Process Capability Indices for the Balanced Random Effects Model

Abrie J. van der Merwe and Delson Chikobvu
Department of Mathematical Statistics
University of Free State, Bloemfontein, 9300, South Africa
e-mail: chikobvu.sci@ufs. ac.za

ABSTRACT

Capability analysis is used in many facets of industrial processes and has been recently introduced into business processes. In this paper a process capability index (P^1_{pl}) is developed for the average of observations from new or unknown batches in the case of a balanced random effects model. Using a Bayesian approach, theoretical and simulation results are derived for the index under two different but related prior distributions. A medical tablet manufacturing example illustrates the flexibility and unique features of the Bayesian simulation method.

Key Words: Bayesian framework, Jeffreys' prior, Medical example, Monte Carlo simulation, Probability matching prior, Process capability index, Process performance index, Reference prior, Variance component model.

1. INTRODUCTION

Process capability indices have been widely used in the manufacturing industry. They measure the ability of a manufacturing process to produce items that meet certain specifications. A capability index relates the voice of the customer (specification limits) to the voice of the process. A large value of the index indicates that the current process is capable of producing items (parts, tablets) that will meet or exceed the customers' requirements. Capability indices are convenient because they reduce complex information about the process to a single number and measure relative variability similar to coefficient of variation.

There is a need to understand and interpret process capability indices. In the literature on statistical quality control there have been some attempts to study the inferential aspects of

these indices. Most of the existing work in this area has been devoted to classical frequentist large sample theory.

As mentioned by Pearn and Wu (2005) a point estimate of the index is not very useful in making reliable decisions. An interval estimation approach is infact more appropriate and widely accepted but the frequency distributions of these estimators are often very complicated which means that the calculation of exact confidence intervals will be difficult.

An alternative approach to the problem of making inferences about capability indices is the Bayesian approach. In the Bayesian approach prior knowledge (or relative ignorance) about the unknown parameters is formally incorporated into the process of inference by assigning a prior distribution to the parameters (Box and Tiao,1973). The information contained in the prior is combined with the likelihood function to obtain the posterior distribution of the parameters. Inferences about the unknown parameters are based on the posterior distribution. If the form of the posterior distribution is complicated, numerical methods or Monte Carlo simulation procedures can be used to solve different complex problems such as credibility intervals (Bayesian confidence intervals), ranking and selection, multiple comparisons and run length for which the frequentist methods are not well developed in the case of capability indices.

There appears to be a general acceptance of the idea that process capability indices can be used only after it has been established that a process is in “statistical control” (for example by the use of control charts). This is reasonable, if it is simply required that there are no irregular changes in quality level. Kotz and Johnson (1993), Herman (1989) and Wolfinger (1998) have drawn attention to the fact that there can be more than one source of variation.

Data arising from multiple sources of variability are very common in practice. Virtually all industrial processes exhibit *between-batch*, as well as *within-batch* components of variation. In some cases the between-batch (or between subgroup) component is viewed as part of the common-cause-system for the process. It therefore seems worthwhile to develop a process capability index in more general settings. To do so, it is necessary to employ a statistical model which adequately handles multiple sources of variability. The variance component model is suitable for this task.

In this paper we look at a version of the most popular process capability index C_{pk} for the balanced random effects model using a Bayesian approach. The process performance index is denoted by P_{pl}^{1} and can be used for averages of observations from new or unknown batches.

To illustrate how and when the index will be used, consider a factory which manufactures medical tablets in very small batches. A small batch in this instance is likely to be a weekly or monthly intake of tablets for an individual patient. The interest is in whether the patient gets on average the required dosage of the drug from the batch in the specified time, given that each patient must get an average dosage of at least ℓ_0. The question therefore is whether the process is capable of producing to this specification.

In the next section definitions, notation and some indices commonly used in process capability analysis are reviewed. In section 3, a Bayesian analysis of the random effects model is considered and a lower performance index (P_{pl}^{1}) is proposed which is based on averages of observations from new or unknown batches. Although there are a few published articles dealing with the Bayesian estimation of capability indices, see for example Pearn and Wu (2005); Lin, Pearn and Yang (2005); Wu and Pearn (2005) and Wu (2007), no one (as far as we know) has worked on capability (performance) indices in the case of the random effects model. An application is provided in section 4. Determination of reasonable non-informative priors in multi-parameter problems is not easy. Common non-informative priors such as Jeffreys' prior can have features that have an unexpectedly dramatic effect on the posterior. In Section 5 reference and probability matching priors are therefore derived for the lower process performance index (P_{pl}^{1}) .In section 6 a weighted Monte Carlo method is described to simulate (P_{pl}^{1}) using the probability matching (reference) prior. This method is especially suitable for computing credibility intervals. The conclusion is given in section 7.

2. DEFINITIONS AND NOTATIONS

Let Y be some characteristics of interest of a manufactured product. The engineering or design specifications for Y are generally stated in terms of a 'nominal' or a 'target value', say T. That is, T is the value of Y which will satisfy the design engineer's criteria for the optimum performance of the product. Now manufacturing the product so that Y exactly equals T is prohibitively expensive, and so it is common practice to specify upper and lower 'specifications' limits, USL and LSL, or simply ℓ_1 and ℓ_0 respectively, and to require that Y be within these limits.

The physical processes that manufacture the part are generally subject to many sources of variation, starting from the quality of raw material to the aging and wear-out of the manufacturing equipment. Consequently, Y is a random quantity (or a random variable), whose distribution if often assumed to be Gaussian with mean, say μ, and a variance, say σ^2. In manufacturing parlance, the variance is referred to as the *natural tolerance* of Y. When working with the process capability indices it is common practice to assume that both μ and σ^2 do not change with time; i.e. the process is stable, or what is known in quality control as being in statistical control.

The question which arises is as to whether the design engineer's compromise in going from the ideal T to the upper and lower specifications limits (the USL and the LSL), is matched by the manufacturer's ability to meet such a compromise vis-á-vis the assumed μ and σ^2 mentioned above. The process capability indices were introduced to address this matter. The quantity $(USL - LSL)$ is known as the specification interval (or *tolerance interval*); it will be denoted by $2d$, where d is the half length of the specification interval. The midpoint of the specification interval, which will be denoted by M, is equal to $(USL + LSL)/2$.

Herman (1989) provided a thought-provoking criticism of the process capability index (PCI) concept, based on engineering considerations. The σ is intended to represent process variability when production is 'in control'. But usually variation often has two components - from *within*-lots and *among*-lots variation. The σ in the denominator of for example

$C_{pk} = \min\left(\frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma}\right)$ is intended to refer to *within*-lot process variation. This σ can be considerably less than the overall standard deviation - σ_{total}, say. The variance component model is used when the variation of group means is more than expected when using a simple within group variance model. Process parameters are known to vary slightly, even when the process is in statistical control, introducing extra variation σ_2^2. The definition of C_{pk} includes as special cases those processes for which only one limit exists, by letting $LSL \to -\infty$ or $USL \to \infty$ in which case it reduces to the appropriate standardized measure. Herman suggests that a different index, the 'process performance index' (PPI), $P_{pk} = \min\left(\frac{USL-\mu}{3\sigma_{total}}, \frac{\mu-LSL}{3\sigma_{total}}\right)$ might 'have more value to a customer than C_{pk}. The measures C_{pk} and P_{pk} again differ only in the estimate of the process standard deviation.

The random effects (variance component) model is suitable for handling multiple sources of variability. In the next sections we will look at the random effects model from a Bayesian perspective. The Bayesian approach is conceptually more appealing than the classical approach since, it allows explicit use of prior information, thereby giving new insights in problems where classical statistics fail.

By using Monte Carlo simulation, random draws from the posterior distribution of the quantities of interest are used to construct the needed inferences. Histograms of the simulations can be constructed. This is precisely the advantage of the sampling based Bayesian approach, where one can create the posterior distributions (in the form of a histogram) based on the samples and hence do inference from the posterior distribution without going through the exact distribution. From the distributions of the performance indices, we are in a positon to obtain quantiles, credible regions and perform other inferential tasks eg. single summary measures of process performance indices. The methods can be generalized to more complicated situations. This however requires computational resourses. The recent increases in the availability of computational resources and the development of computational techniques have led to great advances in the application methods to complicated problems in various disciplines.

3. THE RANDOM EFFECTS MODEL

The random effects (variance component) model with two variance components is of the form:

$$Y_{ij} = \mu + r_i + \varepsilon_{ij} \quad \text{for } i = 1, 2, \cdots, I \text{ and } j = 1, 2, \cdots, J. \tag{3.1}$$

The random variables r_i and ε_{ij} are called random effects and the model in (3.1) is known as the balanced random effects model. Furthermore it is assumed that $\varepsilon_{ij} \sim N(0, \sigma_1^2)$ and $r_i \sim N(0, \sigma_2^2)$.

In addition the single fixed effect μ denotes the overall mean and the random effect r_i denotes the deviation from this mean, specific to batch i. ε_{ij} represents the within group/batch variation. Y_{ij} is known and denotes the j^{th} response value in the i^{th} batch.

From equation (3.1) (see also Box and Tiao 1973, equation (5.2.7)) it follows that the integrated likelihood function is

$$L(\mu, \sigma_1^2, \sigma_2^2 \,|\, \underline{Y}) \propto (\sigma_1^2)^{-\frac{1}{2}\nu_1} (\sigma_1^2 + J\sigma_2^2)^{-\frac{1}{2}(\nu_2+1)} \exp\left\{ -\frac{1}{2}\left[\frac{IJ(\mu - \bar{\bar{Y}}_{..})^2}{\sigma_1^2 + J\sigma_2^2} + \frac{\nu_2 m_2}{\sigma_1^2 + J\sigma_2^2} + \frac{\nu_1 m_1}{\sigma_1^2} \right] \right\} \tag{3.2}$$

where

$$\underline{Y} = [Y_{11,} Y_{12,} \cdots Y_{1J,} Y_{21,} \cdots Y_{2J,} \cdots\cdots Y_{I1,} \cdots Y_{IJ}]'$$

is the $IJ \times 1$ known vector of observed response values.

$\sigma_1^2 > 0, \quad \sigma_2^2 > 0, \quad \sigma_{12}^2 > \sigma_1^2$ and $\sigma_{12}^2 = \sigma_1^2 + J\sigma_2^2$.

The restriction $\sigma_{12}^2 > \sigma_1^2$ is actually part of the prior support given in equation (3.1).

I is the number of groups/batches,

J is the number of observations within each group,

$\nu_1 = I(J-1), \quad \nu_2 = I-1,$

$\nu_1 m_1 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{i.})^2$ is residual sum of squares,

$\nu_2 m_2 = J\sum_{i=1}^{I}(\overline{Y_{i.}} - \overline{\overline{Y}}_{..})^2$ is between groups sum of squares,

$\overline{Y}_{i.} = \frac{1}{J}\sum_{j=1}^{J} Y_{ij}$ the i^{th} batch (group) mean

and

$\overline{\overline{Y}}_{..} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}$ is the overall sample mean.

3.1 Posterior Distributions of μ and the Variance Components:

The first step in a Bayesian approach is to find a prior distribution that summarizes a priori uncertainty about the likely values of the parameters $\mu, \sigma_1^2, \sigma_2^2$. The prior distribution needs to be formulated based on prior knowledge. This is usually a difficult task because such prior knowledge may not be available. In such situations, usually a "non-informative" prior distribution is used. The basic idea behind formulating such a prior distribution is that it should be non-informative so that the likelihood (the density $p(\underline{Y}|\mu, \sigma_1^2, \sigma_2^2)$, evaluated at the observed value of $\underline{Y}$) plays a dominant role in the construction of a posterior density. Jeffreys (1961) formulated such prior distributions based on certain invariance arguments.

The non-informative joint prior for the variance component model as defined by Box and Tiao (1973), page 251, viz:

$$p(\mu, \sigma_1^2, \sigma_2^2) \propto \frac{1}{\sigma_1^2(\sigma_1^2 + J\sigma_2^2)} \qquad (3.3)$$

The prior may easily be obtained by applying Jeffreys' rule. Jeffreys' rule states that the prior distribution for a set of parameters is taken to be proportional to the square root of the determinant of the Fisher information matrix. Equation (3.3) is obtained from the Fisher information matrix for (σ_1^2, σ_2^2), i.e. treating the location parameter μ separately from the variance components. By combining the prior with the likelihood, the joint posterior distribution of μ, σ_1^2 and σ_2^2 can be obtained,

$$p(\mu|\underline{Y}, \sigma_1^2, \sigma_2^2) = \frac{1}{\sqrt{2\pi \frac{(\sigma_1^2 + J\sigma_2^2)}{IJ}}} \exp\left(-\frac{1}{2} \frac{IJ(\mu - \overline{\overline{Y}}_{..})^2}{(\sigma_1^2 + J\sigma_2^2)} \right) \tag{3.4}$$

i.e. $\mu|\underline{Y}, \sigma_1^2, \sigma_2^2 \sim N\left(\overline{\overline{Y}}_{..}, \frac{(\sigma_1^2 + J\sigma_2^2)}{IJ} \right)$.

The joint posterior distribution of the variance components σ_1^2, σ_2^2 is

$$p(\sigma_1^2, \sigma_2^2|\underline{Y}) \propto (\sigma_1^2)^{-\frac{1}{2}(\nu_1+2)} (\sigma_1^2 + J\sigma_2^2)^{-\frac{1}{2}(\nu_2+2)} \exp\left\{ -\frac{1}{2}\left[\frac{\nu_1 m_1}{\sigma_1^2} + \frac{\nu_2 m_2}{(\sigma_1^2 + J\sigma_2^2)} \right] \right\} \tag{3.5}$$

where as mentioned $\sigma_1^2 > 0, \quad \sigma_2^2 > 0, \quad \sigma_{12}^2 > \sigma_1^2$ and $\sigma_{12}^2 = \sigma_1^2 + J\sigma_2^2$.

If the restriction $\sigma_{12}^2 > \sigma_1^2$ did not apply, then the posterior distribution for σ_1^2 and σ_{12}^2 would be independent, each proportional to an inverse gamma distribution. The joint posterior distribution for σ_1^2 and σ_{12}^2 would be the product of these two distributions:

$$p(\sigma_1^2, \sigma_{12}^2|\underline{Y}) \propto (\sigma_1^2)^{-\frac{1}{2}(\nu_1+2)} \exp\left\{ -\frac{1}{2} \frac{\nu_1 m_1}{\sigma_1^2} \right\} \times (\sigma_{12}^2)^{-\frac{1}{2}(\nu_2+2)} \exp\left\{ -\frac{1}{2} \frac{\nu_2 m_2}{\sigma_{12}^2} \right\} \quad . \tag{3.6}$$

However the restrictions do apply. Nevertheless, using a two-step rejection sampling procedure it is straight forward to generate samples from the joint posterior distribution.

(a) Generate values from the two inverse gamma distributions.

(b) Retain those sets of values that conform to the restricted parameter space $\sigma_{12}^2 > \sigma_1^2$.

(c) Substitute each pair of simulated values (σ_1^2, σ_2^2) in equation (3.4) to simulate μ.

(d) Repeat steps (a), (b) and (c) until $\tilde{\ell}$ permissible values are obtained. For our example $\tilde{\ell}$ was taken as 10 000.

The illustrated Monte Carlo simulation procedure is preferable to Gibbs sampling since it generates independent samples from the joint posterior distribution.

3.2 Posterior Distribution of the lower Process Performance Index (P^1_{pl}):

Consider a new (future or unknown) batch of J observations $Y_{f1,}Y_{f2,}\cdots Y_{fJ}$.

Each observation is normally distributed with mean μ and variance $\sigma_1^2 + \sigma_2^2$. In other words

$$Y_{fj}\Big|\mu,\sigma_1^2,\sigma_2^2 \sim N(\mu,\sigma_1^2+\sigma_2^2) \quad (j=1,2,\cdots J) \tag{3.7}$$

and

$$\overline{Y}_{f.}\Big|\mu,\sigma_1^2,\sigma_2^2 \sim N(\mu,\frac{\sigma_1^2+J\sigma_2^2}{J}) \tag{3.8}$$

where

$$\overline{Y}_{f.}=\frac{1}{J}\sum_{j=1}^{J}Y_{fj},$$

the arithmetic mean of the new sample. Equation (3.8) therefore describes the distribution of averages from new or unknown batches.

From (3.8) a lower process performance index, $P_{pl}^{\ 1}$ can be defined as

$$P_{pl}^{\ 1} = \frac{\mu - l_0}{3\left(\frac{\sigma_1^{\ 2} + J\sigma_2^{\ 2}}{J}\right)^{\frac{1}{2}}} \tag{3.9}$$

where

$\mu =$ mean of future observations from a new or unknown batch

$l_0 =$ lower specification limit

$J =$ batch size.

Using the Bayesian simulation procedure as described in section 3.1 an approximation of the exact posterior distribution of $P_{pl}^{\ 1}$ can be obtained. As far as we know a posterior analysis for this form of index does not exist.

To illustrate how and when this index (equation (3.9) will be used, consider a factory that manufactures medical tablets in very small batches. A small batch in this instance is likely to be a weekly or monthly intake of tablets for an individual patient. The interest is in whether the patient gets on average the required dosage of the drug from the batch in the specified time, given that each patient must get an average dosage of at least l_0. The question therefore is whether the process is capable of producing to this specification.

The above mentioned index will be contrasted with the following index:

$$P_{pl} = \frac{\mu - l_0}{3\left(\sigma_1^{\ 2} + J\sigma_2^{\ 2}\right)^{\frac{1}{2}}} \tag{3.10}$$

This index follows from equation (3.7) and is not depended on J, the subgroup size. This index now asses whether the process is capable of producing **each** tablet to specification as opposed to **mean** of the batch.

4. AN APPLICATION

The Bayesian simulation procedure will now be applied to the following data set. The data in the Table 4.1 below are amounts of drug per tablet measurements. The data are assumed to arise from a normal distribution with unknown parameters, but it has more structure than a simple random sample because it is clustered in fifteen batches and each batch contains ten tablets.

Table 4.1 Amount of drug per tablet

Batch	Measurements									
1	150.52	150.39	150.31	150.49	150.47	150.67	150.17	150.45	150.42	150.37
2	150.35	150.47	150.72	150.56	150.53	150.62	150.60	150.52	150.51	150.63
3	150.48	150.79	150.63	150.46	150.71	150.67	150.70	150.48	150.48	150.58
4	150.41	150.45	150.40	150.33	150.24	150.39	150.28	150.36	150.27	150.33
5	150.58	150.54	150.30	150.54	150.50	150.32	150.58	150.46	150.41	150.49
6	150.49	150.83	150.66	150.63	150.72	150.79	150.64	150.62	150.71	150.73
7	150.33	150.44	150.48	150.34	150.50	150.42	150.37	150.54	150.39	150.52
8	150.39	150.52	150.35	150.52	150.47	150.54	150.51	150.37	150.54	150.53
9	150.64	150.78	150.51	150.69	150.51	150.47	150.60	150.50	150.69	150.72
10	150.61	150.49	150.60	150.50	150.68	150.56	150.59	150.73	150.62	150.62
11	150.48	150.25	150.49	150.43	150.40	150.44	150.31	150.36	150.30	150.40
12	150.35	150.41	150.36	150.39	150.34	150.37	150.51	150.32	150.25	150.32
13	150.54	150.67	150.57	150.45	150.57	150.48	150.39	150.38	150.67	150.42
14	150.41	150.54	150.57	150.73	150.47	150.72	150.72	150.49	150.66	150.58
15	150.60	150.45	150.66	150.72	150.45	150.51	150.69	150.62	150.55	150.45

The lower specification limit is $l_0 = 150$. The data and above limit is selected solely for illustrative purposes. In practice, fixed in advance limits are often determined from medical or regulatory considerations. See for example Wolfinger (1998). Based on the data, the quantities needed for the simulation procedure are $I = 15, \quad J = 10, \quad \nu_1 = I(J-1) = 135,$

$\nu_2 = I - 1 = 14, \quad \overline{\overline{Y}}_{..} = 150.5076, \quad \nu_1 m_1 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{i.})^2 = 1.26552, \quad \nu_2 m_2 = J\sum_{i=1}^{I}(\overline{Y}_{i.} - \overline{\overline{Y}}_{..})^2 = 1.469816$

Ten thousand $(\mu, \sigma_1^2, \sigma_2^2)$ values that met the restriction $\sigma_{12}^2 > \sigma_1^2$ were simulated from the posteriors (3.6) and (3.4). In Figures 4.1 and 4.2 histograms of the posterior distributions of σ_1^2 and σ_2^2 are illustrated. The histogram of intraclass correlation coefficient $\rho = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$ is illustrated in Figure 4.3. The means, medians and 95% credibility (Bayesian confidence) intervals are also given.

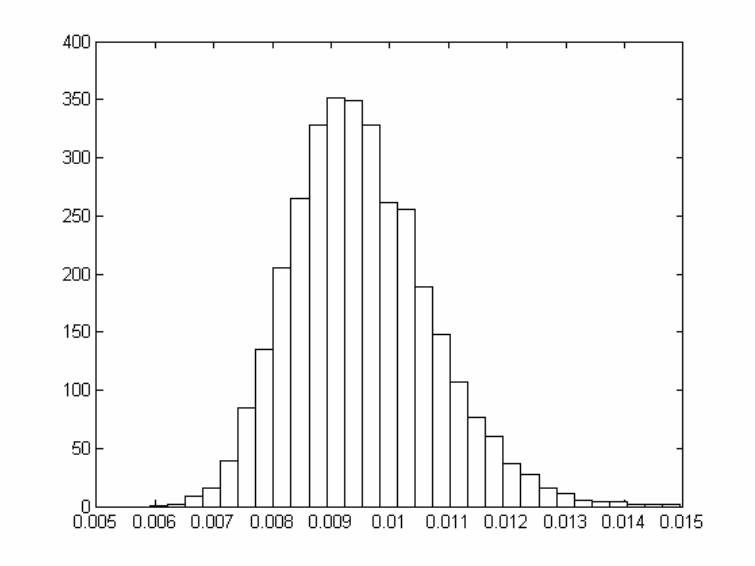
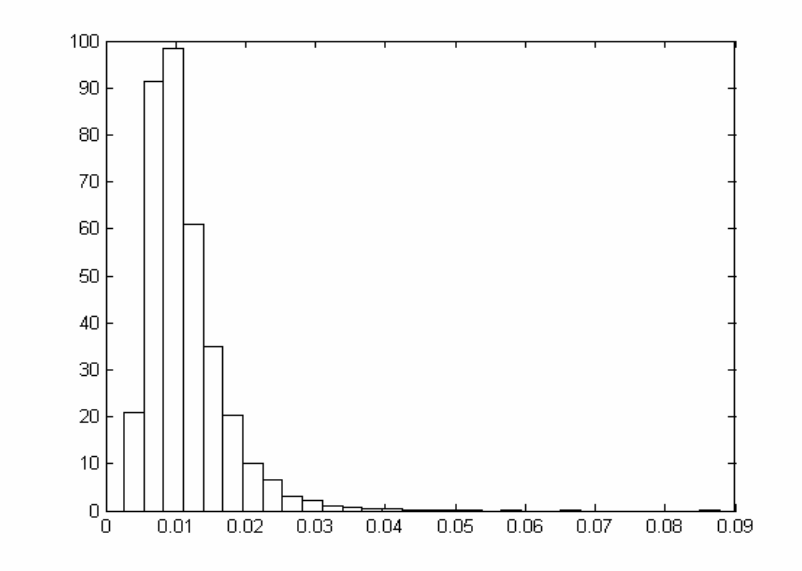


Figure 4.1: Posterior Distribution of σ_1^2 (error variance).
Mean:0.0095, Median:0.0094, 95% Credibility Interval is [0.0075;0.0121].

As is often the case the posterior distribution of σ_1^2, the within batch variance, is quite symmetrical. The reason for this is the large number of degrees of freedom, $\nu_1 = 135$ associated with it. The 95% credibility interval is reasonably small. It has to be remembered that the variance σ_1^2 is a squared entity and that the corresponding interval for the standard deviation σ_1 is [0.0867; 0.1100], which can easily be interpreted. The posterior distribution of σ_2^2 (the between batch variance) on the other hand, is quite skew. The reason for this is the small number of degrees of freedom $(\nu_2 = 14)$ associated with it.

$p(\sigma_2^2|\underline{Y})$



σ_2^2

Figure 4.2: Posterior Distribution of σ_2^2 (between batch variance).

Mean:0.0113, Median:0.0100, 95% Credibility Interval is [0.0047;0.0253].

Although the point estimates of, σ_1^2 and σ_2^2 are quite similar, the 95% credibility interval for σ_2^2, which is an indication of uncertainty in the true value of σ_2^2 is more than four times as large as the corresponding interval for σ_1^2. The reason for this is, as mentioned, the small number of batches included in the experiment. This is also the reason for the large credibility interval for the intraclass correlation coefficient, ρ.

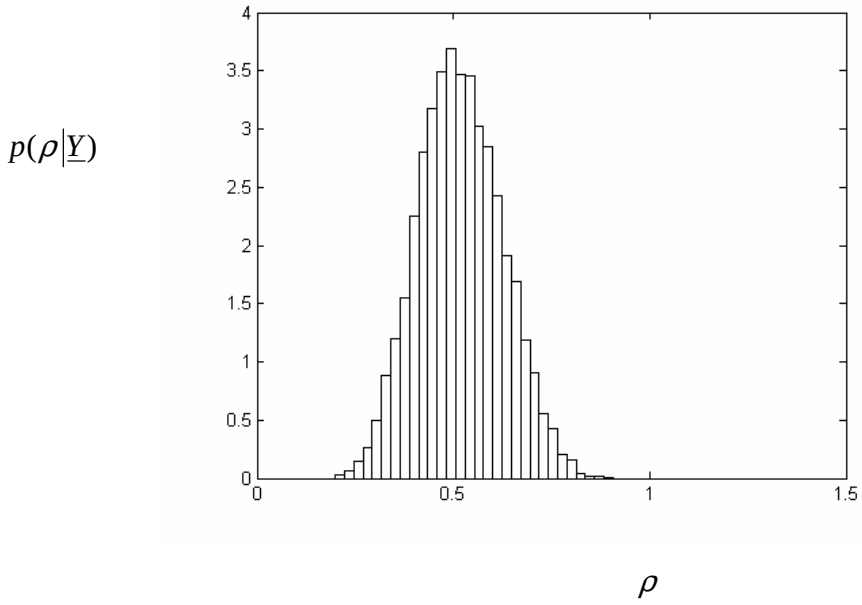


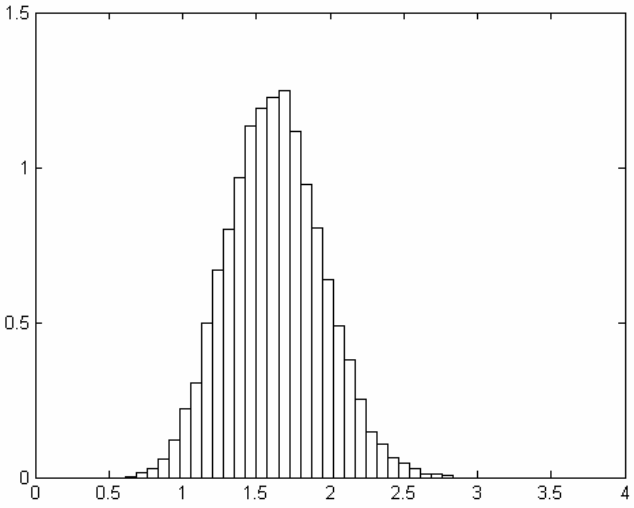
Figure 4.3: Posterior Distribution of ρ (intraclass correlation coefficient).

Mean:0.5202, Median:0.5165, 95% Credibility Interval is [0.3194;0.7382].

In Figures 4.4 and 4.5 the posterior distributions of the process performance indices $P_{pl}{}^{1} = \frac{\mu - l_0}{3\left(\frac{\sigma_1^2 + J\sigma_2^2}{J}\right)^{\frac{1}{2}}}$ and $P_{pl} = \frac{\mu - l_0}{3\left(\sigma_1^2 + J\sigma_2^2\right)^{\frac{1}{2}}}$ are displayed as histograms.

The posterior distributions look quite symmetrical. This is understandable because, conditional on the variance components, $P_{pl}{}^{1}$ and P_{pl} are normally distributed.

$p(P_{pl}^1 | \underline{Y})$

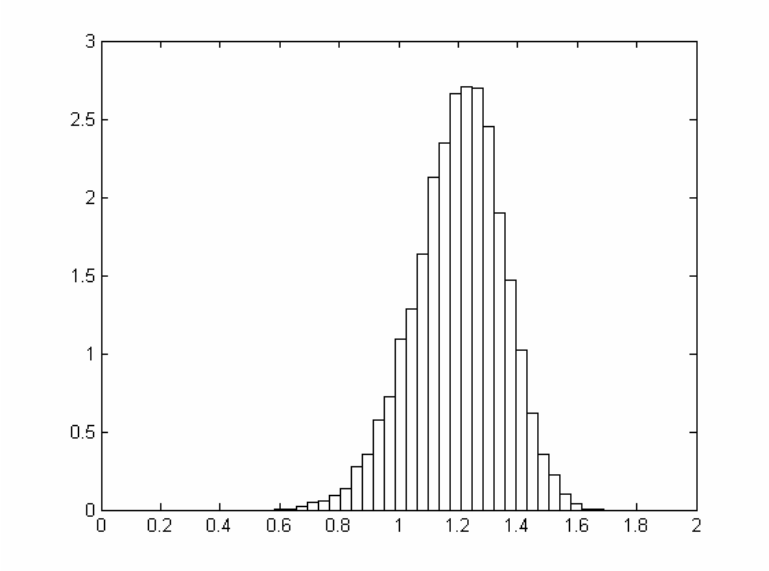


P_{pl}^1

Figure 4.4: Posterior Distribution of Process Performance Index, P_{pl}^1 .

Mean:1.6183, Median:1.6182, 95% Credibility Interval is [1.0100;2.2699].

$p(P_{pl} | \underline{Y})$



P_{pl}

Figure 4.5: Posterior Distribution of Process Performance Index, P_{pl} .

Mean:1.1964, Median:1.2061, 95% Credibility Interval is [0.8663;1.4710].

Despite the widespread use of capability indices in industry and some good review articles such as Gunter (1989 a,b,c,d) there is much confusion and misunderstanding regarding their interpretation and appropriate use. According to Kotz and Johnson (1993) the issue does not generally lie in the validity of the mathematics of the indices, but in their application by those who believe the values are deterministic, rather than random variables. Once the variability is understood and the bias (if any) is known, the use of these indices can be more constructive.

Following Steiner, Bovas and Mackay (1997), the minimum reporting requirements for variables data should be a control chart with limits to show the nature of stability and a process performance index to compare performance to specifications. The performance index P_{pk} is preferable to C_{pk} since it captures all the process variation. They also mentioned that minimum default capability requirements for most characteristics could be given in a simple statement such as $P_{pk} > 1.33$.

From Figures 4.4 and 4.5 it can be seen that the mean of P_{pl}^{1} is 1.62 and for P_{pl} it is 1.21. Therefore according to Steiner etal (1997) patients will get on average, the required dosage of the drug but there is some doubt whether the manufacturing process is capable of producing each tablet to specification.

In table 4.2 certain probabilities (relative frequencies) are given. These probabilities are obtained from the Monte Carlo simulation method.

Table 4.2 Probabilities for specific or larger process performance index values

$P(P_{pl}^{1} > 0) = \frac{10000}{10000} = 1.000$	$P(P_{pl} > 0) = \frac{10000}{10000} = 1.0000$
$P(P_{pl}^{1} > 1.00) = \frac{9770}{10000} = 0.9770$	$P(P_{pl} > 1.00) = \frac{8945}{10000} = 0.8945$
$P(P_{pl}^{1} > 1.33) = \frac{8108}{10000} = 0.8108$	$P(P_{pl} > 1.33) = \frac{1925}{10000} = 0.1925$

Although the average $P_{pl}{}^{1}$ value is 1.62 it is highly probable to get in future, values smaller than 1.33 $(P(P_{pl}^{1} \le 1.33) = 0.189)$ and its not impossible to get $P_{pl}{}^{1}$ values smaller than 1 $(P(P_{pl}^{1} \le 1.00) = 0.023)$.

Typically a "future" batch of $J = 10$ tablets is taken repeatedly from the process, and it is of interest to assess the distribution of the "run length", that is the number of such batches, r, until the control chart signals for the first time. (Note that r here does not include the batch when the control chart signals). Given θ (where for example $\theta = P(P_{pl}^{1} \le 1.00)$) and a stable process, the distribution of the run length r is geometric with parameter θ.

The mean and variance of r is $E(r|\underline{Y}) = \frac{1}{\theta}$ and $Var(r|\underline{Y}) = \frac{1-\theta}{\theta^2}$. In the case of $\theta = P(P_{pl}^{1} < 1.00)$, the average run length is $(0.023)^{-1} = 43$ batches, and for $\theta = P(P_{pl}^{1} < 1.33)$ the average run length is $(0.189)^{-1} = 5$ batches. From Figures 4.4 and 4.5 and Table 4.2 it is however clear that although the values of the indices can easily go below 1.33 and in some cases even below 1.00, it is highly unlikely that they will become negative. This means that very seldom the amount of drug in a tablet will be less than the lower specification limit of $l_0 = 150$.

4.1 The Process Performance Index P^{1}_{pk} :

In the previous section our interest was in whether a patient gets on average the required dosage of a drug from a batch in the specified time. If the avoidance of an over-dose to a patient would be as important as avoiding an under-dose then both problems could be assessed simultaneously using the performance index

$$P_{pk}{}^{1} = \min\left(\frac{l_1 - \mu}{3\left(\frac{\sigma_1^{\,2} + J\sigma_2^{\,2}}{J}\right)^{\frac{1}{2}}}, \frac{\mu - l_0}{3\left(\frac{\sigma_1^{\,2} + J\sigma_2^{\,2}}{J}\right)^{\frac{1}{2}}} \right)$$

which is a version of the most popular process capability index C_{pk}. Let the upper specification limit be $l_1 = 151$, which means that on average a person should not get more than an amount of 151 of the dosage. In Table 4.3, measures of location and dispersion for the posterior distribution of $P_{pk}{}^{1}$ are given, and in Table 4.4 certain probabilities (relative frequencies) are given. As before these probabilities are obtained from the Bayesian simulation method.

Table 4.3 Measures of location and dispersion for the posterior distribution of P^1_{pk}

Index	Mean	Variance	Median	95% credibility Interval
P^1_{pk}	1.533544	0.094221	1.528683	(0.968975;2.159735084)

Table 4.4 Probabilities for specific and larger process performance index values

$P(P^1_{pk} > 0) = \frac{10000}{10000} = 1.000$
$P(P^1_{pk} > 1.00) = \frac{9661}{10000} = 0.9661$
$P(P^1_{pk} > 1.33) = \frac{7402}{10000} = 0.7402$

It is obvious that the mean of P^1_{pk} will always be smaller than that of P^1_{pl}. From Table 4.4 the same conclusion can be made as before namely that although the value of P^1_{pk} can easily go below 1.33 and in some cases below 1, it is unlikely that it will become negative. It therefore seems unlikely that patient on average will be under-dosed or over-dosed.

4.2 A Comparison of Process Performance Indices:

In this section the problem of comparing two process performance indices $P^1_{pl(1)}$ and $P^1_{pl(2)}$ is addressed. Such a problem occurs when selection between two manufacturers or where

assessing the impact of process improvement. In the case of comparing two manufacturers, the objective is to assess the abilities of the manufacturers to meet the medical requirements.

The problem that will be considered is the test for the hypotheses

$$H_0 : P^1_{pl(1)} = P^1_{pl(2)} \quad vs \quad H_a : P^1_{pl(1)} \neq P^1_{pl(2)}$$

based on the $100(1-\frac{\alpha}{2})\%$ two-sided Bayesian confidence (credibility) interval for the difference $D_{12} = P^1_{pl(1)} - P^1_{pl(2)}$ where $\alpha = 0.05$.

Reject H_0 if the lower limit is larger than zero or the upper limit is smaller than zero.

We will consider two manufacturers. The sample data for manufacturer 1 is given in Table 4.1 and the summary statistics for a sample of the same size from manufacturer 2 are

$$I = 15, \quad J = 10, \quad \nu_1 = I(J-1) = 135, \qquad \nu_2 = I - 1 = 14, \quad \overline{\overline{Y}}_{..} = 150.49255, \quad l_0 = 150,$$

$$\nu_1 m_1 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \overline{Y}_{i.})^2 = 1.1977335, \qquad \nu_2 m_2 = J\sum_{i=1}^{I}(\overline{Y_{i.}} - \overline{\overline{Y}}_{..})^2 = 1.3969914.$$

The mean of the second sample $(\overline{\overline{Y}}_{..} = 150.49255)$ is somewhat closer to the lower limit $(l_0 = 150)$ than that of the first sample $(\overline{\overline{Y}}_{..} = 150.5076)$ but the sum of squares $\nu_1 m_1$ and $\nu_2 m_2$ for the second sample are somewhat smaller.

In Figure 4.6 the posterior distribution of the difference $D_{12} = P^1_{pl(1)} - P^1_{pl(2)}$ is illustrated.

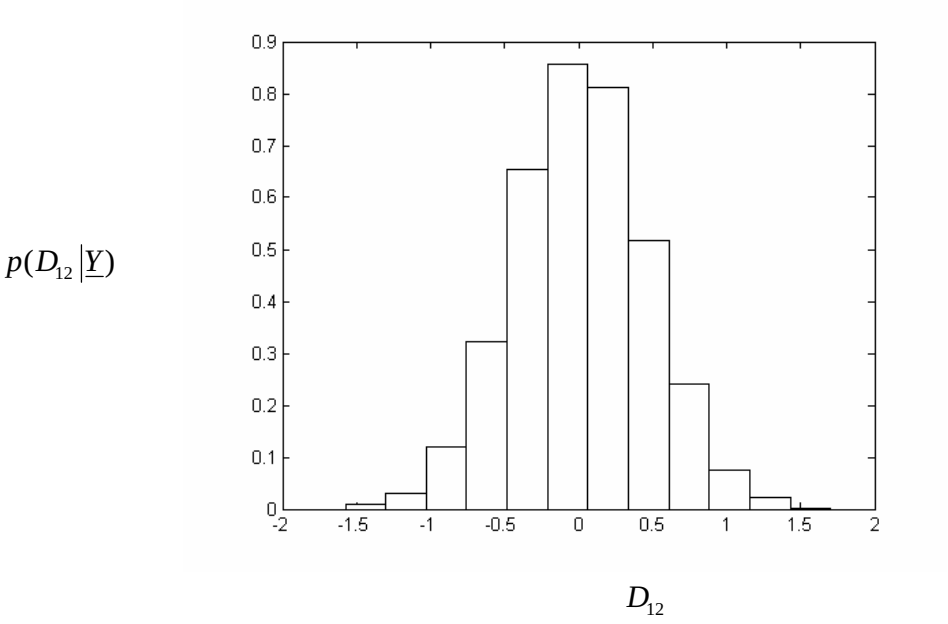


Figure 4.6: Posterior Distribution of the Difference between two Process Performance Indices $D_{12} = P^1_{pl(1)} - P^1_{pl(2)}$.

Mean:0.016006, Median:0.016007, 95% Credibility Interval is [-0.8681;0.9007].

From the figure it is clear that zero is included in the 95% credibility interval. H_0 will therefore not be rejected. It seems that there is no real difference between the two manufacturers. If a choice has to be made between the two manufacturers it will be manufacturer 1. The reason for this is that $E(D_{12}|\underline{Y}) = 0.016$ which is positive.

The problem of selecting the best manufacturer can also be looked at from a ranking and selection perspective. In the past 30 years, beginning with the fundamental papers of Bechhofer (1954) and Gupta (1956), ranking and selection procedures have been developed to overcome the inadequacy of testing procedures.

From a Bayesian point of view ranking and selection is quite simple. To calculate the probability that manufacturer 1 is the best or second best, ranks are assigned to each simulation of the process performance indices $P^1_{pl(1)}$ and $P^1_{pl(2)}$. The higher $P^1_{pl(1)}$ value is assigned the rank of 1 and the smaller one the rank of 2. Repeating the simulation ranking

procedure 10 000 times it was found that in 51.28% of the cases manufacturer 1 was ranked first and in 48.72% of the cases manufacturer 2. Manufacturer 1 is therefore somewhat better than manufacturer 2.

5. PROBABILITY MATCHING AND REFERENCE PRIORS FOR THE LOWER PROCESS PERFORMANCE INDEX P^1_{pl}

Probability matching and reference priors often lead to procedures with good frequency validity while retaining the Bayesian flavour. The fact that the resulting Bayesian posterior intervals of level $1-\alpha$ are also good frequentist confidence intervals at the same level is a very desirable situation. See also Bayarri and Berger (2004) and Sevirini, Mukerjee and Ghosh (2002) for general discussion.

5.1 The Probability Matching Prior for the Lower Process Performance Index, P^1_{pl} :

As mentioned the Bayesian paradigm emerges as attractive in many types of statistical problems - especially in capability and performance index problems but the choice of an appropriate non-informative prior distribution has been controversial. Common non-informative priors in multi-parameter problems, such as Jeffreys' prior, can have features that have an unexpectedly dramatic effect on the posterior. Recently Datta and Ghosh (1995) derived the differential equation that a prior must satisfy if the posterior probability of a one sided credibility interval for a parametric function and its frequentist probability agree up to $0(n^{-1})$ where n is the sample size. Using the method of Datta and Ghosh (1995), the following theorem can be proved:

Theorem 5.1

The probability matching prior for the P^1_{pl} Index in the case of the balanced random effects model defined in equation (3.1) is:

$$\pi_a(\underline{\theta}) = \pi_a(\mu, {\sigma_2}^2, {\sigma_1}^2) \propto {\sigma_1}^{-2}\left({\sigma_1}^2 + J{\sigma_2}^2\right)^{-\frac{3}{2}}\left(1 + \frac{J(\mu - l_0)^2}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)}\right)^{-\frac{1}{2}}$$

where $\underline{\theta} = [\mu, \sigma_2^2, \sigma_1^2]'$.

Proof

The proof is given in the Appendix.

$\pi_a(\mu, \sigma_2^2, \sigma_1^2)$ is the same as the reference prior (see section 5.2).

It is proved in Chikobvu and Van der Merwe (2007) that the probability matching prior leads to a proper posterior distribution.

5.2 The Reference Prior for the Lower Process Performance Index:

The Jeffreys' and probability matching priors are but two methods to obtain useful non-informative priors. As mentioned, the Jeffreys' prior is not always suitable for multi-parameter problems. In recognition of this problem Berger and Bernado (1992), proposed the *reference prior* approach to the development of non-informative priors, the key feature of which was a possible dependence of the reference prior on specification of parameters of interest and nuisance parameters. As mentioned by Pearn and Wu (2005) the reference prior maximizes the difference in information (entropy) about the parameter provided by the prior and posterior. In other words the reference prior is derived in such a way that it provides as little information as possible about the parameter. In this section the reference prior of Berger and Bernado (1992) will be derived for the process performance index (P_{pl}^1). The solution depends on the ordering of the parameters and how the parameter vector is partitioned into sub-vectors. In spite of these difficulties, there is growing evidence, mainly through examples, that reference priors provide "sensible" answers from a Bayesian point of view and more limited evidence that frequentist properties of inference from reference posteriors are asymptotically "reasonable".

As is the case of the Jeffreys' prior, the reference prior method is derived from the Fisher information matrix. Note that the reference priors depend on the group ordering of the parameters. Berger and Bernado (1992) suggested that multiple groups, ordered in terms of inferential importance, are allowed, with the reference prior being determined through a succession of analyses for the implied conditional problems. They particularly recommended the reference prior based on having each parameter in its own group, i.e., having each

conditional reference prior be one dimensional. Notations such as $\{\mu, \sigma_2^2, \sigma_1^2\}$ will be used to specify the groups and importance of parameters; $\{\mu, \sigma_2^2, \sigma_1^2\}$ means that there are three groups, with μ being the most important and σ_1^2 the least important.

We will also examine whether the reference priors satisfy the probability-matching criterion. The following theorem can now be stated.

Theorem 5.2

For the lower process performance index, P_{pl}^1, the reference prior relative to the ordered parametrization $\{\mu, \sigma_2^2, \sigma_1^2\}$ is given by:

$$P_R(\mu, \sigma_2^2, \sigma_1^2) \propto \sigma_1^{-2} \left(\sigma_1^2 + J\sigma_2^2\right)^{-\frac{3}{2}} \left(1 + \frac{J(\mu - l_0)^2}{2\left(\sigma_1^2 + J\sigma_2^2\right)}\right)^{-\frac{1}{2}}.$$

Proof

The proof is given in the Appendix.

Corollary 5.1

The reference prior is the same as the probability matching prior. It is proved in Chikobvu and Van der Merwe (2007) that this is also the reference prior for the group ordering $\{\mu, \sigma_1^2, \sigma_2^2\}$.

6. THE WEIGHTED MONTE CARLO METHOD - SAMPLING - IMPORTANCE RE-SAMPLING

In this section a weighted Monte Carlo method is described which will be used for simulation from the posterior distribution in the case of the probability matching (reference) prior. This method is especially suitable for computing Bayesian confidence (credibility) intervals. It does not require knowing the closed form of the marginal posterior distribution of P_{pl}^1, only the kernel of the posterior distribution of $\{\mu, \sigma_1^2, \sigma_2^2\}$ is needed.

As mentioned by Smith and Gelfand (1992), Guttman and Menzefricke (2003), Skare, Bølviken and Holden (2003), Kim (2006) and Li (2007) the weighted Monte Carlo (sampling-importance re-sampling (SIR)) algorithm aims at drawing a random sample from a target distribution π by first drawing a sample from a proposal distribution q, and from this a smaller sample is drawn with sample probabilities proportional to the importance ratios π/q. For the algorithm to be efficient, it is important that q is a good approximation for π. This means that q should not have too light tails when compared to π. For further details, see Skare et al (2003).

In the case of credibility intervals it is not even necessary to draw the smaller sample. The weights (sample probabilities) are however important. For the Jeffreys' prior

$$P_J(\mu,\sigma_2^{\,2},\sigma_1^{\,2}) \propto \sigma_1^{\,-2}\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)^{-1},$$

the joint posterior of the parameters $\mu,\sigma_1^{\,2},\sigma_2^{\,2}$ is

$$P_J\left(\mu,\sigma_1^{\,2},\sigma_2^{\,2}\,|\,\underline{Y}\right) \propto (\sigma_1^{\,2})^{-\frac{1}{2}(\nu_1+2)}(\sigma_1^{\,2}+J\sigma_2^{\,2})^{-\frac{1}{2}(\nu_2+3)}\exp\left\{-\frac{1}{2}\left[\frac{IJ(\mu-\overline{\overline{Y}}_{..})^2}{(\sigma_1^{\,2}+J\sigma_2^{\,2})}+\frac{\nu_1 m_1}{\sigma_1^{\,2}}+\frac{\nu_2 m_2}{(\sigma_1^{\,2}+J\sigma_2^{\,2})}\right]\right\} \quad (6.1)$$

Equation (6.1) is our proposal distribution q. Although the Jeffreys' prior is improper, the proposal distribution q is proper. In the case of the reference (probability matching) prior

$$P_R(\mu,\sigma_2^{\,2},\sigma_1^{\,2}) \propto \sigma_1^{\,-2}\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)^{-\frac{3}{2}}\left(1+\frac{J(\mu-l_0)^2}{2\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)}\right)^{-\frac{1}{2}}, \quad (6.2)$$

the joint posterior distribution is

$$P_R\left(\mu,\sigma_1^{\,2},\sigma_2^{\,2}\,|\,\underline{Y}\right) \propto (\sigma_1^{\,2})^{-\frac{1}{2}(\nu_1+2)}(\sigma_1^{\,2}+J\sigma_2^{\,2})^{-\frac{1}{2}(\nu_2+4)}\times$$
$$\left(1+\frac{J(\mu-\overline{\overline{Y}}_{..})^2}{2(\sigma_1^{\,2}+J\sigma_2^{\,2})}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\frac{IJ(\mu-\overline{\overline{Y}}_{..})^2}{(\sigma_1^{\,2}+J\sigma_2^{\,2})}+\frac{\nu_1 m_1}{\sigma_1^{\,2}}+\frac{\nu_2 m_2}{(\sigma_1^{\,2}+J\sigma_2^{\,2})}\right]\right\}. \quad (6.3)$$

Equation (6.3) is the target distribution π. It is proved in Chikobvu and Van der Merwe (2007) that π is also a proper distribution. The sample probabilities are therefore proportional to

$$\pi/q = \frac{P_R\left(\mu,\sigma_1^{\,2},\sigma_2^{\,2}\right)}{P_J\left(\mu,\sigma_1^{\,2},\sigma_2^{\,2}\right)} \propto (\sigma_1^{\,2}+J\sigma_2^{\,2})^{-\frac{1}{2}}\left(1+\frac{J(\mu-l_0)^2}{2(\sigma_1^{\,2}+J\sigma_2^{\,2})}\right)^{-\frac{1}{2}}$$

and the normalized weights are

$$W_{(\ell)} = \frac{\left(\sigma_1^{\,2(\ell)}+J\sigma_2^{\,2(\ell)}\right)^{-\frac{1}{2}}\left(1+\frac{J(\mu^{(\ell)}-l_0)^2}{2\left(\sigma_1^{\,2(\ell)}+J\sigma_2^{\,2(\ell)}\right)}\right)^{-\frac{1}{2}}}{\sum_{\ell=1}^{\tilde{\ell}}\left(\sigma_1^{\,2(\ell)}+J\sigma_2^{\,2(\ell)}\right)^{-\frac{1}{2}}\left(1+\frac{J(\mu^{(\ell)}-l_0)^2}{2\left(\sigma_1^{\,2(\ell)}+J\sigma_2^{\,2(\ell)}\right)}\right)^{-\frac{1}{2}}} \qquad (\ell=1,2,\cdots,\tilde{\ell}). \tag{6.4}$$

Details of the Monte Carlo method are as follows:

Step 1:

Obtain a Monte Carlo sample

$\left\{(\mu^{(\ell)},\sigma_1^{\,2(\ell)},\sigma_2^{\,2(\ell)}),\ell=1,2,\cdots,\tilde{\ell}\right\}$ from the proposal distribution q and calculate the performance index

$$P_{pl}^{\,1(\ell)} = \frac{\mu^{(\ell)}-l_0}{3\left(\frac{\sigma_1^{\,2(\ell)}+J\sigma_2^{\,2(\ell)}}{J}\right)^{\frac{1}{2}}} \qquad (\ell=1,2,\cdots,\tilde{\ell}).$$

Step 2:

Sort $\left\{P_{pl}^{\,1(\ell)},(\ell=1,2,\cdots,\tilde{\ell})\right\}$ to obtain the ordered values $P_{pl}^{\,1(1)} \le P_{pl}^{\,1(2)} \le \cdots \le P_{pl}^{\,1(\tilde{\ell})}$.

Step 3:

Each simulated process performance index has an associated weight. Therefore compute the weighted function $W_{(\ell)}$ associated with the ℓ th ordered $P_{pl}^{\,1(\ell)}$ value.

Step 4:

Add the weights up from left to right (from the first on) till you get $\sum_{\ell=1}^{k_1} W_{(\ell)} = \alpha/2$.Write down the corresponding ordered $P_{pl}^{1(k_1)}$ value and denote it as $P_{pl}^{1}{}_{(\alpha/2)}$. Add the weights up from right to left (from the last back) till you get $\sum_{\ell=k_2}^{\tilde{\ell}} W_{(\ell)} = \alpha/2$. Write down the corresponding ordered value $P_{pl}^{1(k_2)}$ and denote it as $P_{pl}^{1}{}_{(1-\alpha/2)}$.

Step 5:

The $(1-\alpha)100\%$ Bayesian confidence interval is $\left(P_{pl}^{1}{}_{(\alpha/2)}, P_{pl}^{1}{}_{(1-\alpha/2)}\right)$.

For the data in Table 4.1 the 95% Bayesian confidence interval in the case of the probability matching (reference) prior for P_{pl}^{1} is $(1.030; 2.2698)$. The interval for the Jeffreys' prior is $(1.010; 2.2699)$ (See Figure 4.4). The two intervals are for all practical purposes the same. The reason is the relative large sample size.

7. CONCLUSION

Data arising from multiple sources of variability are very common in practice. Virtually all industrial processes exhibit between-batch as well as within-batch components of variation. In some cases the between-batch component is viewed as part of the common-cause-system for the process. It therefore seems worthwhile to develop a process capability index in more general settings.

In this paper we look at a version of the most popular process capability index C_{pk}, for the balanced random effects model using a Bayesian approach. The process performance index is denoted by P_{pl}^{1} and can be used for averages of observations from new or unknown

batches. A medical tablet manufacturing example illustrates the flexibility and unique features of the Bayesian simulation method for solving different complex problems such as Bayesian confidence intervals, ranking and selection, hypothesis testing and run length.

Determination of reasonable non-informative priors in multi-parameter problems is not easy. Common non-informative priors such as Jeffreys' prior can have features that have an unexpectedly dramatic effect on the posterior. Therefore reference and probability matching priors are derived for the lower process performance index, . Sampling-importance re-sampling is used to simulate from the posterior distribution in the case of the probability matching (reference) prior.

Appendix

Proof of Theorem 5.1

$\pi(\underline{\theta})$ is a probability-matching prior for $\underline{\theta} = [\mu, \sigma_2^{\,2}, \sigma_1^{\,2}]'$, the vector of unknown parameters, if the following differential equation is satisfied.

$$\sum_{\alpha=1}^{m} \frac{\partial}{\partial \theta_\alpha} \{\eta_\alpha(\underline{\theta})\pi(\underline{\theta})\} = 0$$

where

$$\eta(\underline{\theta}) = \frac{F^{-1}(\underline{\theta})\nabla_t(\underline{\theta})}{\sqrt{\nabla'_t(\underline{\theta})F^{-1}(\underline{\theta})\nabla_t(\underline{\theta})}} = [\eta_1(\underline{\theta}),\ldots,\eta_m(\underline{\theta})]',$$

$$\nabla_t(\underline{\theta}) = \left[\frac{\partial}{\partial \theta_1} t(\underline{\theta}),\ldots,\frac{\partial}{\partial \theta_m} t(\underline{\theta})\right],$$

$t(\underline{\theta})$ is a function of $\underline{\theta}$. In our case $t(\underline{\theta}) = P_{pl}^{\,1}$, the process performance index.

$F^{-1}(\underline{\theta})$ is the inverse of $F(\underline{\theta})$, the Fisher information matrix of $\underline{\theta}$.

The probability matching prior $\pi(\underline{\theta})$, is derived from the inverse of the Fisher information matrix. To obtain the Fisher information matrix, the negative of the expected values of the second derivatives (with respect to the parameters of the log–likelihood) must be calculated.

The Fisher information matrix is given by

$$F(\mu,\sigma_2^{\,2},\sigma_1^{\,2}) = F(\underline{\theta}) = \begin{pmatrix} \dfrac{IJ}{(\sigma_1^{\,2}+J\sigma_2^{\,2})} & 0 & 0 \\ 0 & \dfrac{IJ^2}{2(\sigma_1^{\,2}+J\sigma_2^{\,2})^2} & \dfrac{IJ}{2(\sigma_1^{\,2}+J\sigma_2^{\,2})^2} \\ 0 & \dfrac{IJ}{2(\sigma_1^{\,2}+J\sigma_2^{\,2})^2} & \dfrac{I(J-1)}{2(\sigma_1^{\,2})^2}+\dfrac{I}{2(\sigma_1^{\,2}+J\sigma_2^{\,2})^2} \end{pmatrix}$$

and its inverse by

$$F^{-1}(\underline{\theta}) = F^{-1}(\mu,\sigma_2^{\,2},\sigma_1^{\,2}) = \begin{pmatrix} \dfrac{(\sigma_1^{\,2}+J\sigma_2^{\,2})}{IJ} & 0 & 0 \\ 0 & \dfrac{2\left\{(J-1)(\sigma_1^{\,2}+J\sigma_2^{\,2})^2+(\sigma_1^{\,2})^2\right\}}{IJ^2(J-1)} & \dfrac{-2(\sigma_1^{\,2})^2}{IJ(J-1)} \\ 0 & \dfrac{-2(\sigma_1^{\,2})^2}{IJ(J-1)} & \dfrac{2(\sigma_1^{\,2})^2}{I(J-1)} \end{pmatrix}.$$

We are interested in the probability matching prior for ($P_{pl}^{\,1}$), the lower process performance index.

Let $\underline{\theta} = [\mu,\sigma_2^{\,2},\sigma_1^{\,2}]'$. The index is

$$P_{pl}^{\,1} = t(\underline{\theta}) = \frac{\mu - l_0}{3\left(\dfrac{\sigma_1^{\,2}+J\sigma_2^{\,2}}{J}\right)^{\frac{1}{2}}} = \frac{(\mu-l_0)J^{\frac{1}{2}}}{3\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)^{\frac{1}{2}}}.$$

Therefore

$$\frac{\partial t(\underline{\theta})}{\partial \mu} = \frac{J^{\frac{1}{2}}}{3\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)^{\frac{1}{2}}},\ \frac{\partial t(\underline{\theta})}{\partial \sigma_2^{\,2}} = \frac{-(\mu-l_0)J^{\frac{3}{2}}}{6\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)^{\frac{3}{2}}},\ \frac{\partial t(\underline{\theta})}{\partial \sigma_1^{\,2}} = \frac{-(\mu-l_0)J^{\frac{1}{2}}}{6\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)^{\frac{3}{2}}}.$$

As mentioned

$$\nabla'_t(\underline{\theta}) = \left[\frac{\partial t(\underline{\theta})}{\partial \mu} \quad \frac{\partial t(\underline{\theta})}{\partial {\sigma_2}^2} \quad \frac{\partial t(\underline{\theta})}{\partial {\sigma_1}^2} \right] = \frac{J^{\frac{1}{2}}}{3\left({\sigma_1}^2 + J{\sigma_2}^2\right)^{\frac{1}{2}}} \left[1 \quad \frac{-(\mu - l_0)J}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)} \quad \frac{-(\mu - l_0)}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)} \right].$$

Further

$$\nabla'_t(\underline{\theta}) F^{-1}(\underline{\theta}) = \frac{J^{\frac{1}{2}}}{3\left({\sigma_1}^2 + J{\sigma_2}^2\right)^{\frac{1}{2}}} \left[\frac{({\sigma_1}^2 + J{\sigma_2}^2)}{IJ} \quad \frac{-(\mu - l_0)\left({\sigma_1}^2 + J{\sigma_2}^2\right)}{IJ} \quad 0 \right]$$

and

$$\left\{ \nabla'_t(\underline{\theta}) F^{-1}(\underline{\theta}) \nabla_t(\underline{\theta}) \right\}^{\frac{1}{2}} = \frac{1}{3I^{\frac{1}{2}}} \left(1 + \quad \frac{J(\mu - l_0)^2}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)} \right)^{\frac{1}{2}}.$$

Define as before

$$\eta'(\underline{\theta}) = \frac{\nabla'_t(\underline{\theta}) F^{-1}(\underline{\theta})}{\sqrt{\nabla'_t(\underline{\theta}) F^{-1}(\underline{\theta}) \nabla_t(\underline{\theta})}} = [\eta_1(\underline{\theta}) \quad \eta_2(\underline{\theta}) \quad \eta_3(\underline{\theta})]$$

which means that

$$\eta_1(\underline{\theta}) = \frac{\left({\sigma_1}^2 + J{\sigma_2}^2\right)^{\frac{1}{2}}}{3IJ^{\frac{1}{2}}} 3I^{\frac{1}{2}} \left(1 + \quad \frac{J(\mu - l_0)^2}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)} \right)^{-\frac{1}{2}},$$

$$\eta_2(\underline{\theta}) = \frac{-(\mu - l_0)\left({\sigma_1}^2 + J{\sigma_2}^2\right)^{\frac{1}{2}}}{3IJ^{\frac{1}{2}}} 3I^{\frac{1}{2}} \left(1 + \quad \frac{J(\mu - l_0)^2}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)} \right)^{-\frac{1}{2}}$$

and

$\eta_3(\underline{\theta}) = 0$.

For a prior $\pi(\underline{\theta})$ to be a probability matching prior, the differential equation

$$\frac{\partial}{\partial \mu}\{\eta_1(\underline{\theta})\pi(\underline{\theta})\} + \frac{\partial}{\partial {\sigma_2}^2}\{\eta_2(\underline{\theta})\pi(\underline{\theta})\} + \frac{\partial}{\partial {\sigma_1}^2}\{\eta_3(\underline{\theta})\pi(\underline{\theta})\} = 0$$

must be satisfied.

The prior

$$\pi_a(\underline{\theta}) = \pi_a(\mu, {\sigma_2}^2, {\sigma_1}^2) \propto {\sigma_1}^{-2}\left({\sigma_1}^2 + J{\sigma_2}^2\right)^{-\frac{3}{2}}\left(1+ \quad \frac{J(\mu - l_0)^2}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)}\right)^{-\frac{1}{2}}$$

will be a probability matching prior since

$$\frac{\partial}{\partial\mu}\{\eta_1(\underline{\theta})\pi_a(\underline{\theta})\} \quad = \left\{\frac{-J(\mu - l_0)\sigma_1^{-2}}{I^{\frac{1}{2}}J^{\frac{1}{2}}}\left({\sigma_1}^2 + J{\sigma_2}^2\right)^{-2}\left(1+\frac{J(\mu - l_0)^2}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)}\right)^{-2}\right\},$$

$$\frac{\partial}{\partial{\sigma_2}^2}\{\eta_2(\underline{\theta})\pi_a(\underline{\theta})\} = \left\{\frac{J(\mu - l_0)\sigma_1^{-2}\left({\sigma_1}^2 + J{\sigma_2}^2\right)^{-2}}{I^{\frac{1}{2}}J^{\frac{1}{2}}}\left[1+\frac{J(\mu - l_0)^2}{2\left({\sigma_1}^2 + J{\sigma_2}^2\right)}\right]^{-2}\right\}$$

and

$$\frac{\partial}{\partial{\sigma_1}^2}\{\eta_3(\underline{\theta})\pi_a(\underline{\theta})\} = \frac{\partial}{\partial{\sigma_1}^2}\{0\} = 0\,.$$

Therefore

$$\frac{\partial}{\partial\mu}\{\eta_1(\underline{\theta})\pi_a(\underline{\theta})\} + \frac{\partial}{\partial{\sigma_2}^2}\{\eta_2(\underline{\theta})\pi_a(\underline{\theta})\} + \frac{\partial}{\partial{\sigma_1}^2}\{\eta_3(\underline{\theta})\pi_a(\underline{\theta})\} = 0\,.$$

Proof of Theorem 5.2

We are interested in the Fisher information matrix for $t(\underline{\theta}), \nu$ and ${\sigma_1}^2$. This will be obtained in two stages. Substituting $\nu = \frac{{\sigma_2}^2}{{\sigma_1}^2}$, the Fisher information matrix is given by

$$F(\mu, {\sigma_2}^2, {\sigma_1}^2) = \begin{pmatrix} \frac{IJ}{{\sigma_1}^2(1+J\nu)} & 0 & 0 \\ 0 & \frac{IJ^2}{2({\sigma_1}^2)^2(1+J\nu)^2} & \frac{IJ}{2({\sigma_1}^2)^2(1+J\nu)^2} \\ 0 & \frac{IJ}{2({\sigma_1}^2)^2(1+J\nu)^2} & \frac{I(J-1)}{2({\sigma_1}^2)^2} + \frac{I}{2({\sigma_1}^2)^2(1+J\nu)^2} \end{pmatrix}$$

To derive the Fisher information matrix for $\mu, \nu, {\sigma_1}^2$, let

$$A=\frac{\partial(\mu,\sigma_2^{\,2},\sigma_1^{\,2})}{\partial(\mu,\nu,\sigma_1^{\,2})}=\begin{pmatrix}\frac{\partial\mu}{\partial\mu} & \frac{\partial\mu}{\partial\nu} & \frac{\partial\mu}{\partial\sigma_1^{\,2}}\\ \frac{\partial\sigma_2^{\,2}}{\partial\mu} & \frac{\partial\sigma_2^{\,2}}{\partial\nu} & \frac{\partial\sigma_2^{\,2}}{\partial\sigma_1^{\,2}}\\ \frac{\partial\sigma_1^{\,2}}{\partial\mu} & \frac{\partial\sigma_1^{\,2}}{\partial\nu} & \frac{\partial\sigma_1^{\,2}}{\partial\sigma_1^{\,2}}\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & \sigma_1^{\,2} & \nu\\ 0 & 0 & 1\end{pmatrix}.$$

Now $F(\mu,\nu,\sigma_1^{\,2})=A'F(\mu,\sigma_2^{\,2},\sigma_1^{\,2})A=\begin{pmatrix}\frac{IJ}{\sigma_1^{\,2}(1+J\nu)} & 0 & 0\\ 0 & \frac{IJ^2}{2(1+J\nu)^2} & \frac{IJ}{2(\sigma_1^{\,2})(1+J\nu)}\\ 0 & \frac{IJ}{2(\sigma_1^{\,2})(1+J\nu)} & \frac{IJ}{2(\sigma_1^{\,2})^2}\end{pmatrix}.$

At this second stage the Fisher information matrix for μ,ν and $t(\underline{\theta})$ will be derived.

Therefore

$$t(\underline{\theta})=\frac{\mu-l_0}{3\left(\frac{\sigma_1^{\,2}+J\sigma_2^{\,2}}{J}\right)^{\frac{1}{2}}}=\frac{\mu-l_0}{3\left(\frac{\sigma_1^{\,2}(1+J\nu)}{J}\right)^{\frac{1}{2}}}$$

$$\frac{\partial\mu}{\partial t(\underline{\theta})}=3\frac{(\sigma_1^{\,2})^{\frac{1}{2}}}{J^{\frac{1}{2}}}(1+J\nu)^{\frac{1}{2}},$$

$$\frac{\partial\mu}{\partial\nu}=\frac{3}{2}J^{\frac{1}{2}}(\sigma_1^{\,2})^{\frac{1}{2}}(1+J\nu)^{-\frac{1}{2}}t(\underline{\theta}),$$

$$\frac{\partial\mu}{\partial\sigma_1^{\,2}}=\frac{3}{2}\frac{(\sigma_1^{\,2})^{-\frac{1}{2}}}{J^{\frac{1}{2}}}(1+J\nu)^{\frac{1}{2}}t(\underline{\theta}),$$

and

$$\tilde{A}=\begin{pmatrix} \frac{\partial\mu}{\partial t(\underline{\theta})} & \frac{\partial\mu}{\partial\nu} & \frac{\partial\mu}{\partial\sigma_1^{\,2}} \\ \frac{\partial\nu}{\partial t(\underline{\theta})} & \frac{\partial\nu}{\partial\nu} & \frac{\partial\nu}{\partial\sigma_1^{\,2}} \\ \frac{\partial\sigma_1^{\,2}}{\partial t(\underline{\theta})} & \frac{\partial\sigma_1^{\,2}}{\partial\nu} & \frac{\partial\sigma_1^{\,2}}{\partial\sigma_1^{\,2}} \end{pmatrix}=\begin{pmatrix} 3\frac{(\sigma_1^{\,2})^{\frac{1}{2}}}{J^{\frac{1}{2}}}(1+J\nu)^{\frac{1}{2}} & \frac{3}{2}J^{\frac{1}{2}}(\sigma_1^{\,2})^{\frac{1}{2}}(1+J\nu)^{-\frac{1}{2}}t(\underline{\theta}) & \frac{3}{2}\frac{(\sigma_1^{\,2})^{-\frac{1}{2}}}{J^{\frac{1}{2}}}(1+J\nu)^{\frac{1}{2}}t(\underline{\theta}) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The Fisher Information matrix for $t(\underline{\theta}),\nu$ and $\sigma_1^{\,2}$ is given by

$$F(t(\underline{\theta}),\nu,\sigma_1^{\,2})=\tilde{A}'F(\mu,\nu,\sigma_1^{\,2})\tilde{A}.$$

$$\tilde{A}'F(\mu,\nu,\sigma_1^{\,2})\tilde{A}=\begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix}=\begin{pmatrix} 9I & \frac{9}{2}\frac{IJ}{(1+J\nu)}t(\underline{\theta}) & \frac{9}{2}\frac{I}{(\sigma_1^{\,2})}t(\underline{\theta}) \\ \frac{9}{2}\frac{IJ}{(1+J\nu)}t(\underline{\theta}) & \frac{IJ^2\left(\frac{9}{2}t^2(\underline{\theta})+1\right)}{2(1+J\nu)^2} & \frac{IJ\left(\frac{9}{2}t^2(\underline{\theta})+1\right)}{2(\sigma_1^{\,2})(1+J\nu)} \\ \frac{9}{2}\frac{I}{(\sigma_1^{\,2})}t(\underline{\theta}) & \frac{IJ\left(\frac{9}{2}t^2(\underline{\theta})+1\right)}{2(\sigma_1^{\,2})(1+J\nu)} & \frac{I\left(\frac{9}{2}t^2(\underline{\theta})+1\right)}{2(\sigma_1^{\,2})^2}+\frac{I(J-1)}{2(\sigma_1^{\,2})^2} \end{pmatrix}.$$

As mentioned a reference prior depends on the group ordering of the parameters and it is determined through a succession of analysis for the implied conditional problems. Berger and Bernado (1992) particularly recommended the reference prior based on having each parameter in its own group, i.e. having each conditional reference prior being one dimensional.

Therefore consider the sub-matrix $\tilde{F}=\begin{pmatrix} f_{22} & f_{23} \\ f_{32} & f_{33} \end{pmatrix}$ and its inverse

$$\tilde{F}^{-1}=\begin{pmatrix} \frac{2(1+J\nu)^2}{IJ^2(J-1)}+\frac{2(1+J\nu)^2}{IJ^2\left(\frac{9}{2}t^2(\underline{\theta})+1\right)} & -\frac{2(\sigma_1^{\,2})(1+J\nu)}{IJ(J-1)} \\ -\frac{2(\sigma_1^{\,2})(1+J\nu)}{IJ(J-1)} & \frac{2(\sigma_1^{\,2})^2}{I(J-1)} \end{pmatrix}$$

The Reference prior in the $(t(\underline{\theta}),\nu,\sigma_1^2)$ parametrisation

Let $\widetilde{F}^{-1}=\begin{pmatrix} f_{22} & f_{23} \\ f_{32} & f_{33} \end{pmatrix}^{-1}$ from the section above.

Now

$$h_1 = f_{11.2} = f_{11} - \begin{bmatrix} f_{12} & f_{13} \end{bmatrix} \widetilde{F}^{-1} \begin{bmatrix} f_{21} \\ f_{31} \end{bmatrix} = 18I\left(9t^2(\underline{\theta})+2\right)^{-1},$$

$$h_2 = f_{22} - \frac{1}{f_{33}} f_{23} f_{32} = \frac{IJ^2\left(\frac{9}{2}t^2(\underline{\theta})+1\right)}{2(1+J\nu)^2}\left\{1-\frac{\frac{9}{2}t^2(\underline{\theta})+1}{\frac{9}{2}t^2(\underline{\theta})+J}\right\} \propto (1+J\nu)^{-2}$$

and

$$h_3 = f_{33} = \frac{I\left(\frac{9}{2}t^2(\underline{\theta})+1\right)}{2(\sigma_1^2)^2} + \frac{I(J-1)}{2(\sigma_1^2)^2} = \frac{\left(\frac{9}{2}It^2(\underline{\theta})+I+IJ-I\right)}{2(\sigma_1^2)^2} = \frac{\left(\frac{9}{2}It^2(\underline{\theta})+IJ\right)}{2(\sigma_1^2)^2} \propto (\sigma_1^2)^{-2}.$$

From the above it follows that

$$p(t(\underline{\theta})) \propto h_1^{\frac{1}{2}} = \left(9t^2(\underline{\theta})+2\right)^{-\frac{1}{2}},$$

$$p(\nu\,|\,t(\underline{\theta})) \propto h_2^{\frac{1}{2}} = \left(1+J\nu\right)^{-1},$$

$$p(\sigma_1^2\,|\,t(\underline{\theta}),\nu) \propto h_3^{\frac{1}{2}} = \sigma_1^{-2}.$$

Therefore the reference prior relative to the ordered parametrisation $(t(\underline{\theta}),\nu,\sigma_1^2)$ is given by

$$p(t(\underline{\theta}),\nu,\sigma_1^2) = p(t(\underline{\theta}))\,p(\nu\,|\,t(\underline{\theta})\,p(\sigma_1^2\,|\,t(\underline{\theta}),\nu),$$

$$p(t(\underline{\theta}),\nu,\sigma_1^2) = \left(9t^2(\underline{\theta})+2\right)^{-\frac{1}{2}}\left(1+J\nu\right)^{-1}\sigma_1^{-2}.$$

The reference prior in the $(\mu,\sigma_2^{\,2},\sigma_1^{\,2})$ parametrisation

As defined

$$t(\underline{\theta})=\frac{\mu-l_0}{3\left(\frac{\sigma_1^{\,2}+J\sigma_2^{\,2}}{J}\right)^{\frac{1}{2}}},$$

$$\frac{\partial t(\underline{\theta})}{\partial\mu}=\frac{1}{3}\left(\frac{\sigma_1^{\,2}+J\sigma_2^{\,2}}{J}\right)^{-\frac{1}{2}},\ \nu=\frac{\sigma_2^{\,2}}{\sigma_1^{\,2}}\ \text{and}\ \frac{\partial\nu}{\partial\sigma_2^{\,2}}=\frac{1}{\sigma_1^{\,2}}.$$

The reference prior for the group ordering $(\mu,\sigma_2^{\,2},\sigma_1^{\,2})$ is

$$p_R(\mu,\sigma_2^{\,2},\sigma_1^{\,2})\propto\left(\frac{J(\mu-l_0)^2}{2\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)}+1\right)^{-\frac{1}{2}}\left(1+\frac{J\sigma_2^{\,2}}{\sigma_1^{\,2}}\right)^{-1}\sigma_1^{-2}\frac{J^{\frac{1}{2}}}{3}\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)^{-\frac{1}{2}}\sigma_1^{-2}.$$

Therefore

$$p_R(\mu,\sigma_2^{\,2},\sigma_1^{\,2})\propto\left(\frac{J(\mu-l_0)^2}{2\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)}+1\right)^{-\frac{1}{2}}\left(\sigma_1^{\,2}+J\sigma_2^{\,2}\right)^{-\frac{3}{2}}\sigma_1^{-2}$$

which corresponds to the probability matching prior $\pi(\mu,\sigma_2^{\,2},\sigma_1^{\,2})$.

References

Bayarri, M.J. and Berger, J. (2004). The Interplay between Bayesian and Frequentist Analyis. Statist. Sci. **19**:58-80. MR2082147.

Bechhofer, R.E. (1954). A Single-sample Multiple Decision Procedure for Ranking Means of Normal populations with known Variances. Annals of Mathematical Statistics, **25**:16-39.

Berger, J.O. and Bernado, J.M. (1992). On the Development of Reference Priors in Bayesian Statistics IV, Eds. J.M. Bernardo, J.O. Berger, A.P. David and A.F.M. Smith, Oxford University Press, 35 - 70.

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

Chikobvu, D. and van der Merwe A.J. (2007). A process Capability Index for Averages of Observations from New Batches in the case of the Balanced Random Effects Model. Technical Report, no 378, Department of Mathematical Statistics, University of the Free State.

Datta, G.S. and Gosh, J.K. (1995). On Priors providing Frequentist Validity for Bayesian Inference. Biometrika, **82**:37-45.

Gunter, B.H. (1989a), The Use and Abuse of C_{pk} Quality Progress, January, 72-73.

Gunter, B.H. (1989b), The Use and Abuse of C_{pk} Part 2, Quality Progress, March, 108-109.

Gunter, B.H. (1989c), The Use and Abuse of C_{pk} Part 3, Quality Progress, May, 79-80.

Gunter, B.H. (1989d), The Use and Abuse of C_{pk} Part 4, Quality Progress, July, 86-87.

Gupta, S.S. (1956). On a Decision Rule for a Problem in Ranking Means. PhD thesis, Institute of Statistics, University of Norh Carolina, Chapel Hill.

Guttman, I. and Menzefrieke, U. (2003). Posterior Distributions for Functions of Variance Components. Sociedad de Estadistica e Investigacio 'n Operativa Test, **12** :115-123.

Herman, J.T. (1989). Capability Index-Enough for Process Industries? Transactions of the ASQC Quality Congress, Toronto 670-675.

Kim, H. (2006). On Bayesian Estimation of the Product of Poisson Rates with Application to Reliability. Communications in Statistics- Simulation and Computation, **35**:47-59.

Kotz, S. and Johnson, N.L. (1993), Process Capability Indices, Chapman and Hall, New York.

Li, K. (2007). Pool Size Selection for the Sampling/Importance Resampling Algorithm. Statistica Sinica **17** :895-907.

Lin, G.H., Pearn, W.L. and Yang, Y.S. (2005). A Bayesian Approach to Obtain a Lower Bound for the C_{pm} Capability Index. Quality and Reliability Engineering International, **21:**655-668.

Pearn, W.L. and Wu, C.W. (2005). Process Capability Assessment for Index C_{pk} based on Bayesian Approach. Metrika, **61**:221-234.

Severini, T.A., Mukerjee, R. and Ghosh, M. (2002). On An Exact Probability Matching Property of Right-invariant Priors. Biometrika **89***:*952-957. MR1946524.

Skare, O., Bølviken, E. and Holden, L. (2003). Improved Sampling-Importance Resampling and Reduced Bias Importance Sampling. J. Scand. Statist,. **30**:719-737.

Smith, A. and Gelfand, A. (1992). Bayesian Analysis Statistics without Tears: A Sampling-Resampling Perspective. Amer. Statist., **46(2)**:84-88.

Steiner S., Abraham B. and MacKay J. (1997). Understanding Process Capability Indices. Institute for Improvement of Quality and Productivity, Department of Statistics and Actuarial Science University of Waterloo, Waterloo, Ontario N2L 3G1.

Wolfinger, R.D. (1998). Tolerance Intervals for Variance Component Models Using Bayesian Simulation. Journal of Quality Technology, **30**:18-32.

Wu, C.W. & Pearn, W.L. (2005). Capability Testing Based on C_{pm} with Multiple Samples. Quality and Reliability Engineering International, **21**:29-42.

Wu, C.W. (2007). An Alternative Approach to test Process Capability for Unilateral Specification with Subsamples. International Journal of Production Research, **45**: 5397-5415.