave{\imath}_2 \geq 0 \qquad H_1 : \grave{\imath}_1 - \grave{\imath}_2 < 0$$

A standard test statistic and P-value for H_0 are

$$T = \sqrt{\tfrac{R(R-1)}{s_1+s_2}}\left(\overline{O_1}-\overline{O_2}\right)$$

$$P-value = \Phi_{R-1}\left(\sqrt{\tfrac{R(R-1)}{s_1+s_2}}\left(\overline{O_1}-\overline{O_2}\right) \right)$$

where $\overline{O_k} = \sum_{r=1}^{R} \tfrac{o_{k,r}}{R}$ and $s_k = \sum_{r=1}^{R}\left(o_{k,r}-\overline{O_k}\right)^2$

Consider a special case for the Bayesian approach. Assume normal-gamma conjugate priors for $(\grave{\imath}_1,\ddot{e}_1)$ and $(\grave{\imath}_2,\ddot{e}_2)$: the conditional distribution of $\grave{\imath}_1$ given $\ddot{e}_1$ is normal with mean $\grave{\imath}_{10}$ and precision $n_{10}\ddot{e}_1$, and the marginal distribution of $\ddot{e}_1$ is $Gamma(\acute{a}, \hat{a}_1)$ and the conditional distribution of $\grave{\imath}_2$ given $\ddot{e}_2$ distributed conditionally normal with mean $\grave{\imath}_{20}$ and precision $n_{20}\ddot{e}_2$ with $\ddot{e}_2$ distributed with $Gamma(\acute{a}, \hat{a}_2)$. Then the posterior distribution of $\grave{\imath}_1 - \grave{\imath}_2$ is

$$p(\grave{\imath}_1 - \grave{\imath}_2 / D_1, D_2) \sim Student(\grave{\imath}_R, \ddot{e}_R, 2\acute{a} + R)$$

where $\grave{\imath}_R = \tfrac{n_{10}\grave{\imath}_{10}+R\overline{O_1}}{n_{10}+R} - \tfrac{n_{20}\grave{\imath}_{20}+R\overline{O_2}}{n_{20}+R}$

$$\ddot{e}_R = \left(\tfrac{\hat{a}_{1R}}{(n_{10}+R)(\acute{a}+R/2)} + \tfrac{\hat{a}_{2R}}{(n_{20}+R)(\acute{a}+R/2)} \right)^{-1}$$

With reference prior defined in Section 2.1.3,

$$p(\grave{\imath}_1 - \grave{\imath}_2 / D_1, D_2) \sim Student\left(\overline{O_1}-\overline{O_2}, \tfrac{R(R-1)}{s_1+s_2}, R-1 \right)$$

This leads the following:

Result 4.3 *Given the normality assumptions (Section 2.1), the Bayesian posterior P(CS) with improper prior is exactly same as the frequentist P-value.*

Proof Results follows like Result 4.1 and 4.2.

Inoue (1998) discusses many other cases dealing with two system cases with known precision, unknown but equal precision, and unequal replications.

4.2.2 Dependent Case

For $K = 2$ systems with correlated output, assume that each pair of outputs has a bivariate normal with unknown mean vector and unknown precision matrix. In one frequentist approach, for the one-sided hypothesis, the test statistic is

$$T_d = \sqrt{\frac{R(R-1)}{s_d}}\left(\overline{d}\right)$$

where
$$d_r = o_{1r} - o_{2r} \ (r = 1,..., R)$$
$$\overline{d} = \sum_{r=1}^{R}\frac{(o_{1r}-o_{2r})}{R} = \overline{O_1} - \overline{O_2}$$
$$s_d = \sum_{r=1}^{R}\left(d_r - \overline{d}\right)^2$$

In the frequentist sense, P-value for H_0 is

$$P - value = \Phi_{R-1}\left(\sqrt{\frac{R(R-1)}{s_d}}\left(\overline{d}\right)\right)$$

This is *different* from the analogous Bayesian result even when a reference prior is used.

Result 4.4 *For K = 2, the Bayesian posterior P(CS) with a reference prior has the different degrees of freedom from the frequentist P(CS).*
   Proof From Equation (8) of Fact 3.4,

$$p(\boldsymbol{\imath}\,|\boldsymbol{D}) \sim Student_2\left(\overline{\boldsymbol{O}}, \frac{R(R-2)}{s_d}, R-2\right)$$

The marginal distribution of $\imath_1 - \imath_2$ is also Student with $R-2$ degrees of freedom (deGroot, 1970). The probability that system 1 is the best is then

$$P(CS) = \Phi_{R-2}\left(\sqrt{\frac{R(R-2)}{s_d}}\left(\overline{d}\right)\right)$$

4.3 More Than Two (*K* > 2) Systems

Suppose that we would like to measure evidence that the system i is best. In frequentist approaches, a wide variety of P-values have been used for MCB and multiple hypotheses testing (Hsu, 1996).

Here we use Bonferroni inequality to obtain a conservative P-value for the hypothesis that the system i is best with $(K - 1)$ one-sided hypotheses.

$$H_{0,i,j} : \imath_i \geq \imath_j \quad H_{1,i,j} : \imath_i < \imath_j \quad (j = 1,....,K\!:\!j \neq i)$$

The overall P-value is based on Bonferroni inequality, which gives a lower bound on the probability for the joint occurrence of a finite number of events.

$$P(H_{0,i,j} \ is \ true) = P - value(H_{0,i,j})$$

then

$$P(all \ H_{0,i,j} \ is \ true) \geq 1 - \sum_{j \neq i}^{K}\left[1 - \left(P - value(H_{0,i,j})\right)\right] \quad (9)$$

The use of Bonferroni inequity is familiar in simultaneous statistical inference (Charnes and Kelton, 1988). So we define the right hand side of Inequality (9) as the overall P-value of the joint hypothesis $H_{0,i} : \imath_i \geq \imath_j$ $(j = 1,..., K : j \neq i)$:

$$P - value(H_{0,i}) = 1 - \sum_{j \neq i}^{K}\left[1 - \left(P - value(H_{0,i,j})\right)\right]$$

4.4 General Observations

For independent replications, a Bayesian approach with reference prior and a frequentist approach have the same relevant distributions for pairwise comparisons. For a Bayesian approach with conjugate prior, the sizes of credible sets are asymptotically of the same order of magnitude. So if the posterior P(CS) is approximated by the Bonferroni bound not Monte Carlo simulation, frequentist approach is justified asymptotically (in limit of parameters of priors or the large number of replications) by the Bayesian approach (Inoue, 1998). For unknown precision case, we need $R_k > 1$ replications for each system because both Bayesian and frequentist approaches are based on the Student distribution with $R_k - 1$ degrees of freedom.

For CRN, the story is different. First, to obtain proper posterior distribution, we need $R > (K-1)/2$ replications for each system. As the number of system increases, so does the number of replications required to justifiable

inference. Second, the degree of freedom for hypothesis tests differs for a Bayesian approach with reference prior $R-K$ and a frequentist approach $R-1$. For a small number of replications, this indicates that frequentist assessments of evidence of correct selection are optimistic.

Formally, the reference prior is improper and therefore incoherent behavior. It is not clear that a justification of P-values based on them, therefore, is to be considered advantage for frequentists. The Bayesian P(CS) formulation may be estimated by MC simulation form the posterior for the unknown mean. This is not for frequentist P-values.

5 NUMERICAL EXAMPLE

In this section, we present experimental results from the application of Bayesian and frequentist approach to a typical simulation problem. We consider five (s, S) inventory policies treated by Law and Kelton (1991, pp. 592). The objective here is to determine which inventory system has the minimum expected cost per period for 120 periods.

Table 1: Simulation results for five (s, S) policies

Independent Reps.

System	s	S	Mean	S.D
1	20	40	125.27	4.57
2	20	80	121.28	3.37
3	40	60	125.83	2.64
4	40	100	131.90	2.48
5	60	100	144.65	2.00

Dependent Reps

System	s	S	Mean	S.D
1	20	40	125.29	4.20
2	20	80	120.74	2.66
3	40	60	124.96	2.16
4	40	100	131.10	1.99
5	60	100	143.53	1.66

We run $R = 100$ replications for each of the five systems with both independent and synchronized common random numbers were performed (Table 1). In this example, CRN means that all inventory systems were evaluated with the same times between demands and the same sizes of demands.

We start with independent case. For Fact 3.2, assume that a prior mean for μ_{k0} is an approximate value (Nelson, Schmeiser, Taaffe, and Wang, 1997 used queuing approximation in a similar context) using (Q,R) inventory method (Namhias, 1989): $\mu_{10} = 130$, $\mu_{20} = 120$, $\mu_{30} = 135$, $\mu_{40} = 140$, $\mu_{50} = 155$ and that all systems has

the expected value of precision of the prior distribution of 0.33 with $n_{k0} = 1$, $\acute{a} = 1.5$, and $\hat{a}_k = 4.5$ $(k = 1,..,5)$. Then the posterior distribution have the Student distribution whose mean, precision, degrees of freedom given in Table 2.

Table 2: Posterior Distribution with Independent Case

	Mean	Precision	D.F
$\mu_1 \vert D_1$	125.32	4.70	103
$\mu_2 \vert D_2$	121.27	9.14	103
$\mu_3 \vert D_3$	125.92	10.86	103
$\mu_4 \vert D_4$	131.97	14.97	103
$\mu_5 \vert D_5$	144.75	19.65	103

For dependent replications, we will use the same value for prior for μ :

$$\mu_0 = \begin{bmatrix} 130 & 120 & 135 & 140 & 155 \end{bmatrix}^t$$
$$n_0 = 1$$
$$\acute{a} = 1.5 \quad \hat{a} = diag\begin{bmatrix} 4.5 & 4.5 & 4.5 & 4.5 & 4.5 \end{bmatrix}^t$$

Using the covariance of the output and the above prior distribution, the posterior distribution of μ is the multivariate Student t distribution defined as

$$p(\mu \,/ D) \sim Student_5 \left(\mu_R, \tfrac{(101)(103)}{2} \hat{a}_R^{-1}, 103 \right)$$

where

$$\mu_R = \begin{bmatrix} 125.34 & 120.73 & 125.06 & 131.19 & 143.64 \end{bmatrix}^t$$

$$\hat{a}_R = \begin{bmatrix} 903.06 & 405.84 & 361.53 & 307.44 & 213.95 \\ 405.84 & 355.05 & 209.30 & 216.58 & 113.02 \\ 361.53 & 209.30 & 284.80 & 192.45 & 173.17 \\ 307.44 & 216.58 & 192.45 & 240.47 & 139.18 \\ 213.95 & 113.02 & 173.17 & 139.18 & 206.83 \end{bmatrix}$$

We examined the probability of correct selection given a selection of system 2 as best using (1) estimates of P(CS), $\hat{P}\{-\mu_2 \geq -\mu_j \, (j \neq 2) \vert D\}$, by sampling from the Bayesian posterior distribution and (2) the Bonferroni inequality of frequentist P-values for $-\mu_2 \geq -\mu_j$ $(j \neq 2)$. A summary of results is presented in Table 3.

Table 3: Summary of Comparison

# REPS	Variates	P(CS) (sampling from posterior)	P-value (Bonferroni)
10	Indep.	0.9737	0.8847
	CRN	0.9999	0.9996
30	Indep.	0.9992	0.9953
	CRN	1.0000	1.0000
50	Indep.	0.9999	0.9994
	CRN	1.0000	1.0000
100	Indep.	1.0000	1.0000
	CRN	1.0000	1.0000

Therefore, at the expense of sampling from the posterior distribution of mean, one can obtain tighter assessments than Bonferroni with pairwise comparison. Results were somewhat insensitive to our choice of prior distribution.

REFERENCES

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. Wiley, Chichester, UK.

Casella, G. and R. L. Berger (1987) Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *Journal of the American Statistical Association*, 82: 106-111

Charnes, J. M. and W. D. Kelton (1988) A comparison of confidence region estimators for multivariate simulation output. In M. Abrams, P. Haigh, and J. Comforts (Eds.), *Proceedings of the Winter Simulation Conference*, pp. 458-465. Institute of Electrical and Electronics Engineers Inc.

Chick, S. E. (1997). Selecting the best system: A decision-theoretic approach. In S. Andradottir, K. Healy, D. Whithers, and B. Nelson (Eds.), *Proceedings of the Winter Simulation Conference*. Institute of Electrical and Electronics Engineers Inc.

deGroot, M. H. (1970). *Optimal Statistical Decisions.* New York: MacGraw-Hill, Inc.

Goldsman, D. and B. L. Nelson (1994). Ranking, selection, and multiple comparisons in computer simulation. In J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila (Eds.), *Proceedings of the Winter Simulation Conference*, pp. 192-199. Institute of Electrical and Electronics Engineers Inc.

Gupta, S. S. and K. J. Miescke (1988). On the problem of finding the largest normal mean under hetroscedasity. In S. S. Gupta and J. O. Berger (Eds.), *Statistical Decision Theory and Related Topics IV,* Volume 2. New York: Springer Verlag.

Gupta, S. S. and K. J. Miescke (1994). Bayesian look ahead one-stage sampling allocation for selecting the best population. *Journal of Statistical Planning and Inference,* 54: 229-244.

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. Chapman & Hall, New York

Inoue, K. (1998). Simulation output analysis in a Bayesian decision-theoretic framework. *working Ph.D. dissertation.*

Law A. M. and W. D. Kelton (1991). *Simulation Modeling & Analysis*. McGraw-Hill, Inc, New York, 2nd edition.

Namhias, S. (1989). *Production and Operations Analysis*, Irwin, Homewood, Ill.

Nelson, B. L., B. W. Schmeiser, M. R., Taaffe, and J. Wang (1997). Approximation-assisted point estimation. *Operations Research Letters*, 20: 109-118.

AUTHOR BIOGRAPHIES

KOICHIRO INOUE is a Ph.D. candidate in the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. His working dissertation is application of Bayesian statistics to simulation output analysis. His research also includes decision analysis with multi-attributes.

STEPHEN E. CHICK is an assistant professor of the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. In addition to stochastic simulation, his research interests include engineering probability, Bayesian statistics, reliability, decision analysis, and computational methods in statistics, and applications to health care, particularly epidmiology. His work experience includes several years of material handling system design for the automotive industry using simulation analysis.