1-1-2006

On computing Bayesian credible sets for the coefficient of variation of a normal population

Skanda Pokkunuri
University of Nevada, Las Vegas

ON COMPUTING BAYESIAN CREDIBLE SETS FOR THE COEFFICIENT OF VARIATION OF A NORMAL POPULATION

by

Skanda Pokkunuri

Master of Science Degree in Electrical Engineering
South Dakota School of Mines and Technology
2003

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science Degree in Mathematical Sciences
Department of Mathematical Sciences
College of Sciences

Graduate College
University of Nevada, Las Vegas
May 2006

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

Thesis Approval
The Graduate College
University of Nevada, Las Vegas

12/28, 2005

The Thesis prepared by

SKANDA POKKUNURI

Entitled

ON COMPUTING BAYESIAN CREDIBLE SETS FOR THE COEFFICIENT OF VARIATION OF A NORMAL POPULATION

is approved in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN MATHEMATICAL SCIENCES

Examination Committee Chair

Dean of the Graduate College

Examination Committee Member

Examination Committee Member

Graduate College Faculty Representative

1017-53

ABSTRACT

On Computing Bayesian Credible Sets for the Coefficient of Variation of a Normal Population

by

Skanda Pokkunuri

Dr. A.K. Singh, Examination Committee Chair
Professor of Statistics
University of Nevada, Las Vegas

The use of Coefficient of Variation (CV) is quite common in many disciplines, yet its estimation has not received much attention from statisticians. In this paper, we consider the problem of confidence interval estimation of CV for a normal population using the Bayesian approach. The method of Gibbs Sampler is used for numerical integration in order to compute the Bayes credible sets for CV for two different joint prior distributions – the natural conjugate prior, and the non-informative prior. Several simulated examples are included to demonstrate the proposed procedure.

Key Words: Natural conjugate prior, non-informative prior, Gibbs sampler, numerical integration.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to dedicate this work to my grandmother, mother and sister who have placated me during my failures and cherished my success.

A project of this magnitude can never be done in a vacuum. I'm very grateful to Dr.Ashok Singh, my mentor for spending many hours teaching me how to do things right and better, for his immense support, expert guidance and infallible tutelage.

I wish to thank Dr.Rohan Dalpatadu, Dr.Xin Li and Dr.Laxmi Gewali for being on my graduate committee despite giving them a short notice.

My thanks to Devarshi for his Buddha-like persistence that helped me with formatting the work and timely help whenever needed.

CHAPTER1

INTRODUCTION

1.1 Definition

Population standard deviation (σ) is a measure of dispersion or uncertainty in variable. Coefficient of variation is a relative measure of dispersion defined as the ratio σ/μ (provided population mean μ is different from zero); CV clearly has no unit of measurement. CV sometimes is expressed as a percentage. Even though the use of CV is quite common in many disciplines such as climatology (Singh *et al.,* 1987; Ananthakrishnan and Soman, 1989), business (De *et al.,* 1996; Hillier and So, 1991), immunology (Reed *et al.,* 2002), and engineering (Kwang, 1995; Zeevi, 1999), CV does not seem to get the attention of statisticians. Standard statistics text books may include the definition of CV, but do not mention how a confidence interval can be computed for population CV from available sample data. There is some literature available on conducting tests of hypotheses involving the CV (Doornbos and Dijkstra, 1983; Rao and Bhatt, 1989 and 1995; Rao and Vidya, 1992; Singh, 1993; Sharma and Krishnan, 1994; Ahmed, 1995; Rani, 1996; Gupta and Ma, 1996). The coefficient of variation is also often used to compare numerical distributions measured on different scales. For example, in a diet study, when the intent is to compare the variability in the ratio of total/HDL cholesterol with the variability in vessel diameter change, a comparison of standard deviation makes no sense because cholesterol and vessel diameter are measured in different scales. A sensible comparison can be made using the coefficient of variation since the coefficient of variation measures the relative spread of data and therefore adjusts the scale. For inferences about the coefficient of variation for a single sample from a normal population, an exact method for confidence intervals based on the non-central t distribution is available, but it is computationally cumbersome. For multiple sample cases, there exist several statistical tests for testing the equality of coefficients of variation. Fung and Tsang (1998) reviewed several parametric and non-parametric tests for the equality of coefficients of variations. Ahmed (1995) considered the problem of estimating the coefficients of variation when it is a priori suspected that two coefficients of variations are the same. However there is lack of literature on the inference procedures concerning the common population coefficient of variation based on several independent samples. Under many circumstances, we need to estimate the confidence intervals or perform hypothesis testing about the common population coefficient of variation from several examples.

1.2 Bayesian Approach

In this paper, the problem of computing a confidence interval for CV is considered using the Bayesian approach. In the Bayesian inference, the parameter of a distribution is a random variable; the Bayesian estimation of the unknown (random) parameter is typically more complicated than the classical estimation problem. The computation and interpretation of the confidence interval of an unknown parameter, however, is simpler in the Bayesian approach. The computation of the classical confidence interval typically begins with the search for a pivotal quantity involving the unknown parameter of interest. The computation of the Bayesian confidence interval, on the other hand, is quite straightforward.

The underlying population in this thesis is assumed to follow a normal distribution. The mean and variance of the normal distribution are considered to be random. The Bayesian credible sets for two different prior distributions have been derived. The joint posterior distribution of CV given the sample information turns out to be numerically intractable, and we resort to the method of Gibbs Sampler (Mary Ann Gregurich and Lyle D.Broemeling, 1997) for numerical integration in order to compute the Bayesian credible sets. The R programming language (http://www.r-project.org/) is used for writing a code for the Gibbs sampler. Several examples have been provided in the paper to illustrate the proposed method.

CHAPTER 2

PROBLEM DESCRIPTION AND DERIVATION OF BAYESIAN CREDIBLE SETS

2.1 Notations and Preliminaries

Let $\{x_1, x_2, \ldots, x_n\}$ be an independent random sample of size n from a normally distributed population with mean μ and standard deviation σ.

In the Bayesian framework, the parameters (μ,σ) are assumed random with a joint prior probability distribution g (μ,σ). In this paper, we will consider two different forms of the joint prior distribution (Berger, 1985): a natural conjugate prior, and a non-informative prior.

2.2 The Case of the Natural Conjugate Prior

Natural Conjugate Prior: For a given class of densities, a conjugate family can frequently be determined by examining the likelihood functions $lx(\theta) = f(x \mid \theta)$, and choosing, as a conjugate family, the class of distributions with the same functional form as these likelihood functions. The resulting priors are frequently called natural conjugates.

Suppose that $x=(x_{1}\ldots x_n)$ is a sample from a N (θ, σ^2) distribution, where both θ and σ^2 are unknown. The prior density of θ and σ^2 is given by

$\Pi(\theta, \sigma^2) = \Pi_1(\theta|\sigma^2) * \Pi_2(\sigma^2)$, where $\Pi_1(\theta|\sigma^2)$ is a N(μ, $\tau\sigma^2$) and $\Pi_2(\sigma^2)$ is an inverse gamma(α,β) density.

The joint posterior density of θ and σ^2 given x is

$g^*(\theta,\sigma^2 | \underline{x}) = \Pi_1(\theta|\sigma^2,\underline{x}))^* \Pi_2(\sigma^2|\underline{x})$. Where $\Pi_1(\theta|\sigma^2,\underline{x})$ is a normal density N

(μ^*,σ^{*2}), with mean μ^* and variance σ^{*2} where $\mu^* = \dfrac{\mu + n\tau\bar{x}}{n\tau+1}$, $\sigma^* = \sigma\sqrt{\dfrac{1}{\tau}+n}$ and

$\Pi_2(\sigma^2|\underline{x})$ is an inverted gamma density with parameters (α+n/2) and β' where

$$\beta' = [\beta^{-1} + ½ \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x}-\mu)^2/2(1+n\tau)]^{-1}.$$

2.3 The Case of the Non-informative Prior

Non-informative Prior: If prior information about the parameter is not available and we want a prior with minimal influence or no influence at all, such a prior is called noninformative prior.

Consider a random variable X from a normal distribution with a probability density function f(x) specified as:

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\Pi\sigma^2}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

As prior for μ and σ, we will use the non-informative prior (Berger, 1985):

$$g(\mu,\sigma) = \frac{1}{\sigma^2}.$$

Combining this prior with the likelihood function yields the posterior pdf for (μ and σ)

$g^*(\mu,\sigma \mid x) \;\alpha\; f(x \mid \mu,\sigma) g(\mu,\sigma)$

The joint posterior density of (μ and σ) is given by:

$$g^*(\mu,\sigma \mid x) \,\alpha\, \left(\frac{1}{\sigma^2}\right)^{(n+2)/2} \exp-\left[\sum_{i=1}^{n}(x_i-\mu)^2 / 2\sigma^2\right]$$

$$=\left(\frac{1}{\sigma^2}\right)^{(n+2)/2} \exp-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 + n(\mu-\bar{x})^2\right]$$

where $\bar{x}$ is the sample mean of x_i.

The above joint posterior is easily seen to be a product of

$$g^*(\theta, \sigma^2 \mid x_1, x_2, \ldots, x_n) = g^*_1(\theta \mid \sigma^2, x_1, x_2, \ldots, x_n) * g^*_2(\sigma^2 \mid x_1, x_2, \ldots, x_n)$$

where

$$g^*_1(\theta \mid \sigma^2, x_1, x_2, \ldots, x_n) = N(\bar{x}, \frac{\sigma^2}{n})$$

and

$$g^*_2(\sigma^2 \mid x_1, x_2, \ldots, x_n) = IG(\frac{n-1}{2}, \frac{2}{(n-1)s^2})$$

2.4 Gibbs Sampler for Computing the Bayesian Credible Set

In situations where the integration of the joint density is extremely difficult, Gibbs sampler has proven to be a good alternative. The Gibbs sampler generates a sample from the joint density by sampling instead from the conditional densities which are often known. According to Casella and George (1992), by generating a large enough sample, characteristics of the marginal density and even the density itself can be obtained.

2.5 Bootstrapping

The bootstrap is a resampling method for statistical inference which is commonly used to compute confidence intervals, but it can also be used for estimating bias and precision of an estimator or conduct hypothesis tests. Bootstrap methods have been used in environmental applications(Singh *et al.,* 1997), toxicology (Bailer and Oris, 1994), pollution modelling (Archer and Giovannoni, 1994; Cooley, 1997), chemometrics (Wehrens and Van der Linden, 1997), ecological studies (Dixon, 2001), and fisheries (Smith, 1997). For details of the various bootstrap methods, see Efron and Tibshirani (1993) or Davison and Hinkley (1997).

Let $\{x_1, x_2, \ldots, x_n\}$ be an independent random sample from a population with parameter θ for which a confidence interval needs to be computed. Let $\hat{\theta} = T(x_1, x_2, \ldots, x_n)$ be an estimator of θ. The bootstrap method is described below (see Singh *et al.,* 1997):

1) Generate a simple random sample $\{x^*_1, x^*_2, \ldots, x^*_n\}$ with replacement from the original sample $\{x_1, x_2, \ldots, x_n\}$, and compute $\hat{\theta}_i = T(x^*_1, x^*_2, \ldots, x^*_n)$ for the bootstrap sample $i, i = 1, 2, \ldots, K$.

2) Compute $\bar{\theta}_B = \dfrac{\sum_{i=1}^{K} \hat{\theta}_i}{K}$, $\quad \hat{\sigma}_B = \sqrt{\dfrac{\sum_{i=1}^{K} (\hat{\theta}_i - \bar{\theta}_B)^2}{K-1}}$.

3) The 100(1-α)% percentile bootstrap confidence interval is obtained by sorting the $\hat{\theta}_i = T(x^*_1, x^*_2, \ldots, x^*_n)$ values in increasing order, and then computing:

$$L = [\frac{\alpha}{2} \times K]$$

$$U = [\frac{1-\alpha}{2} \times K]$$

where

[a] = largest integer in the real number a.

4) The 100(1-α)% standard bootstrap confidence interval is given by the formula:

$$L = \hat{\theta} - z_{1-\frac{\alpha}{2}} \hat{\sigma}_B$$

$$U = \hat{\theta} + z_{1-\frac{\alpha}{2}} \hat{\sigma}_B$$

CHAPTER 3

ALGORITHM DESCRIPTION

In this chapter, we give algorithmic descriptions of the proposed procedure for computing the Bayes credible set via the Gibbs sampler.

3.1 The Case of Natural Conjugate Prior:

1. Input the prior parameter values α, β, μ, τ

2. Generate one value σ_0 from inverse gamma density with (α, β) as parameters.IG (α, β).

3. Generate one value θ_0 from the conditional pdf of θ given $\sigma = \sigma_0$, which is N ($\mu, \tau\sigma^2_0$).

4. Generate $x_1, x_2 \ldots x_n$ from the conditional pdf N (θ_0, σ_0).

5. Compute $CV_0 = \sigma_0 / \theta_0$, which is the CV to be estimated.

6. We next generate K samples (θ_i, σ_i^2) from the joint posterior distribution

$g^*(\theta, \sigma^2 \mid x_1, x_2, \ldots, x_n) = g^*_1(\theta \mid \sigma^2, x_1, x_2, \ldots, x_n)\ g^*_2(\sigma^2 \mid x_1, x_2, \ldots, x_n)$

where $g^*_1(\theta \mid \sigma^2, x_1, x_2, \ldots, x_n) = N(\mu^*, \sigma^{*2})$, with

$$\mu^* = \frac{\mu + n\tau\bar{x}}{n\tau + 1}, \qquad \sigma^* = \sigma\sqrt{\frac{1}{\tau} + n}$$

and

$g^*_2(\sigma^2 | x_1, x_2, \ldots, x_n)$ = Inverted Gamma pdf with parameters (α^*, β^*)

with

$$\alpha^* = \alpha + \frac{n}{2}, \quad \beta^* = 1/[\frac{1}{\beta} + \frac{(n-1)s^2}{2} + \frac{n(\bar{x}-\mu)^2}{2(1+n\tau)}]$$

and for each generated sample, we compute $CV_i = \dfrac{\sigma_i}{\theta_i}$ for each i = 1, 2, …, K.

7. The K values of CV are sorted in increasing order:

$CV_{(1)} \le CV_{(2)} \le \ldots \le CV_{(K)}$.

8. We used K = 1000 in this paper, in which case the percentile bootstrap confidence interval (Bayesian credible set) for 95% confidence was calculated as:

$L_P = CV_{(25)}, \quad U_P = CV_{(975)}$

9. We also compute the standard bootstrap confidence interval as follows:

$L_S = \bar{C} - 1.96 \times s_{CV}$
$U_S = \bar{C} + 1.96 \times s_{CV}$
where
$\bar{C} =$ sample mean of $CV_1, CV_2, \ldots, CV_K$
$s_{CV} =$ sample standard deviation of $CV_1, CV_2, \ldots, CV_K$

3.2 The Case of Non-informative Prior

INPUT: Sample size n, population mean θ, population standard deviation σ

1. Generate $x_1, x_2 \ldots x_n$ from the conditional pdf N (θ,σ), and compute sample mean $\bar{x}$ and sample standard deviation s.

2. We next generate K samples (θ_i, σ_i^2) from the joint posterior distribution

$g^*(\theta, \sigma^2 | x_1, x_2, \ldots, x_n) = g^*_1(\theta | \sigma^2, x_1, x_2, \ldots, x_n) * g^*_2(\sigma^2 | x_1, x_2, \ldots, x_n)$

where

$g^*_1(\theta \mid \sigma^2, x_1, x_2, \ldots, x_n) = N(\bar{x}, \frac{\sigma^2}{n})$

and

$g^*_2(\sigma^2 \mid x_1, x_2, \ldots, x_n) = IG(\frac{n-1}{2}, \frac{2}{(n-1)s^2})$

by first generating $\sigma_i^2 \sim IG(\frac{n-1}{2}, \frac{2}{(n-1)s^2})$ and then generating $\theta_i \sim N(\bar{x}, \frac{\sigma_i^2}{n})$.

and for each sample, we compute $CV_i = \frac{\sigma_i}{\theta_i}$ for each i = 1, 2, …, K.

3. The K values of CV are sorted in increasing order:

$CV_{(1)} \le CV_{(2)} \le \ldots \le CV_{(K)}$.

4. We used K = 1000 in this paper, in which case the percentile bootstrap confidence interval (Bayesian credible set) for 95% confidence was calculated as:

$L_P = CV_{(25)}, \quad U_P = CV_{(975)}$

CHAPTER 4

SIMULATIONS

4.1 Examples

The R programming language (http://www.r-project.org/) was used for writing a code for the Gibbs sampler. The codes are included in Appendices A and B. We now present the results obtained for a few simulated examples. For each example, we have used the following values for the parameters of the natural conjugate prior: $\alpha = 10$, $\beta = 2$, $\mu = 5$,

$\tau = 1.2$.

EXAMPLE 1: In this example, sample size n was taken to be 20. The value of σ_0^2 generated from IG(10, 2) prior distribution was 0.2081527, and the value of θ_0 generated from conditional distribution $N(\mu, \tau\sigma_0^2)$ was 4.534036, resulting in a $CV_0 = 0.1006250$.

The sample of size n = 20 generated from $N(\theta_0, \sigma_0^2)$ is:

3.938558 4.783378 4.109437 3.954994 4.334185 4.089939 4.516363 4.264087

4.622989 5.042685 4.374686 4.435917 4.936095 4.927059 4.711183 4.371529

4.960497 4.595111 3.842991 4.781038

The sample mean and sample sd of the above sample are .479636 and 0.3687259, respectively.

The 95% Bayes credible set from the percentile bootstrap method turned out to be:

L95 = 0.05632081

U95 = 0.08760798

The 95% Bayes credible set from the standard bootstrap method turned out to be:

L95 = 0.05384001

U95 = 0.0856592

Histogram of CV values for this example (n = 20) is shown in Figure 1.

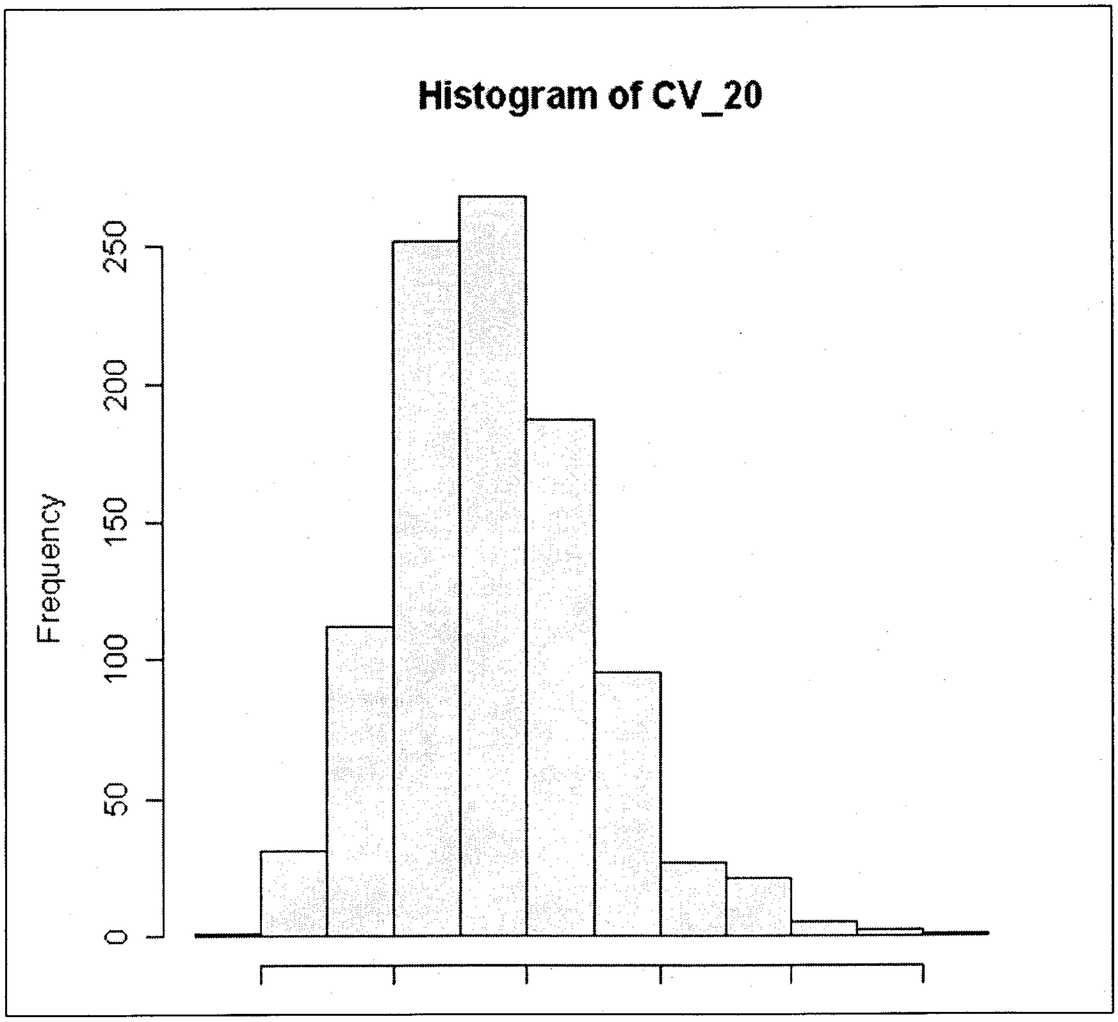

Figure 1: Histogram of 1000 CV values generated for n = 20 for data of example 1.

EXAMPLE 2: In this example, sample size n was taken to be 40. The value of σ_0^2 generated from IG(10, 2) prior distribution was 0.1761724, and the value of θ_0 generated from conditional distribution $N(\mu, \tau\sigma_0^2)$ was5.123818, resulting in a CV_0 = 0.08191722.

The sample of size n = 40 generated from $N(\theta_0, \sigma_0^2)$ is:

5.195091 4.901161 5.904814 4.703082 5.277199 4.421394 5.015085 5.133922

4.989313 4.533240 5.196308 5.167518 5.397499 5.779703 4.642173 4.772840

4.491819 4.932537 5.676259 5.641911 5.607806 4.674985 5.119727 4.851226

5.697781 4.907040 5.638857 5.318878 5.114785 4.511287 5.261230 4.403297

4.360544 4.895447 4.939569 4.886122 4.556046 4.634467 5.253661 4.596254

The sample mean and sample sd of the above sample are 5.025047 and 0.4209097, respectively.

The 95% Bayes credible set from the percentile bootstrap method turned out to be:

L95 =0.05902769

U95 = 0.08735178

The 95% Bayes credible set from the standard bootstrap method turned out to be:

L95 = 0.06070841

U95 = 0.08948844

Histogram of CV values for this example (n = 40) is shown in Figure 2.

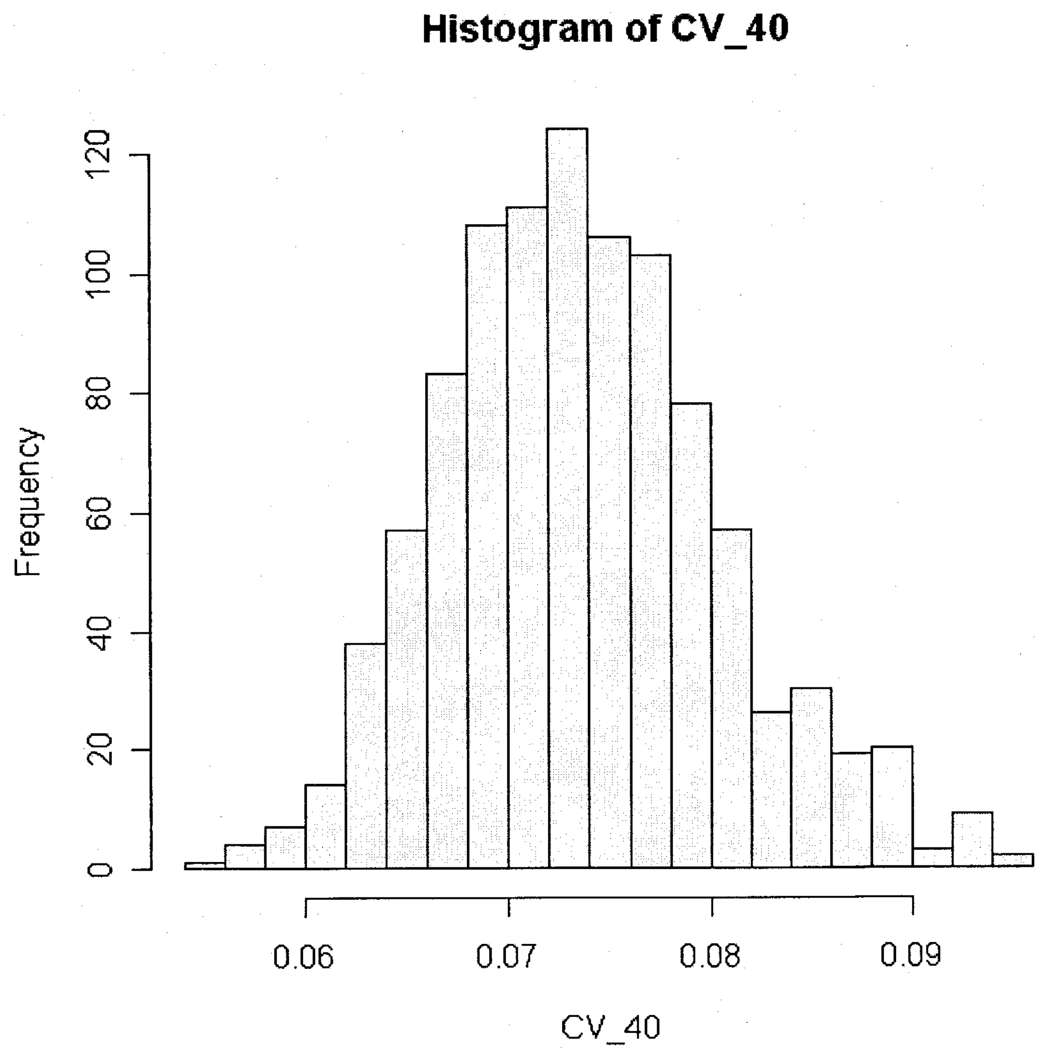

Figure 2: Histogram of 1000 CV values generated for n = 40 for data of example 2.

EXAMPLE 3: In this example, sample size n was taken to be 60. The value of σ_0^2 generated from IG(10, 2) prior distribution was 0.1725316, and the value of θ_0 generated from conditional distribution $N(\mu, \tau\sigma_0^2)$ was5.444339 , resulting in a CV_0 = 0.07629379.

The sample of size n = 60 generated from $N(\theta_0, \sigma_0^2)$ is:

5.334515 4.919074 4.587078 4.847087 4.832990 5.980627 5.251696 5.010982
5.335300 4.790179 4.590599 5.117517 6.349936 5.370224 5.791562 5.273810
5.787132 4.950041 5.952299 5.040438 5.603921 4.712257 4.974523 5.243142
5.097183 5.422368 5.177294 5.446500 6.009141 5.455782 5.930678 4.776485
5.924238 6.489499 5.262082 4.501793 5.624766 5.124291 5.644083 5.612753
6.022135 5.635454 6.080871 5.923877 4.908210 6.085459 6.206678 6.099297
4.651845 5.428960 5.508126 5.010765 5.804394 5.057767 5.759914 4.966113
5.666293 5.385872 5.225384 5.065815

The sample mean and sample sd of the above sample are 5.393985 and 0.4870752 , respectively.

The 95% Bayes credible set from the percentile bootstrap method turned out to be:

L95 = 0.06808912

U95 = 0.09515453

The 95% Bayes credible set from the standard bootstrap method turned out to be:

L95 = 0.0692972

U95 = 0.09601882

Histogram of CV values for this example (n = 60) is shown in Figure 3.

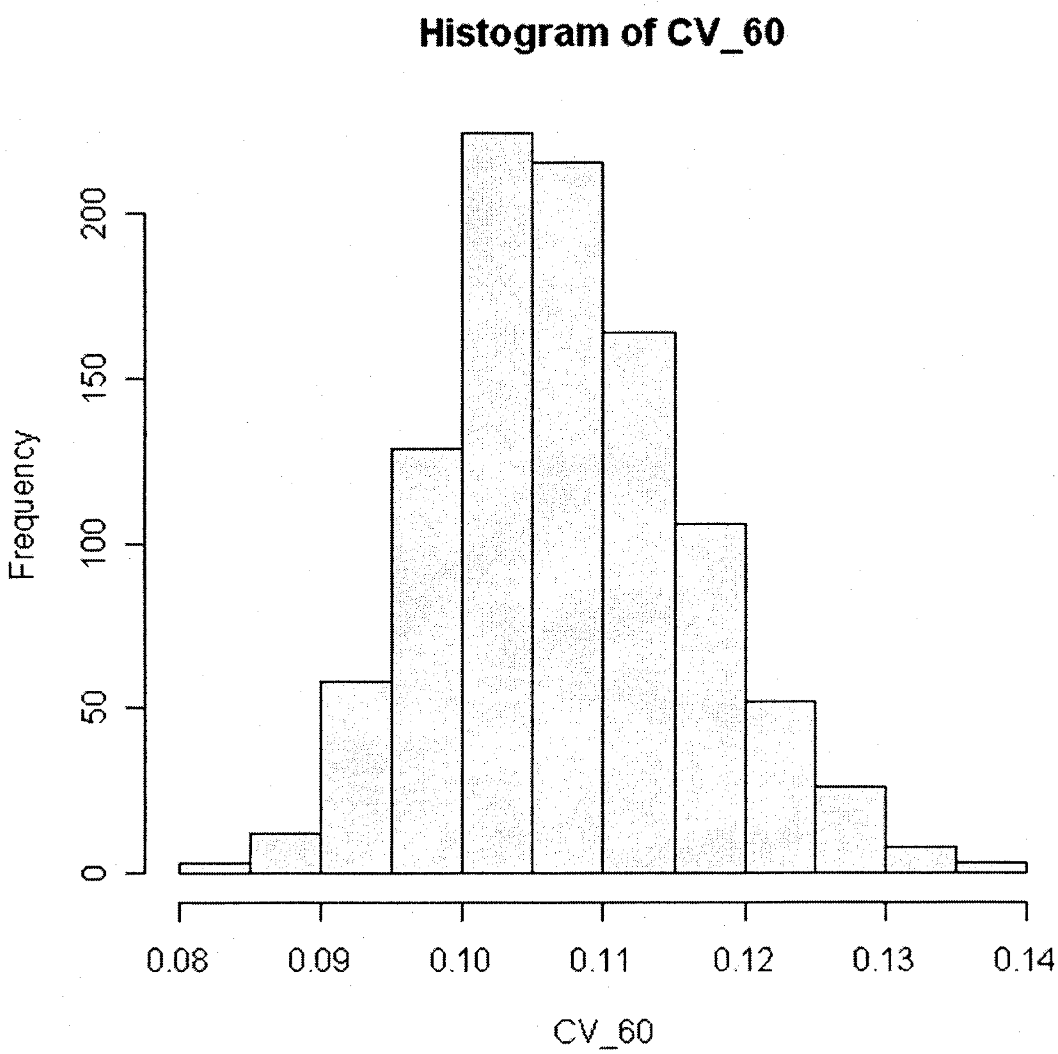

Figure 3: Histogram of 1000 CV values generated for n = 60 for data of example 3.

EXAMPLE 4: In this example, sample size n was taken to be 80. The value of σ_0^2 generated from IG(10, 2) prior distribution was 0.2108031, and the value of θ_0 generated from conditional distribution $N(\mu, \tau\sigma_0^2)$ was 3.929149, resulting in a CV_0 = 0.1168530.

The sample of size n = 80 generated from $N(\theta_0, \sigma_0^2)$ is:

3.999052 4.408115 3.298024 3.968417 3.076839 3.958106 3.700058 4.678890
3.505135 3.555600 4.065826 3.189667 4.448869 5.022847 4.114886 4.329122
4.548362 3.520170 2.998336 4.432888 4.088477 4.846503 3.399125 3.575701
3.431566 4.575279 4.324783 4.736411 3.836589 3.989737 4.054211 4.294270
4.582821 3.402545 3.640554 4.497816 3.406731 3.004748 4.034646 4.306560
4.356074 3.463942 3.980844 3.768406 4.433256 4.008923 3.822448 4.029907
4.123775 3.848241 5.121104 3.149490 4.440706 3.797326 4.599612 4.291882
3.951572 3.729198 4.703178 3.613449 2.976407 4.345485 4.440940 3.947941
3.663664 4.045704 3.900019 4.147354 3.856574 4.175532 4.135150 3.697708
3.511834 3.973835 4.675573 4.775284 4.088820 4.729956 3.509026 3.506224

The sample mean and sample sd of the above sample are 4.002308 and 0.4975223, respectively.

The 95% Bayes credible set from the percentile bootstrap method turned out to be:

L95 = 0.09938383

U95 = 0.1328529

The 95% Bayes credible set from the standard bootstrap method turned out to be:

L95 = 0.1016182

U95 = 0.1339338

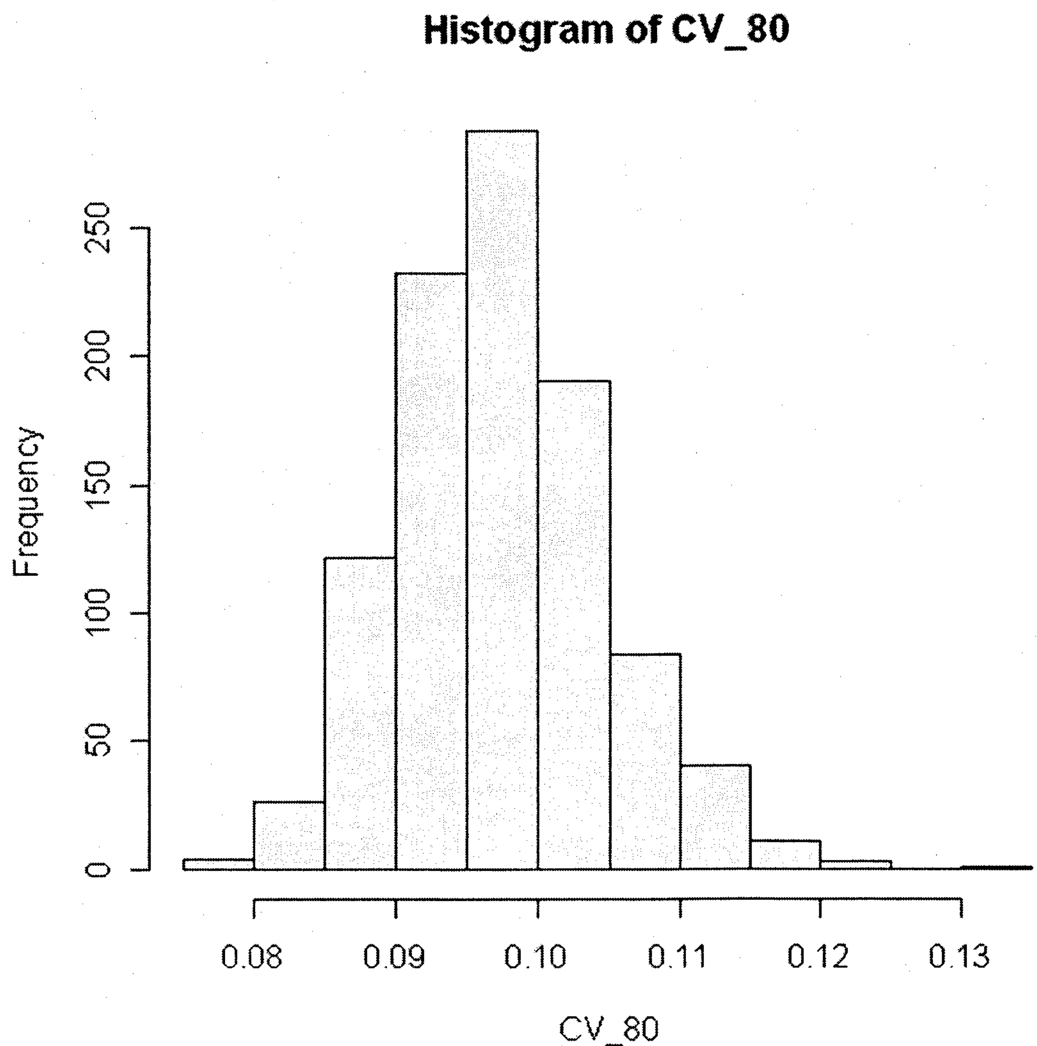

Figure 4: Histogram of 1000 CV values generated for n = 80 for data of example 4.

EXAMPLE 5: In this example, sample size n was taken to be 100. The value of σ_0^2 generated from IG(10, 2) prior distribution was0.2168264 , and the value of θ_0 generated from conditional distribution $N(\mu, \tau\sigma_0^2)$ was 4.702729, resulting in a CV_0 = 0.09901618.

The sample of size n =100 generated from $N(\theta_0, \sigma_0^2)$ is:

3.956778 4.442984 4.211406 4.418205 4.544092 4.457706 4.034471 3.855637
4.637847 5.447485 4.007945 4.518593 3.988190 4.425846 5.035207 3.814567
5.230892 4.412919 4.389110 4.561933 3.976510 4.375612 4.165671 5.025080
5.232845 4.342760 5.439756 4.385766 4.846016 4.066312 4.812685 4.035827
4.592830 4.942523 4.501829 5.021291 4.291405 5.004731 4.955971 5.050640
4.761112 3.959176 6.215485 4.898647 4.993158 4.699749 4.332323 4.394696
4.542854 4.398722 4.110449 4.728602 4.824840 3.946871 5.177165 4.367280
4.502533 4.376623 4.017201 5.286946 4.171814 5.428869 5.091261 4.571737
4.219236 5.529405 4.674859 3.677564 4.949650 4.879857 4.533523 4.922928
4.774677 4.996976 4.161086 4.343213 4.371833 4.515768 5.057718 3.958922
4.719948 4.576818 4.260100 4.613944 4.858702 4.672026 4.781582 4.539652
5.121245 4.490817 5.276982 5.787626 3.638489 5.125076 3.788655 5.076870
3.855569 4.567685 4.378434 5.191193

The sample mean and sample sd of the above sample are 4.601446 and0.4841811, respectively.

The 95% Bayes credible set from the percentile bootstrap method turned out to be:

L95 = 0.08594355

U95 = 0.1111672

The 95% Bayes credible set from the standard bootstrap method turned out to be:

L95 = 0.08639479

U95 = 0.1114016

Histogram of CV values for this example (n = 100) is shown in Figure 5.

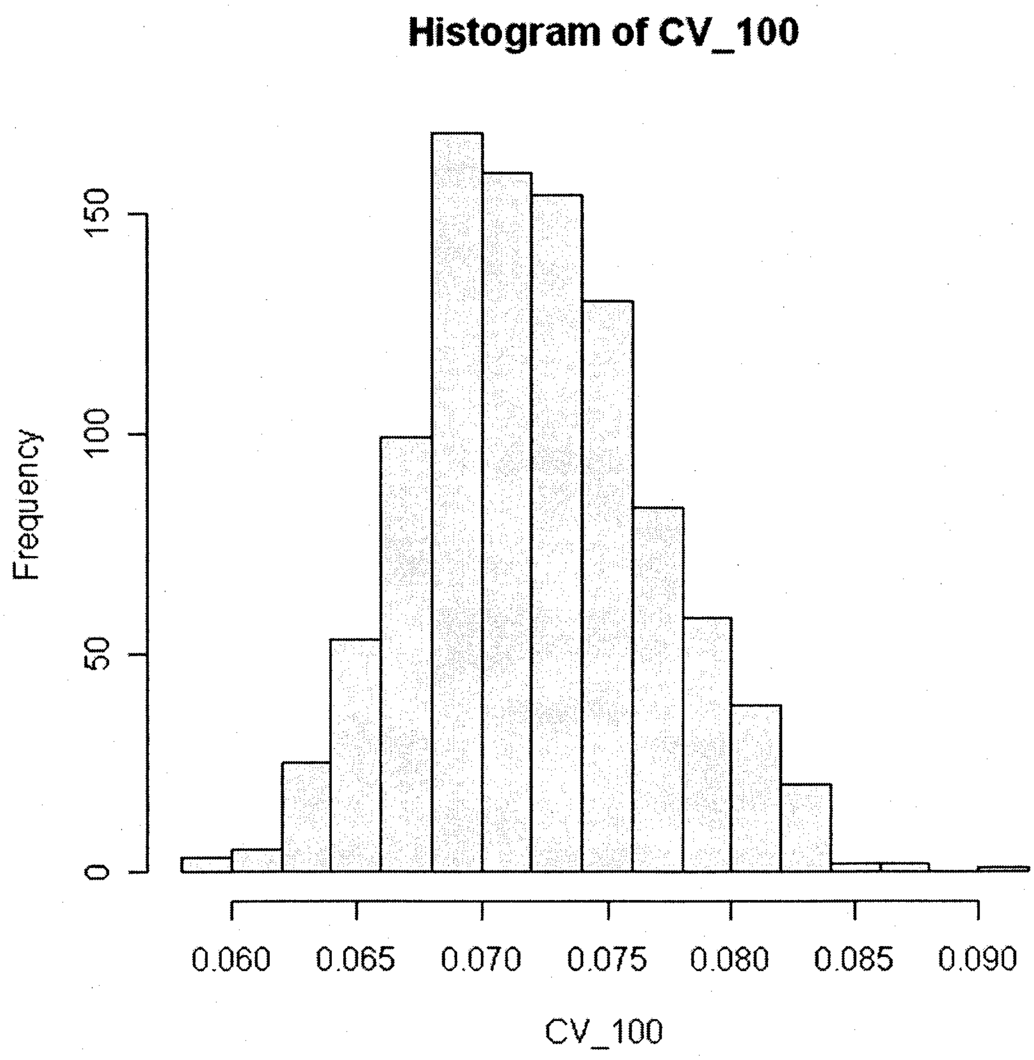

Figure 5: Histogram of 1000 CV values generated for n = 100 for data of example 5.

Examples of non-informative prior

EXAMPLE 1: In this example, sample size n was taken to be 30. The value of $\sigma_0 = 5$ and

the value of θ_0 is 50, resulting in a $CV_0 = 0.1$.

The sample mean and sample sd of the above sample are 49.72061and 4.357471 and respectively.

From the sorted values obtained it can be seen that CV0 is captured between

L= 0.05850385

U=0.13388827.

Where L and U are the lower and upper values of the computed CV.

Histogram of CV values for this example (n=30) is shown in Figure 6.

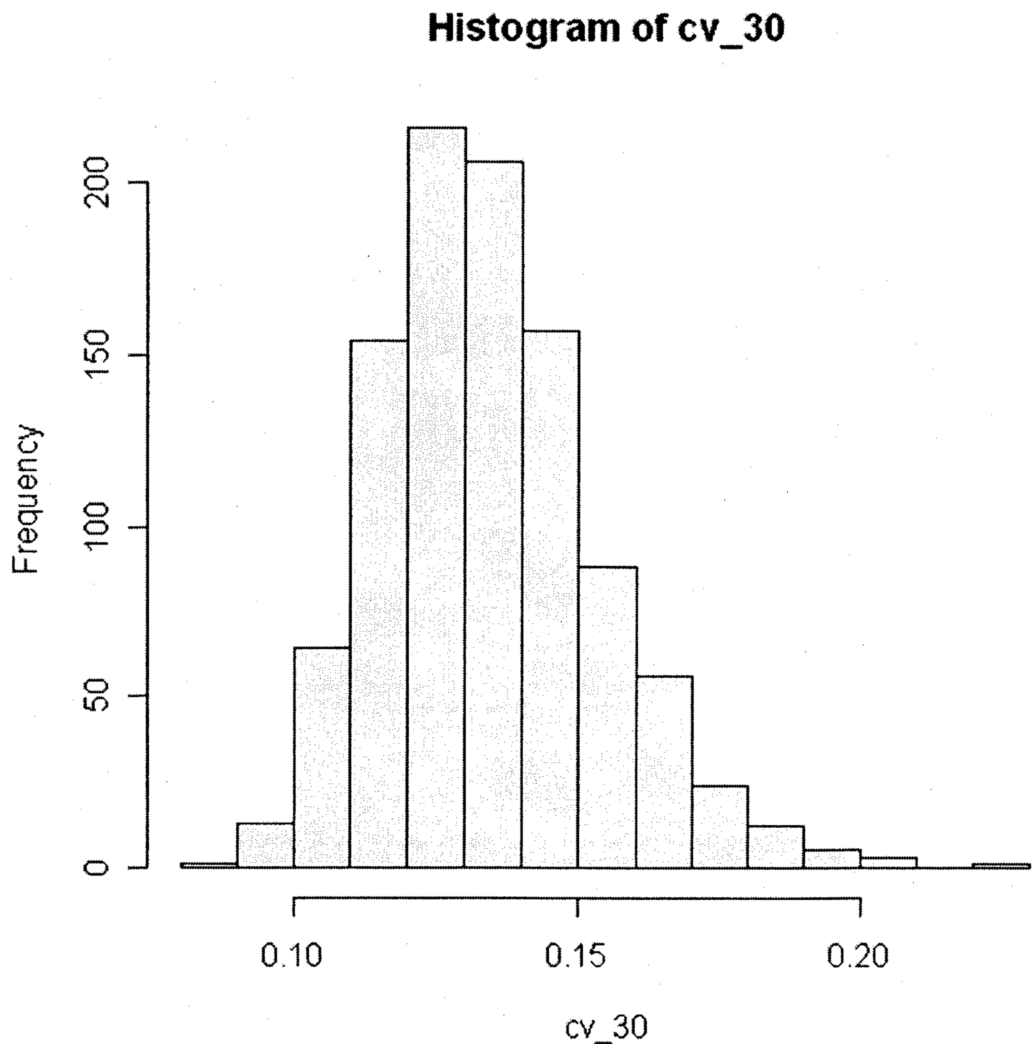

Figure 6: Histogram of 1000 CV values generated for n = 30 for data of example 6.

EXAMPLE 2: In this example, sample size n was taken to be 50. The value of $\sigma_0 = 5$ and

the value of θ_0 is 25 , resulting in a $CV_0 = 0.2$.

The sample mean and sample sd of the above sample are 24.38152 and 5.242306 respectively.

From the sorted values obtained it can be seen that CV0 is captured between

L= 0.1563942

U=0.3081829.

Where L and U are the lower and upper values of the computed CV.

Histogram of CV values for this example (n=50) is shown in Figure 7.

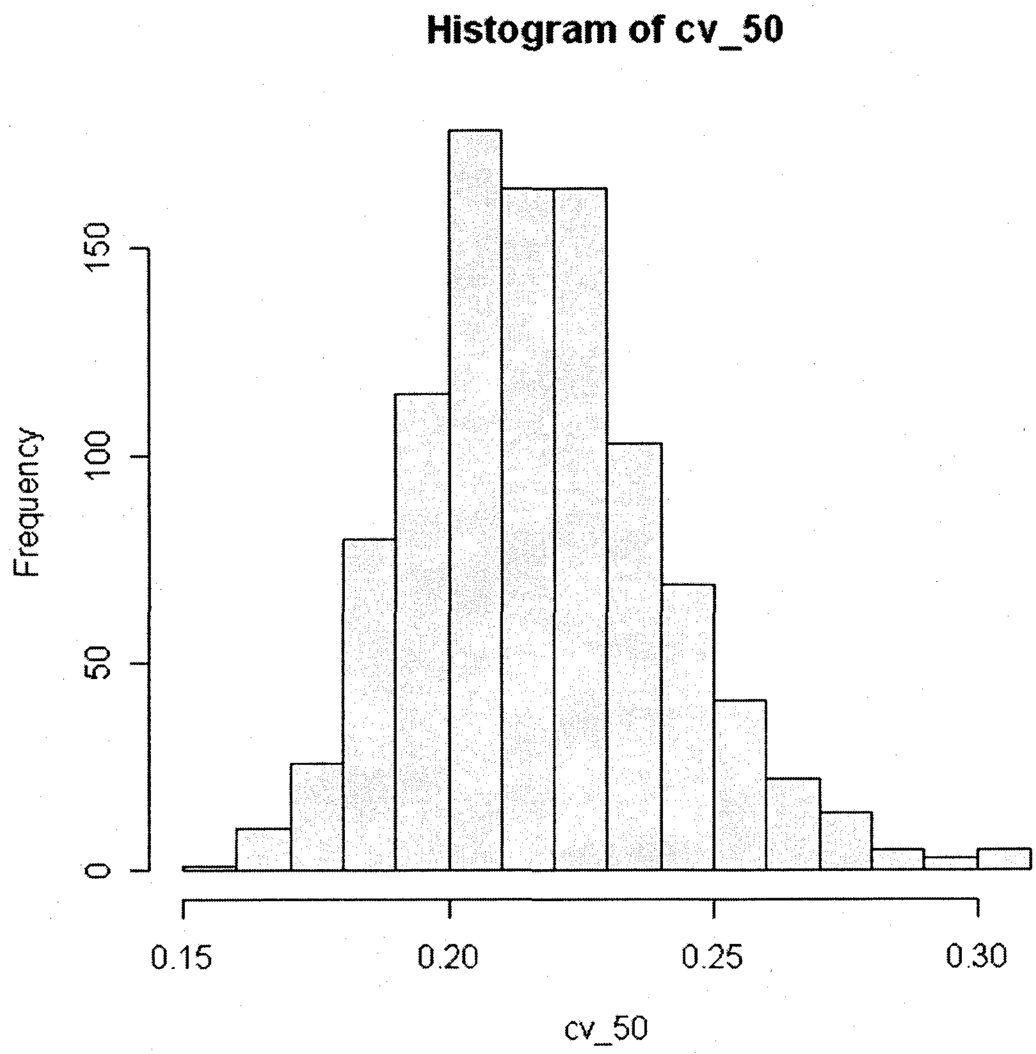

Figure 7: Histogram of 1000 CV values generated for n = 50 for data of example 7.

EXAMPLE 3: In this example, sample size n was taken to be 70. The value of $\sigma_0 = 5$ and the value of θ_0 is 12.5 , resulting in a $CV_0 = 0.4$.

The sample mean and sample sd of the above sample are 12.80022 and 4.55282 respectively.

From the sorted values obtained it can be seen that CV0 is captured between

L= 0.2695582

U= 0.5235603.

Where L and U are the lower and upper values of the computed CV.

Histogram of CV values for this example (n=70) is shown in Figure 8.

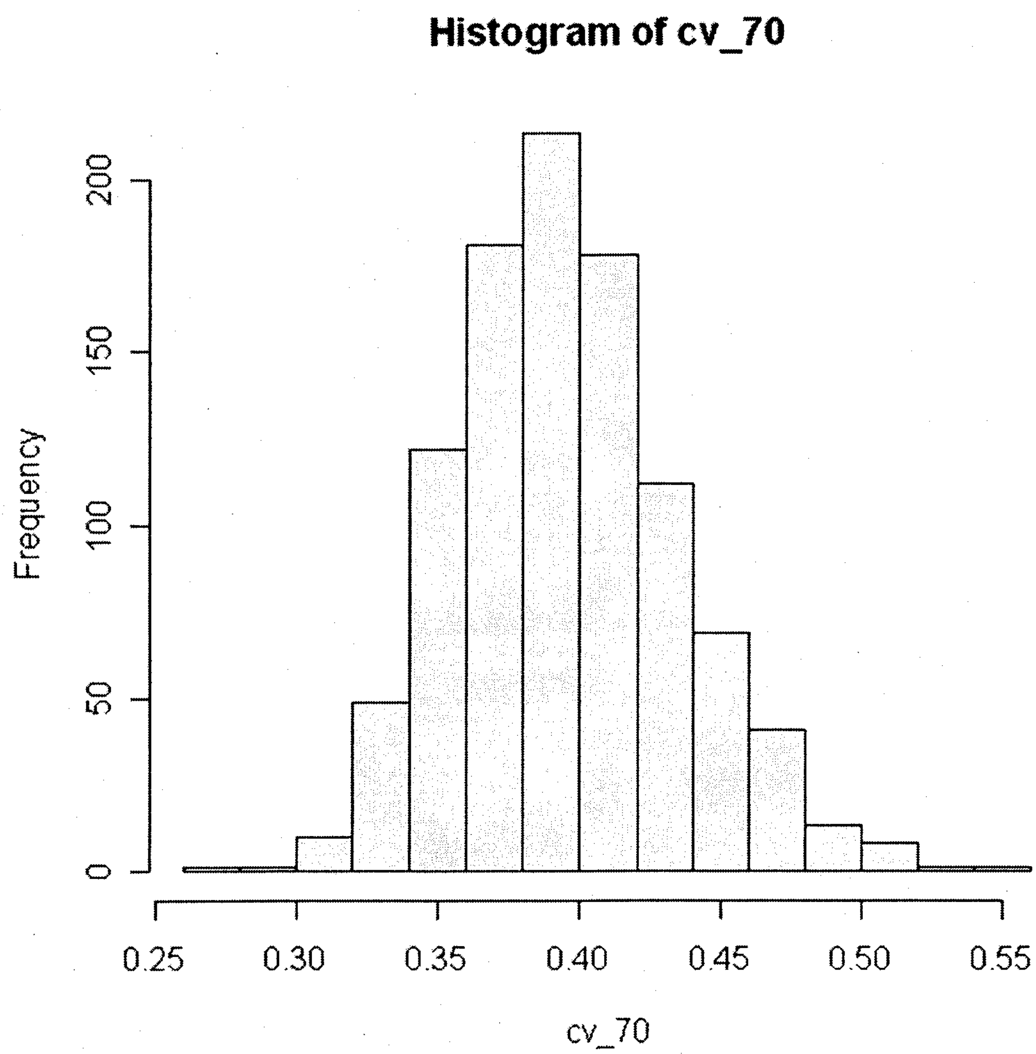

Figure 8: Histogram of 1000 CV values generated for n = 70 for data of example 8.

EXAMPLE 4: In this example, sample size n was taken to be 90. The value of $\sigma_0 = 10$ and the value of θ_0 is 10 , resulting in a $CV_0 = 1$.

The sample mean and sample sd of the above sample are 9.433877 and 9.910639 respectively.

From the sorted values obtained it can be seen that CV0 is captured between

L= 0.7692960.

U= 1.8238918.

where L and U are the lower and upper values of the computed CV.

Histogram of CV values for this example (n=90) is shown in Figure 9.

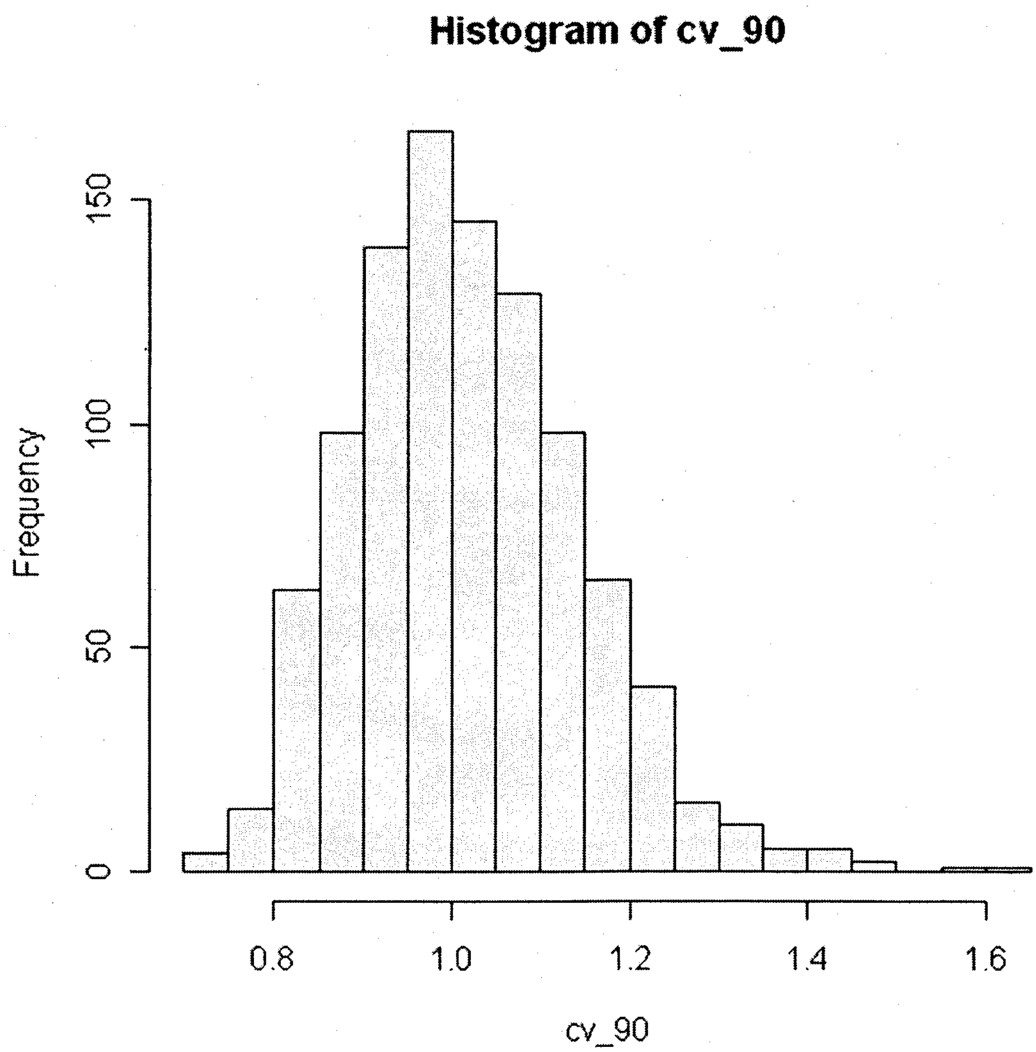

Figure 9: Histogram of 1000 CV values generated for n = 90 for data of example 9.

4.2 Conclusions

In this paper, we have shown how to compute Bayes credible sets using the Gibbs sampler for the coefficient of variation (CV) when the underlying conditional distribution is normal. From the simulated examples presented in this paper, we can see that the Bayes credible set captures the true CV value. The approach proposed used can also be used, with slight modifications, for the log-normal distribution, which is very commonly used in environmental applications.

APPENDIX A

CODE FOR NATURAL CONJUGATE PRIOR

Joint prior pdf is natural conjugate g(theta, sigma_sqr)

g(theta, sigma_sqr) = g1(theta|sigma_sqr) X g2(sigma_sqr)

g1(theta|sigma_sqr) = N(mu, tauXsigma_sqr), (mu, tau) are known

g2(sigma_sqr) = IG(alpha, beta)

Joint posterior pdf is:

g*(theta, sigma_sqr|x1, ...,xn) = g1*(theta|sigma_sqr, x1, ...,xn)Xg2*(sigma_sqr| x1, ...,xn)

Joint prior pdf is natural conjugate g(theta, sigma_sqr)

g(theta, sigma_sqr) = g1(theta|sigma_sqr) X g2(sigma_sqr)

g1(theta|sigma_sqr) = N(mu, tauXsigma_sqr), (mu, tau) are known

g2(sigma_sqr) = IG(alpha, beta)

Joint posterior pdf is:

g*(theta, sigma_sqr|x1, ...,xn) = g1*(theta|sigma_sqr, x1, ...,xn)Xg2*(sigma_sqr| x1, ...,xn)

g1*(theta|sigma_sqr, x1, ...,xn) = N(mu*, sigma_sqr*)

mu* = (mu + n X tau X xbar)/(nXtau + 1)

sigma_sqr* = sigma_sqr X (1/tau + n)

and g2*(sigma_sqr| x1, ...,xn) = IG(alpha*, beta*) alpha* = alpha + n/2,

1/beta* = 1/beta + (n-1)*s_sqr/2 + n(xbar-mu)X(xbar-mu)/[2(1+nXtau)]

Input prior parameters

alpha<-10

beta<-2

mu<-5

tau<-1.2

Generate sigma0_sqr ~ g2(sigma_sqr)

vinv<-rgamma(1, alpha, beta)

sigma0_sqr<-1/vinv

print(sigma0_sqr)

psd = prior conditional sd

psd<-sqrt(tau*sigma0_sqr)

Generate theta0 ~ g1(theta|sigma_sqr) = N(mu, tauXsigma_sqr),

theta0<-rnorm(1, mu, psd)

print(theta0)

CV0<-sqrt(sigma0_sqr)/theta0

print('CV0')

print(CV0)

generate x1, …, xn ~ N(theta0, sigma0) conditional pdf

```
n <- 20
x1<-rnorm(20, theta0, sqrt(sigma0_sqr))
print(x1)
mean1<-mean(x1)
print(mean1)
std1<-sd(x1)
print(std1)
cv1 <-array(0, c(0,1000))
srtcv1 <- array(0, c(0,1000))
mu1 = (mu + n * tau * mean1)/(n*tau + 1)
print(mu1)
alpha1 <- alpha + n/2
beta1_inv<-(1/beta) + (n-1)*std1*std1/2 + n*(mean1 - mu)*(mean1 -
mu)/(2*(1+n*tau))
beta1 <- beta1_inv
in R, gamma scale parameter is inverse of scale parameter in Berger's book
print(alpha1)
print(beta1)
generate K samples from g*(theta, sigma_sqr|sample), posterior pdf
k <- 1000
for (i in 1:k)
{ v <- rgamma(1, alpha1, beta1)
 print(v)
```

```
 sigma1_sqr<- 1/v
  sigma1 <- sqrt(sigma1_sqr)
   post_sd <- sigma1_sqr/(1/tau + n)
  theta1<- rnorm(1, mu1, sqrt(post_sd))
  cv1[i] <- sigma1/theta1
}
print(cv1)
srtcv1 <- sort(cv1)
print(srtcv1)
L95<-srtcv1[25]
U95<-srtcv1[975]
print("95% Credible LImits for CV")
print(L95)
print(U95)
cvmean<-mean(cv1)
cvmean1<-mean(srtcv1)
print(cvmean1)
cvstd<-sd(srtcv1)
print(cvmean)
print(cvstd)
print(L95)
print(U95)
cvmean<-mean(cv1)
```

```
sdcv<-sd(cv1)
print(cvmean)
print(sdcv)
L95a<-cvmean-1.96*cvstd
U95a<-cvmean+1.96*cvstd
print('approx 95% Credible Limits for CV')
print(L95a)
print(U95a)
```

APPENDIX B

CODE FOR NON-INFORMATIVE PRIOR

```
cv<-array(0, c(0,1000))
srtcv<-array(0, c(0,1000))
n<-70 # sample size
x<-rnorm(n, 12.5, 5)
cv0<-5/12.5
xbar<-mean(x)
s<-sd(x)
print(xbar)
print(s)
n1<-(n-1)/2
print(n1)
beta<-n1*s*s
#: NOTE: in R, if X~gamma(alpha, beta), then E(X)=alpha/beta
print(beta)
for (i in 1:1000)
{
# generate sigma2~IG( (n-1)/2, 2/[(n-1) s_square] )
xgamma<-rgamma(1,n1, beta)
```

```
sigma2<-1/xgamma
sigma<-sqrt(sigma2)
#now that sigma2 has been generated from IG( (n-1)/2, 2/[(n-1)s_square] ),
# generate theta (population mean) ~ N(xbar, sigma2/n)
sd_theta<-sqrt(sigma2/n)
theta<-rnorm(1, xbar, sd_theta)
cv[i]<-sigma/theta
# print(sigma)
# print(theta)
# print(cv[i])
}
srtcv<-sort(cv)
print(srtcv)
```

REFERENCES

Ahmed SE (1995). A pooling methodology for coefficient of variation. Sankhya. 57:57-75.

Archer, G. & Giovannoni, J.-M. (1998). Statistical analysis with bootstrap diagnostics of atmospheric pollutants predicted in the APSIS experiment, *Water, Air,and Soil Pollution* 106, 43–81.

Bailer, A.J. & Oris, J.T. (1994). Assessing toxicity of pollutants in aquatic systems, in *Case Studies in Biometry*, N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest & J. Greenhouse, eds, Wiley, New York, pp. 25–40.

Berger, James O (1985). Statistical Decision theory and Bayesian Analysis, Springer-verlag NewYork Inc. Second Edition.

Casella , G and George ,E.I. (1992). Explaining the Gibbs Sampler .The American Statistician, 46, 167-174.

Chaturvedi , A., Rani, U.(1996). Fixed width confidence interval estimation of the inverse coefficient of variation in a normal population. Microelectronics and reliability 36:1305-1308.

De, P., Ghosh (1996). Scheduling to minimize the coefficient of variation. International Journal of Production Economics 44:429-253.

Dixon, P.M. (2001). The bootstrap and the jackknife: Describing the precision of ecological studies, in *Design and Analysis of Ecological Experiments*, 2nd Edition, S. Scheiner & J. Gurevitch, eds, Oxford University Press, Oxford.

Doornbos,R., Dijkstra , J.B (1983). A multi sample test for the equality of coefficients of variation in normal populations. Commun.statist.- simula.computa 12:147-158.

Fung WK, Tsang TS (1998). A simulation study comparing tests for the equality of coefficients of variation. Statistics in medicine ; 17:2003-2014.

Gregurich, Mary Ann, and Lyle D.Broemeling (1997). A Bayesian analysis for estimating the common mean af independent normal populations using the Gibbs sampler; Commun.Statist.-Simula., 26, 35-51.

Gupta, C.R., Ma, S. (1996). Testing the equality of coefficients of variation in K normal populations. Common. Statist. Theory meth. 25:115-132.

Hillier, F.S., So K.C. (1991). The effect of the coefficient of variation of operation times on the allocation of storage space. I.I.E. Transactions 23:198-206 .

Kwang, I.A. (1995). On use of coefficient of variation for uncertainty analysis in fault tree analysis. Reliability engineering and system safety 47:229-230.

Sharma, K.K., Krishnan (1994): Asymptotic sampling distribution of inverse coefficient of variation and its applications. IEEE-Transactions on reliability 43:630-633.

Singh, W., Soman, M.K., Kumar (1987).Use of coefficient of variation in determining rainfall probabilities in a humid region . Mausam 38:261-264

Rao, K. A., Bhatt (1989). Tests for coefficient of variation. Journal of the Indian society of agricultural statistics XLVII :225-229.

Rao, K. A ., Vidya, R .(1992). On the performance of a test for coefficient of variation . Calcutta Statistical Association Bulletin 42:87-95

Singh, Ashok K., Singh, Anita, Engelhardt, Max (1997). The Lognormal Distribution in Environmental Applications. EPA/600/R-97/006.

Wehrens, R. & Van der Linden, W.E. (1997). Bootstrapping principal component regression models, *Journal of Chemometrics* 11, 157–171.

VITA

Graduate College

University of Nevada, Las Vegas

Skanda Pokkunuri

Home Address:
3401 Grove gate court
Apt #1628
Richmond VA 23233

Degrees:
Master of Science in Electrical Engineering
South Dakota School of Mines and Technology, 2003

Bachelors in Electrical Engineering, 1999
University of Madras, India

Thesis Title:
On Computing Bayesian Credible Sets for the Coefficient of Variation of a Normal Population

Thesis Examination Committee:
Chairperson, Dr. Ashok Kumar Singh, Ph.D.
Committee Member, Dr. Rohan Dalpatadu, Ph.D.
Committee Member, Dr. Xin Li, Ph.D.
Graduate Faculty Representative, Dr. Laxmi Gewali, Ph.D.