A Bayesian Control Chart for the Coefficient of Variation in the Case of Pooled Samples

R. van Zyl[a,**], A.J. van der Merwe[b]

[a] *PAREXEL International, Bloemfontein, South Africa*
[b] *University of the Free State, Bloemfontein, South Africa*

Abstract

By using the data and results obtained by Kang, Lee, Seong, and Hawkins [7], a Bayesian procedure is applied to obtain control limits for the coefficient of variation. Reference and probability matching priors are derived for the coefficient of variation in the case of pooled samples. By simulating the posterior predictive density function of a future coefficient of variation it is shown that the control limits are effectively identical to those obtained by Kang et al. [7]. This article illustrates the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions, predictive intervals and run lengths in the case of the coefficient of variation.

Keywords: coefficient of variation, control charts, reference prior, probability-matching prior

1. Introduction

The monitoring of variability is a vital part of modern statistical process control (SPC). Shewart control charts are widely used SPC tools for detecting changes in the quality of a process. In most settings where the process is under control the process have readings that have a constant mean (μ) and constant variance (σ^2). In such settings the $\bar{X}$ chart is usually used to monitor the mean, and the R and S control charts the variance of the process.

In practice there are some situations though where the mean is not a constant and the usual SPC control reduces to the monitoring of the variability alone. As a further complication it sometimes happens that the variance of the process is a function of the mean. In these situations the usual R and S charts can also not be used.

The proposed remedy depends on the nature of the relationship between the mean and the variance of the process. One common relationship that we

*Corresponding author
**Principal corresponding author
Email address: ruaan.vanzyl@PAREXEL.com (R. van Zyl)

will look at is that the mean and standard deviation of the process is directly proportional so that the coefficient of variation

$$\gamma = \frac{\sigma}{\mu} \tag{1}$$

is a constant. According to Kang, Lee, Seong, and Hawkins [7] this is often the case in medical research. By using frequentist methods they developed a Shewart control chart, equivalent to the S chart, for monitoring the coefficient of variation using rational groups of observations. The chart is a time-ordered plot of the coefficient of variation for successive samples. It contains three lines:

- A center line;
- The upper control limit (UCL);
- The lower control limit (LCL).

By using the predictive distribution, a Bayesian procedure will be developed to obtain control limits for a future sample coefficient of variation. These limits will be compared to the classical limits obtained by Kang et al. [7].

Bayarri and García-Donato [2] give the following reasons for recommending a Bayesian analysis:

- Control charts are based on future observations and Bayesian methods are very natural for prediction.
- Uncertainty in the estimation of the unknown parameters is adequately handled.
- Implementation with complicated models and in a sequential scenario poses no methodological difficulty, the numerical difficulties are easily handled via Monte Carlo methods;
- Objective Bayesian analysis is possible without introduction of external information other than the model, but any kind of prior information can be incorporated into the analysis, if desired.

2. Frequentist Methods

Assume that X_i $(i = 1, 2, \ldots, n)$ are independently, identically normally distributed with mean μ and variance σ^2. $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ is the sample variance. The sample coefficient of variation is defined as

$$W = \frac{S}{\bar{X}}$$

Kang et al. [7] suggested a control chart for the sample coefficient of variation, similar to that of the $\bar{X}$, R and S charts. They proposed two methods in developing these charts:

1. The use of the non-central t distribution;
2. The use of the canonical form of the distribution of the coefficient of variation.

It can be noted that

$$T = \frac{\sqrt{n}\bar{X}}{S} = \sqrt{n}W^{-1}$$

follows a non-central t distribution with $(n-1)$ degrees of freedom and non-centrality parameter, $\frac{\sqrt{n}}{\gamma}$. The cumulative distribution function of the coefficient of variation can therefore be computed from the non-central t distribution.

In what follows, a more general distribution (than the canonical form of Kang et al. [7]) will be given for $W = \frac{S}{\bar{X}}$. Using a Bayesian procedure this distribution will be used for prediction purposes:

$$f(w|\gamma) = \begin{cases} \frac{A(w)}{(n+fw^2)^{\frac{f+1}{2}}} I_f\left(\frac{n}{\gamma(n+fw^2)^{0.5}}\right) & , w \geq 0 \\ \frac{(-1)^{f-1}A(w)}{(n+fw^2)^{\frac{f+1}{2}}} I_f\left(\frac{n}{\gamma(n+fw^2)^{0.5}}\right) & , w < 0 \end{cases} \tag{2}$$

where $\gamma = \frac{\sigma}{\mu}$, $f = n-1$,

$$A(w|\gamma) = \frac{f^{\frac{f}{2}}\sqrt{n}w^{f-1}}{2^{\frac{1}{2}(f-2)}\Gamma\left(\frac{f}{2}\right)\sqrt{2\pi}} \exp\left\{-\frac{\frac{1}{2}\left(nfw^2\right)}{\gamma^2\left(n+fw^2\right)}\right\}$$

and

$$I_f\left(\frac{n}{\gamma\left(n+fw^2\right)^{0.5}}\right) = \int_0^\infty q^f \exp\left\{-\frac{1}{2}\left[q - \frac{n}{\gamma\left(n+fw^2\right)^{\frac{1}{2}}}\right]\right\} dq$$

is the Airy function (Iglewicz [6]).

2.1. The Data

The example used by Kang et al. [7] was that of patients undergoing organ transplantation, for which Cyclosporine is administered. For patients undergoing immunosuppressive treatment, it is vital to control the amount of drug circulating in the body. For this reason frequent blood assays were taken to find the best drug stabilizing level for each patient. The dataset consist of $m = 105$ patients and the number of assays obtained for each patient is $n = 5$. By doing a regression test they confirmed that there is no evidence that the coefficient of variation depends on the mean which means that the assumption of a constant coefficient of variation is appropriate. They used the weighted root mean square estimator $\hat{\gamma} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} w_i^2} = \sqrt{\frac{0.593515}{105}} = 0.075$ to pool the samples for estimating γ. By substituting $\hat{\gamma}$ in equation (2) and by calculating the lower and upper $\frac{1}{740}$ percentage points, they obtained a LCL = 0.01218 and UCL = 0.15957. The chart was then applied to a fresh data set of 35 samples from a different laboratory.

3. Bayesian Procedure

Since five observations per patient is quite small, groups of five patients will be pooled together to implement the Bayesian procedure. Based on similar means, as presented in AppendixA, the results of the first five patients will therefore be pooled together, similar the results of the second five patients and so forth. $k = 21$ new groups are therefore formed.

By assigning a prior distribution tot he unknown parameters the uncertainty in the estimation of the unknown parameters can adequately be handled. The information contained in the prior is combined with the likelihood to obtain the posterior distribution of γ. By using the posterior distribution the predictive distribution of a future coefficient of variation can be obtained. The predictive distribution on the other hand can be used to determine the distribution of the "run length". Determination of reasonable non-informative priors is however not an easy task. Therefore, in the next section, reference and probability matching priors will be derived for the coefficient of variation in the case of pooled samples.

4. Reference and Probability-Matching Priors for the Coefficient of Variation in the Case of Pooled Samples

As mentioned the Bayesian paradigm emerges as attractive in many types of statistical problems, also in the case of the coefficient of variation.

Prior distributions are needed to complete the Bayesian specification of the model. Determination of reasonable non-informative priors in multi-parameter problems is not easy; common non-informative priors, such as the Jeffreys' prior can have features that have an unexpectedly dramatic effect on the posterior.

Reference and probability-matching priors often lead to procedures with good frequency properties while returning to the Bayesian flavour. The fact that the resulting Bayesian posterior intervals of the level $1 - \alpha$ are also good frequentist intervals at the same level is a very desirable situation.

See also Bayarri and Berger [1] and Severine, Mukerjee, and Ghosh [13] for a general discussion.

4.1. The Reference Prior

In this section the reference prior of Berger and Bernardo [3] will be derived for the coefficient of variation in the case of pooled samples. In general, the derivation depends on the ordering of the parameters and how the parameter vector is divided into sub-vectors. As mentioned by Pearn and Wu [12] 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 possible about the parameter of interest. The reference prior algorithm is relatively complicated and, as mentioned, 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 that frequentist

properties of inference from reference posteriors are asymptotically "good". As in the case of the Jeffreys' prior, the reference prior is obtained from the Fisher information matrix. In the case of a scalar parameter, the reference prior is the Jeffreys' prior.

Berger, Liseo, and Wolpert [4] derived the reference prior for the coefficient of variation in the case of a single sample. From the medical example given in Kang et al. [7] it is clear that the standard deviation of measurements is approximately proportional to the mean; that is, the coefficient of variation is constant across the range of means, which is an indication that the a reference prior for a pooled coefficient of variation should be derived.

Theorem 1. *Let* $x_{pl\tilde{m}} \sim N\left(\mu_{\tilde{m}}, \sigma^2_{\tilde{m}}\right)$ *where* $p = 1, 2, ..., \tilde{p}$, $l = 1, 2, ..., \tilde{l}$, $\tilde{m} = 1, 2, ..., k$ *and the coefficient of variation is* $\gamma = \frac{\sigma_1}{\mu_1} = \frac{\sigma_2}{\mu_2} = ... = \frac{\sigma_k}{\mu_k}$.

The reference prior for the ordering $\left\{\gamma; \left(\sigma^2_1, \sigma^2_2, ..., \sigma^2_k\right)\right\}$ *is given by*

$$p_R\left(\gamma, \sigma^2_1, \sigma^2_2, ..., \sigma^2_k\right) \propto \frac{1}{|\gamma| \sqrt{\gamma^2 + \frac{1}{2}}} \prod_{\tilde{m}=1}^{k} \sigma^{-2}_{\tilde{m}}$$

Proof. The proof is given in AppendixB. □

Note: The ordering $\left\{\gamma; \left(\sigma^2_1, \sigma^2_2, \ldots, \sigma^2_k\right)\right\}$ means that the coefficient of variation is the most important parameter while the k variance components are of equal importance, but not as important as γ. Also, if $k = 1$, equation (B.1) simplifies to the reference prior obtained by Berger et al. [4].

4.2. Probability-Matching Priors

The reference prior algorithm is but one way to obtain a useful non-informative prior. Another type of non-informative prior is the probability-matching prior. This prior has good frequentist properties. Two reasons for using probability-matching priors are that they provide a method for constructing accurate frequentist intervals, and that they could be potentially useful for comparative purposes in a Bayesian analysis.

There are two methods for generating probability-matching priors due to Tibshirani [14] and Datta and Ghosh [5].

Tibshirani [14] generated probability-matching priors by transforming the model parameters so that the parameter of interest is orthogonal to the other parameters. The prior distribution is then taken to be proportional to the square root of the upper left element of the information matrix in the new parametrization.

Datta and Ghosh [5] provided a different solution to the problem of finding probability-matching priors. They 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 $O\left(n^{-1}\right)$ where n is the sample size.

According to Datta and Ghosh [5] $p(\underline{\theta})$ is a probability-matching prior for $\underline{\theta} = \left[\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2\right]'$ the vector of unknown parameters, if the following differential equation is satisfied:

$$\sum_{\alpha=1}^{k+1} \frac{\partial}{\partial \theta_\alpha} \left\{ \Upsilon_\alpha(\underline{\theta}) p(\underline{\theta}) \right\} = 0$$

where

$$\Upsilon(\underline{\theta}) = \frac{F^{-1}(\underline{\theta}) \nabla_t(\underline{\theta})}{\sqrt{\nabla_t'(\underline{\theta}) F^{-1}(\underline{\theta}) \nabla_t(\underline{\theta})}} = \left[\begin{array}{cccc} \Upsilon_1(\underline{\theta}) & \Upsilon_2(\underline{\theta}) & \cdots & \Upsilon_{k+1}(\underline{\theta}) \end{array} \right]'$$

and

$$\nabla_t(\underline{\theta}) = \left[\begin{array}{cccc} \frac{\partial}{\partial \theta_1} t(\underline{\theta}) & \frac{\partial}{\partial \theta_2} t(\underline{\theta}) & \cdots & \frac{\partial}{\partial \theta_{k+1}} t(\underline{\theta}) \end{array} \right]'.$$

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

Theorem 2. *The probability-matching prior for the coefficient of variation γ and the variance components is given by*

$$p_M\left(\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2\right) \propto \frac{1}{|\gamma| (1+2\gamma^2)^{\frac{1}{2}}} \prod_{\tilde{m}=1}^{k} \sigma_{\tilde{m}}^{-2} = \frac{1}{|\gamma| \sqrt{\gamma^2 + \frac{1}{2}}} \prod_{\tilde{m}}^{k} \sigma_{\tilde{m}}^{-2}$$

Proof. The proof is provided in AppendixC. □

From Theorems 1 and 2 it is clear that the reference and probability-matching priors are equal.

4.3. The Joint Posterior Distribution

By combining the prior with the likelihood the joint posterior distribution of $\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ can be obtained.

$$p\left(\gamma, \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2 | data\right) \propto \prod_{\tilde{m}=1}^{k} \left(\sigma_{\tilde{m}}^2\right)^{-\frac{n^*}{2}} \exp\left\{ -\frac{1}{2\sigma_{\tilde{m}}^2} \left[n^* \left(\bar{x}_{\tilde{m}} - \frac{\sigma_{\tilde{m}}}{\gamma} \right)^2 + v_{\tilde{m}} s_{\tilde{m}}^2 \right] \right\} \frac{1}{|\gamma| (1+2\gamma^2)^{\frac{1}{2}}} \prod_{\tilde{m}=1}^{k} \sigma_{\tilde{m}}^{-2} \tag{3}$$

The conditional posterior distributions are given by

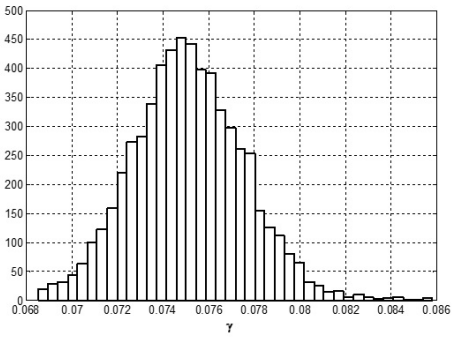
$$p\left(\gamma | \sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2, data\right) \propto \frac{1}{|\gamma| (1+2\gamma^2)^{\frac{1}{2}}} \exp\left\{ -\sum_{\tilde{m}=1}^{k} \frac{1}{2\sigma_{\tilde{m}}^2} \left[n^* \left(\bar{x}_{\tilde{m}} - \frac{\sigma_{\tilde{m}}}{\gamma} \right)^2 \right] \right\} \tag{4}$$

$$p\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2 | \gamma, data\right) \propto \prod_{\tilde{m}=1}^{k} \left(\sigma_{\tilde{m}}^2\right)^{-\frac{1}{2}(n^*+2)} \exp\left\{-\frac{1}{2\sigma_{\tilde{m}}^2}\left[n^*\left(\bar{x}_{\tilde{m}} - \frac{\sigma_{\tilde{m}}}{\gamma}\right)^2 + v_{\tilde{m}} s_{\tilde{m}}^2\right]\right\} \tag{5}$$

For the medical example, $\tilde{p} = 5$, $\tilde{l} = 5$, $n^* = \tilde{p}\tilde{l} = 25$ and $k = 21$. As mentioned the reason for the pooling is that five observations per patient is quite small.

By using the conditional posterior distributions (equations [4] and [5]) and Gibbs sampling the unconditional posterior distribution of the coefficient of variation, $p\left(\gamma | data\right)$ can be obtained as illustrated in Figure 1.

Figure 1: Histogram of the Posterior-Distribution of $\gamma = \frac{\sigma}{\mu}$



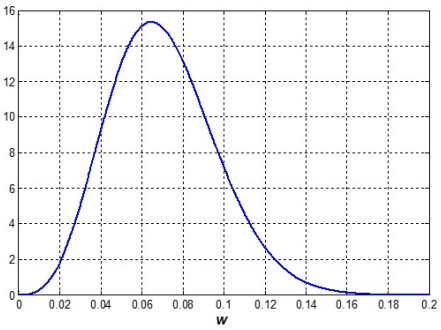
$mean\left(\gamma\right) = 0.0751$, $median\left(\gamma\right) = 0.0750$, $mode\left(\gamma\right) = 0.0748$,
$var\left(\gamma\right) = 5.951e^{-6}$
95% equal-tail interval = $(0.0705; 0.0800)$, length 0.00942
95% HDP interval =$(0.07048; 0.07989)$, length 0.00941

From a frequentist point of view Kang et al. [7] mentioned that the best way to pool the sample coefficients of variation is to calculate the weighted root mean square $\hat{\gamma} = \sqrt{\frac{1}{m_t}\sum_i w_i^2} = \sqrt{\frac{1}{105}(0.593515)} = 0.075$.

It is interesting to note that the weighted root mean square value is equal to the mean (median) of the posterior distribution of γ.

By substituting each of the simulated γ values of the posterior distribution into the conditional predictive density $f\left(w|\gamma\right)$ and using the Rao-Blackwell procedure the unconditional posterior predictive density $f\left(w|data\right)$ of a future sample coefficient of variation can be obtained. This is illustrated in Figure 2 for $n = 5$.

Figure 2: Predictive Density $f(w|data)$ for $n = 5$



$mean(w) = 0.0705$, $median(w) = 0.0686$, $mode(w) = 0.0647$,
$var(w) = 6.6743e^{-6}$
95% equal-tail interval = $(0.0259; 0.1260)$, length = 0.1001
95% HDP interval = $(0.0220; 0.1207)$, length = 0.0987
99.73% equal-tail interval = $(0.0121; 0.1602)$, length = 0.1481
99.73% HDP interval = $(0.0086; 0.1546)$, length = 0.1460

Kang et al. [7] calculated lower (LCL=0.01218) and upper (UCL=0.15957) control limits which are for all practical purposes the same as the 99.73% equal-tail prediction interval.

Kang et al. [7] then applied their control chart to a new dataset of 35 patients from a different laboratory. Eight of the patients' coefficient of variation (based on five observations) lie outside the control limits. Since the 99.73% equal-tail prediction interval is effectively identical to the control limits of Kang et al. [7] our conclusions are the same.

As mentioned the rejection region of size α ($\alpha = 0.0027$) for the predictive distribution is

$$\alpha = \int_{R(\alpha)} p(w|data)\, dw.$$

In the case of the equal-tail interval, $R(\alpha)$ represents those values of w that are smaller than 0.0121 or larger than 0.1602.

Assuming that the process remains stable, the predictive distribution can be used to derive the distribution of the "run length" or "average run length". The "run length" is defined as the number of future coefficients of variation, r until the control chart signals for the first time. (Note that r does not include the coefficient of variation when the control chart signals.) Given γ and a stable Phase I process, the distribution of the run length r is geometric with parameter

$$\Psi(\gamma) = \int_{R(\alpha)} f(w|\gamma)\, dw$$

where $f(w|\gamma)$ is the distribution of the sample coefficient of variation given γ as defined in equation (2).

The value of γ is of course unknown and the uncertainty is described by the posterior distribution.

The predictive distribution of the "run length" or the "average run length" can therefore be easily simulated. The mean and second moment about zero of r given γ are given by

$$E\left(r|\gamma\right) = \frac{1}{\Psi\left(\gamma\right)}$$

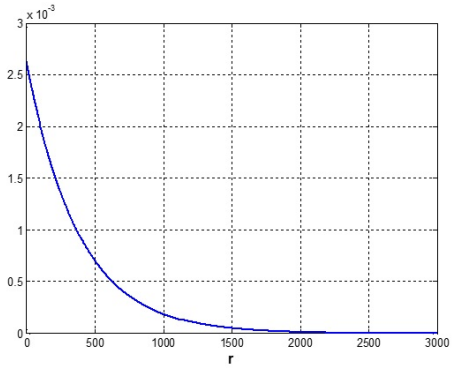
and

$$E\left(r^2|\gamma\right) = \frac{2 - \Psi\left(\gamma\right)}{\Psi\left(\gamma\right)}$$

The unconditional moments $E\left(r|data\right)$, $E\left(r^2|data\right)$ and $Var\left(r|data\right)$ can therefore easily obtained by simulation or numerical integration. For further details see Menzefricke [8, 9, 10, 11].

In Figure 3 the predictive distribution of the "run length" is displayed and in Figure 4, the distribution of the "average run length" is given.

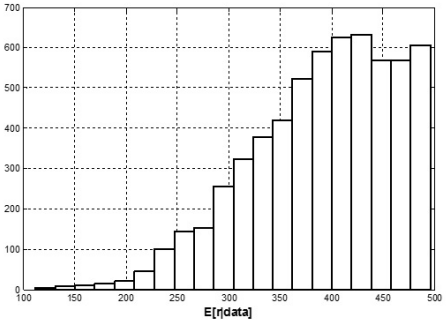
Figure 3: Predictive Distribution of the Run Length $p\left(r|data\right)$ for $n = 5$



$E\left(r|data\right) = 392.7419$, $Median\left(r|data\right) = 265$, $Var\left(r|data\right) = 1.6379e^5$
95% Equal-tail Interval = $(8; 1487)$, Length = 1479
95% HDP Interval = $(0; 1196)$, Length = 1196

As mentioned for given γ, the run length r is geometric with parameter $\Psi\left(\gamma\right)$. The unconditional run length displayed in Figure 3 is therefore obtained using the Rao-Blackwell method, i.e., it is the average of the conditional run lengths.

Figure 4: Distribution of thee Expected Run Length

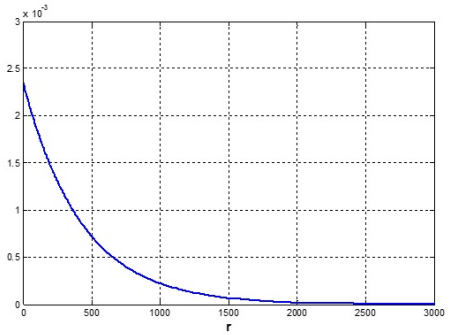


Mean = 392.7488, Median = 402.4477, Variance = $4.9863e^3$
95% Equal-tail Interval = (237.42; 495.26), Length = 257.85
95% HDP Interval = (262.95; 497.74), Length = 234.79

From Figure 3 it can be seen that the expected run length, $E(r|data) = 392.74$, is somewhat larger than the ARL of 370 given by Kang et al. [7]. The median run length $Median(r|data) = 265$ is smaller than he mean run length. This is clear from the skewness of the distribution.

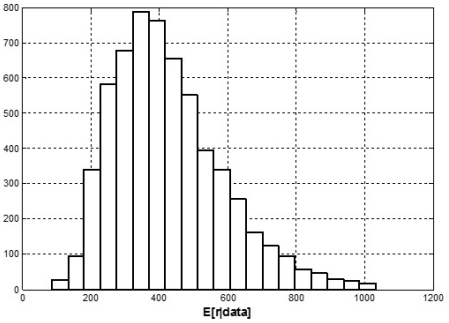
In the case of the HDP limits, $\tilde{R}(\alpha)$ represents those values of w that are smaller than 0.0086 and larger than 0.1546. The predictive distribution of the run length is illustrated in Figure 5 while the distribution of the average run length is given in Figure 6.

Figure 5: Predictive Distribution of the Run Length in the Case of HDP Limits



$E(r|data) = 425.8417$, $Median(r|data) = 267.7$, $Var(r|data) = 2.3200e^5$
95% Equal-tail Interval = (8.35; 1750.20), Length = 1741.85
95% HDP Interval = (0; 1363.08), Length = 1363.08

Figure 6: Distribution of the Expected Run Length for HDP Limits



Mean = 426.2723, Median = 400.5495, Variance = $2.7102e^4$
95% Equal-tail Interval = (180.74; 820.17), Length = 639.43
95% HDP Interval = (169.39; 789.76), Length = 620.37

A comparison of Figure (3) and Figure (5) show that the median run length for equal tail and HDP limits are more or less the same.

5. Conclusion

This paper develops a Bayesian control chart for monitoring the coefficient of variation in the case of pooled samples. In the Bayesian approach prior knowledge about the unknown parameters is formally incorporated into the process of inference by assigning a prior distribution to the parameters. The information contained in the prior is combined with the likelihood function to obtain the posterior distribution. By using the posterior distribution the predictive distribution of a future coefficient of variation can be obtained.

Determination of reasonable non-informative priors in multi-parameter problems is not an easy task. The Jeffreys' prior for example can have a bad effect on the posterior distribution. Reference and probability matching priors are therefore derived for the coefficient of variation int he case of pooled samples. The theory and results are applied to a real problem of patients undergoing organ transplantation for which Cyclosporine is administered. This problem is discussed in detail by Kang et al. [7]. The 99.73% equal tail prediction interval of a future coefficient of variation is effectively identical to the lower and upper control chart limits calculated by Kang et al. [7].

The example illustrates the flexibility and unique features of the Bayesian simulation method for obtaining posterior distributions, prediction intervals and run lengths.

AppendixA. Data for Medical Example

m	$\overline{X}$	W	m	$\overline{X}$	W	m	$\overline{X}$	W
1	31.7	12.4	36	120.3	5.8	71	361.4	8.3
2	37.7	15.3	37	143.7	5.6	72	361.5	13.4
3	40.6	9.1	38	148.6	5.5	73	361.8	6.1
4	50.5	4.6	39	149.1	3.1	74	374.6	5.8
5	52	10.5	40	149.9	2	75	376.3	2.8
6	57.6	6.2	41	151	4.4	76	382.3	5.8
7	58.3	6.6	42	153.6	6.6	77	401.7	7.3
8	58.9	8.4	43	172.2	7.2	78	415.2	15.1
9	61.2	8.1	44	179.8	7.9	79	428.8	4.5
10	64.3	7	45	185.3	7.6	80	442.1	9.9
11	64.5	8.8	46	192.1	5.3	81	450.1	7.4
12	65.6	4.1	47	193.8	5.9	82	496.5	4.8
13	68	3.7	48	195.1	11	83	499.7	10
14	71.8	6.2	49	195.2	5.1	84	504.6	8.4
15	72.1	8.4	50	195.4	9.4	85	523.1	5
16	78.4	6.8	51	196.4	5.6	86	531.7	8.5
17	78.4	4.6	52	199.6	6.8	87	556.4	11.8
18	79.5	5.7	53	204.4	3.7	88	571.4	5.9
19	83.2	10.5	54	207.8	12.4	89	584.1	8.3
20	85.1	4.8	55	219	7.6	90	597.6	4.2
21	85.6	5.4	56	222.9	4.8	91	606.2	8.2
22	86	10.1	57	225.1	5.7	92	609	9.7
23	87.3	7.9	58	227.6	6.5	93	635.4	5.6
24	89.1	10.3	59	240.5	3.8	94	672.2	7.2
25	95.4	6.2	60	241.1	8.4	95	695.9	2.7
26	101.9	4.8	61	252.2	8.3	96	696.4	10.6
27	105.4	5.6	62	262.2	5.8	97	721.3	9.8
28	107.2	2.2	63	277.9	8.7	98	752	4.2
29	108.2	3.3	64	278.3	6.2	99	769.5	9.7
30	112	8.7	65	303.4	8.8	100	772.7	9.6
31	112.3	5.7	66	309.7	3.9	101	791.6	2
32	113.5	9.4	67	323.9	4.1	102	799.9	11.4
33	114.3	3.5	68	328.7	4.1	103	948.6	5.2
34	116.8	6	69	341.2	6.5	104	971.8	11.1
35	117.8	5.7	70	347.3	4.9	105	991.2	8.8

AppendixB. Proof of Theorem 1

Proof. The likelihood function is given by

$$L\left(\gamma,\sigma_1^2,\sigma_2^2,...,\sigma_k^2|data\right)\propto\prod_{\tilde{m}=1}^{k}\left(\sigma_{\tilde{m}}^2\right)^{-\frac{n^*}{2}}\exp\left\{-\frac{1}{2\sigma_{\tilde{m}}^2}\left[n^*\left(\bar{x}_{\tilde{m}}-\frac{\sigma_{\tilde{m}}}{\gamma}\right)^2+v_{\tilde{m}}s_{\tilde{m}}^2\right]\right\}$$

where

$$\bar{x}_{\tilde{m}}=\frac{1}{n^*}\sum_{l=1}^{\tilde{l}}\sum_{p=1}^{\tilde{p}}x_{pl\tilde{m}}$$

$$v_{\tilde{m}}s_{\tilde{m}}^2=\sum_{l=1}^{\tilde{l}}\sum_{p=1}^{\tilde{p}}x_{pl\tilde{m}}^2-n^*\bar{x}_{\tilde{m}}^2$$

and

$$n^*=\tilde{p}\tilde{l}.$$

By differentiating the log likelihood function, l^*, twice with respect to the unknown parameters and taking expected values the Fisher information matrix can be obtained.

$$l^*=\log L\left(\gamma,\sigma_1^2,\sigma_2^2,...,\sigma_k^2|data\right)=-\frac{n^*}{2}\sum_{\tilde{m}=1}^{k}\log\sigma_{\tilde{m}}^2-\frac{1}{2}\sum_{\tilde{m}=1}^{k}\frac{1}{\sigma_{\tilde{m}}^2}\left[n^*\left(\bar{x}_{\tilde{m}}-\frac{\sigma_{\tilde{m}}}{\gamma}\right)^2+v_{\tilde{m}}s_{\tilde{m}}^2\right]$$

and

$$\frac{\partial^2 l^*}{\left(\partial\sigma_{\tilde{m}}^2\right)^2}=\frac{n^*}{2}\left(\frac{1}{\sigma_{\tilde{m}}^2}\right)^2-\frac{2n^*\bar{x}_{\tilde{m}}^2}{2\left(\sigma_{\tilde{m}}^2\right)^3}+\frac{3n^*\bar{x}_m}{4\gamma\sigma_{\tilde{m}}^5}-\frac{v_{\tilde{m}}s_{\tilde{m}}^2}{\left(\sigma_{\tilde{m}}^2\right)^3}.$$

Therefore

$$-E\left[\frac{\partial^2 l^*}{\left(\partial\sigma_{\tilde{m}}^2\right)^2}\right]=\frac{n^*}{2}\left(\frac{1}{\sigma_{\tilde{m}}^2}\right)^2\left\{1+\frac{1}{2\gamma^2}\right\}\text{ where }\tilde{m}=1,2,...,k.$$

Also

$$-E\left[\frac{\partial^2 l^*}{\partial\sigma_{\tilde{m}}^2\partial\sigma_{m^*}^2}\right]=0\text{ where }\tilde{m}=1,2,...,k,\ m^*=1,2,...,k\text{ and }\tilde{m}\neq m^*.$$

Further

$$\frac{\partial^2 l^*}{\left(\partial\gamma\right)^2}=\sum_{\tilde{m}=1}^{k}\left(\frac{2n^*\bar{x}_m}{\sigma_{\tilde{m}}\gamma^3}-\frac{3n^*}{\gamma^4}\right)$$

and

$$-E\left[\frac{\partial^2 l^*}{(\partial\gamma)^2}\right] = \frac{kn^*}{\gamma^4}$$

If we differentiate l^* with respect to $\sigma^2_{\tilde{m}}$ and γ we get

$$\frac{\partial^2 l^*}{\partial\sigma^2_{\tilde{m}}\partial\gamma} = \frac{n^*\bar{x}_{\tilde{m}}}{2\gamma^2\sigma^3_{\tilde{m}}}$$

and

$$-E\left[\frac{\partial^2 l^*}{\partial\sigma^2_{\tilde{m}}\partial\gamma}\right] = \frac{-n^*}{2\gamma^3\sigma^2_{\tilde{m}}}.$$

The Fisher information matrix then follows as

$$F\left(\gamma,\sigma_1^2,\sigma_2^2,...,\sigma_k^2\right) \quad = \quad \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}$$

where

$$F_{11} = \frac{kn^*}{\gamma^4}\ ,\ F_{12} = F_{21}' = \begin{bmatrix} \frac{-n^*}{2\gamma^3\sigma_1^2} & \frac{-n^*}{2\gamma^3\sigma_2^2} & \cdots & \frac{-n^*}{2\gamma^3\sigma_k^2} \end{bmatrix}$$

and

$$F_{22} = \begin{bmatrix} \frac{n^*}{2}\left(\frac{1}{\sigma_1^2}\right)^2\left\{1+\frac{1}{2\gamma^2}\right\} & 0 & \cdots & 0 \\ 0 & \frac{n^*}{2}\left(\frac{1}{\sigma_2^2}\right)^2\left\{1+\frac{1}{2\gamma^2}\right\} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{n^*}{2}\left(\frac{1}{\sigma_k^2}\right)^2\left\{1+\frac{1}{2\gamma^2}\right\} \end{bmatrix}.$$

To calculate the reference prior for the ordering $\left\{\gamma;\left(\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2\right)\right\}$ we must first calculate $F_{11\cdot 2}$ and then $|F_{22}|$. Now

$$F_{11\cdot 2} = F_{11} - F_{12}F_{22}^{-1}F_{21} = \frac{kn^*}{\gamma^4} - \frac{kn^{*2}}{2^2\left(\gamma^3\right)^2}\frac{4\gamma^2}{n^*\left(2\gamma^2+1\right)} = \frac{2kn^*}{\gamma^2\left(2\gamma^2+1\right)} = h_1$$

and

$$p\left(\gamma\right) \propto h_1^{\frac{1}{2}} \propto \frac{1}{|\gamma|\sqrt{\gamma^2+\frac{1}{2}}}.$$

Also

$$|F_{22}| = \left(\frac{n^*}{2}\left\{1+\frac{1}{2\gamma^2}\right\}\right)^k \prod_{\tilde{m}=1}^{k}\left(\frac{1}{\sigma_{\tilde{m}}^2}\right)^2 = h_2$$

which means that

$$p\left(\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2|\gamma\right)\propto h_2^{\frac{1}{2}}\propto\prod_{\tilde{m}}^{k}\left(\frac{1}{\sigma_{\tilde{m}}^2}\right).$$

Therefore the reference prior for the ordering $\left\{\gamma;\left(\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2\right)\right\}$ is

$$p_R\left(\gamma,\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2\right)=p\left(\gamma\right)p\left(\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2|\gamma\right)\propto\frac{1}{|\gamma|\sqrt{\gamma^2+\frac{1}{2}}}\prod_{\tilde{m}=1}^{k}\sigma_{\tilde{m}}^{-2}. \tag{B.1}$$

□

AppendixC. Proof of Theorem 2

Proof. Using the previously derived Fisher information matrix we can calculate

$$F^{-1}\left(\underline{\theta}\right)=F^{-1}\left(\gamma,\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2\right)=\begin{bmatrix}F^{11}&F^{12}&F^{13}&\cdots&F^{1,k+1}\\F^{21}&F^{22}&F^{23}&\cdots&F^{2,k+1}\\\vdots&\vdots&\vdots&\ddots&\vdots\\F^{k+1,1}&F^{k+1,2}&F^{k+1,3}&\cdots&F^{k+1,k+1}\end{bmatrix}.$$

Let

$$t\left(\underline{\theta}\right)=t\left(\gamma,\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2\right)=\gamma.$$

Since

$$\nabla'\left(\underline{\theta}\right)=\begin{bmatrix}\frac{\partial}{\partial\gamma}t\left(\underline{\theta}\right)&\frac{\partial}{\partial\sigma_1^2}t\left(\underline{\theta}\right)&\cdots&\frac{\partial}{\partial\sigma_k^2}t\left(\underline{\theta}\right)\end{bmatrix}=\begin{bmatrix}1&0&\cdots&0\end{bmatrix}$$

we have that

$$\begin{aligned}\nabla'\left(\underline{\theta}\right)&=\begin{bmatrix}F^{11}&F^{12}&\cdots&F^{1,k+1}\end{bmatrix}\\&=\begin{bmatrix}\frac{\gamma^2\left(1+2\gamma^2\right)}{2n^*k}&\frac{\gamma\sigma_1^2}{n^*k}&\frac{\gamma\sigma_2^2}{n^*k}&\cdots&\frac{\gamma\sigma_k^2}{n^*k}\end{bmatrix}\end{aligned}$$

and

$$\sqrt{\nabla_t'\left(\underline{\theta}\right)F^{-1}\left(\underline{\theta}\right)\nabla_t\left(\underline{\theta}\right)}=\left\{\frac{\gamma^2\left(1+2\gamma^2\right)}{2n^*k}\right\}^{\frac{1}{2}}.$$

Further

$$\Upsilon'\left(\underline{\theta}\right)=\frac{\nabla_t'\left(\underline{\theta}\right)F^{-1}\left(\underline{\theta}\right)}{\sqrt{\nabla_t'\left(\theta\right)F^{-1}\left(\underline{\theta}\right)\nabla_t\left(\underline{\theta}\right)}}=\begin{bmatrix}\Upsilon_1\left(\underline{\theta}\right)&\Upsilon_2\left(\underline{\theta}\right)&\cdots&\Upsilon_{k+1}\left(\underline{\theta}\right)\end{bmatrix}$$

where

$$\Upsilon_1(\underline{\theta}) = \frac{\gamma\left(1+2\gamma^2\right)^{\frac{1}{2}}}{(2n^*k)^{\frac{1}{2}}},$$

$$\Upsilon_2(\underline{\theta}) = \frac{(2)^{\frac{1}{2}}\sigma_1^2}{\left\{n^*k\left(1+2\gamma^2\right)\right\}^{\frac{1}{2}}},$$

$$\Upsilon_3(\underline{\theta}) = \frac{(2)^{\frac{1}{2}}\sigma_2^2}{\left\{n^*k\left(1+2\gamma^2\right)\right\}^{\frac{1}{2}}}$$

and

$$\Upsilon_{k+1}(\underline{\theta}) = \frac{(2)^{\frac{1}{2}}\sigma_k^2}{\left\{n^*k\left(1+2\gamma^2\right)\right\}^{\frac{1}{2}}}.$$

The prior

$$p_M(\underline{\theta}) = p_M\left(\gamma,\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2\right) \propto \frac{1}{|\gamma|\left(1+2\gamma^2\right)^{\frac{1}{2}}}\prod_{\tilde{m}=1}^{k}\sigma_{\tilde{m}}^{-2} \tag{C.1}$$

is therefore a probability-matching prior since

$$\frac{\partial}{\partial\gamma}\left\{\Upsilon_1(\underline{\theta})\,p_M(\underline{\theta})\right\} + \frac{\partial}{\partial\sigma_1^2}\left\{\Upsilon_2(\underline{\theta})\,p_M(\underline{\theta})\right\} + \cdots + \frac{\partial}{\partial\sigma_k^2}\left\{\Upsilon_{k+1}(\underline{\theta})\,p_M(\underline{\theta})\right\} = 0.$$

□

---

[1] Bayarri, M., Berger, J., 2004. The interplay of bayesian and frequentist analysis. Statistical Science 19 (1), 58–80.

[2] Bayarri, M., García-Donato, G., 2005. A bayesian sequential look at u-control charts. Technometrics 47(2), 142–151.

[3] Berger, J., Bernardo, J., 1992. On the development of reference priors. Bayesian Statistics 4, 35–60.

[4] Berger, J. O., Liseo, B., Wolpert, R. L., 1999. Integrated likelihood methods for eliminating nuisance parameters. Statistical Science 14, 1–28.

[5] Datta, G., Ghosh, J., 1995. On priors providing frequentist validity for bayesian inference. Biometrika 82 (1), 37–45.

[6] Iglewicz, B., 1967. Some properties of the coefficient of variation. Ph.D. thesis, Virginia Polytechnic Institute.

[7] Kang, C. W., Lee, M. S., Seong, Y. J., Hawkins, D. M., 2007. A control chart for the coefficient of variation. Journal of Quality Technology 39 (2), 151–158.

[8] Menzefricke, U., 2002. On the evaluation of control chart limits based on predictive distributions. Communications in Statistics - Theory and Methods 31(8), 1423–1440.

[9] Menzefricke, U., 2007. Control chart for the generalized variance based on its predictive distribution. Communications in Statistics - Theory and Methods 36(5), 1031–1038.

[10] Menzefricke, U., 2010. Control chart for the variance and the coefficient of variation based on their predictive distribution. Communications in Statistics - Theory and Methods 39(16), 2930–2941.

[11] Menzefricke, U., 2010. Multivariate exponentially weighted moving average chart for a mean based on its predictive distribution. Communications in Statistics - Theory and Methods 39(16), 2942–2960.

[12] Pearn, W., Wu, C., 2005. A bayesian approach for assessing process precision based on multiple samples. European Journal of Operational Research 165 (3), 685–695.

[13] Severine, T., Mukerjee, R., Ghosh, M., 2002. On an exact probability matching property of right-invariant priors. Biometrika 89 (4), 952–957.

[14] Tibshirani, R., 1989. Noninformative priors for one parameter of many. Biometrika 76 (3), 604–608.