

1-1-2002

Confidence intervals for exposure risk assessment

Mark S Kirkham
University of Nevada, Las Vegas

Follow this and additional works at: <https://digitalscholarship.unlv.edu/rtds>

Repository Citation

Kirkham, Mark S, "Confidence intervals for exposure risk assessment" (2002). *UNLV Retrospective Theses & Dissertations*. 1431.

<http://dx.doi.org/10.25669/gzag-edmj>

This Thesis is protected by copyright and/or related rights. It has been brought to you by Digital Scholarship@UNLV with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Thesis has been accepted for inclusion in UNLV Retrospective Theses & Dissertations by an authorized administrator of Digital Scholarship@UNLV. For more information, please contact digitalscholarship@unlv.edu.

14

1 3 4 8 4

U M I
MICROFILMED 2003

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

CONFIDENCE INTERVALS
FOR EXPOSURE RISK
ASSESSMENT

by

Mark S Kirkham
Bachelor of Science
California State University, Fullerton
1992

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science Degree
Department of Mathematical Sciences
College of Sciences

Graduate College
University of Nevada, Las Vegas
December 2002

UMI Number: 1413484

Copyright 2003 by
Kirkham, Mark S.

All rights reserved.

UMI[®]

UMI Microform 1413484

Copyright 2003 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346



Thesis Approval
The Graduate College
University of Nevada, Las Vegas

June 28, _____, 2002

The Thesis prepared by

Mark S. Kirkham

Entitled

Confidence Intervals for Exposure Risk Assessment

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

Masters in Mathematical Sciences

Examination Committee Chair

Dean of the Graduate College

Examination Committee Member

Examination Committee Member

Graduate College Faculty Representative

ABSTRACT

Confidence Intervals for Exposure Risk Assessment by

Mark S Kirkham

Dr. Malwane Ananda, Examination Committee Chair
Professor of Mathematics
University of Nevada, Las Vegas

The problem of estimating exposure to a pollutant involves estimation of chemical intake Q . Typically, Q has the form $Q = \theta_{11}\theta_{12} \cdots \theta_{1r_1} / \theta_{211}\theta_{22} \cdots \theta_{2r_2}$ where all θ_{ij} 's are unknown means of certain random variables. In assessing the risk to human health from exposure to the pollutant of concern, a confidence interval estimate of Q is needed. In the case of independent normal random variables, point estimates and certain types of interval estimates of a product of several parameters exist in the literature, but not for a ratio of products. In many situations, lot of prior knowledge is available regarding the random variables corresponding to these parameters. In the case of independent normal random variables, we look at this problem from a Bayesian approach, and show how to incorporate prior knowledge to calculate confidence interval for Q . In situations with no prior knowledge, generalized Bayes approach with noninformative priors can be used. Several real and simulated examples are provided to demonstrate the proposed procedures.

TABLE OF CONTENTS

ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	v
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 NO PRIOR INFORMATION.....	3
CHAPTER 3 PRIOR INFORMATION.....	12
CHAPTER 4 SIGMA KNOWN OR LARGE SAMPLE SIZES.....	18
CHAPTER 5 EPA EXAMPLE AND CONCLUSION.....	24
APPENDIX SPLUS PROGRAMS	
METHOD 1.....	27
METHOD 2.....	29
METHOD 3.....	32
BIBLIOGRAPHY.....	36
VITA.....	37

ACKNOWLEDGEMENTS

I would like to thank Dr. Malwane Ananda for the many hours of help he has given to assisting me in completing this thesis. I would especially like to thank my wife Sherrie she has been a great source of encouragement and support.

CHAPTER 1

INTRODUCTION

Suppose X_{ij} , $i = 1, 2$; $j = 1, 2, \dots, r_i$ are independent normally distributed random variables with means θ_{ij} , $i = 1, 2$; $j = 1, 2, \dots, r_i$ and variances σ_{ij}^2 , $i = 1, 2$; $j = 1, 2, \dots, r_i$. We are interested in constructing confidence intervals for the statistic

$$Q = \frac{\theta_{11}\theta_{12} \cdots \theta_{1r_1}}{\theta_{11}\theta_{12} \cdots \theta_{1r_2}}. \quad (1)$$

These types of estimation problems arise in many environmental applications, such as exposure assessment and risk modeling, where measurements are assumed to have different independent normal populations with unknown means and variances (cf. United States Environmental Protection Agency document 540189-002, Vol. 1 Chap 6, page 21 - 54).

Suppose a random sample $x_{ij1}, x_{ij2}, \dots, x_{ijn_j}$ of size n_j is available from each population and let (\bar{x}_{ij}, s_{ij}) ; $i = 1, 2$; $j = 1, 2, \dots, r_i$ be their sample mean and sample standard deviation. Then $\{(\bar{x}_{ij}, s_{ij}); i = 1, 2$; $j = 1, 2, \dots, r_i\}$ makes a complete sufficient statistic for $\{\theta_{ij}, i = 1, 2$; $j = 1, 2, \dots, r_i\}$. Furthermore, they are all independent with probability distributions

$$\bar{X}_{ij} \sim N(\theta_{ij}, \sigma_{ij}^2/n_{ij}) \text{ and } (n_{ij} - 1)s_{ij}^2/\sigma_{ij}^2 \sim \chi_{n_{ij}-1}^2.$$

Then the maximum likelihood estimate and the method of moment estimate of Q is given by

$$\hat{Q} = \frac{\bar{x}_{11}\bar{x}_{12} \cdots \bar{x}_{1r_1}}{\bar{x}_{21}\bar{x}_{22} \cdots \bar{x}_{2r_2}}.$$

This is not an unbiased estimate of Q . In the case of the product of three normal means, i.e. $Q = \theta_{11}\theta_{12}\theta_{13}$, the estimate $\bar{x}_{11}\bar{x}_{12}\bar{x}_{13}$ is an unbiased estimate of $\theta_{11}\theta_{12}\theta_{13}$. Confidence interval estimates for the product of several normal means exist in the literature, but not for a ratio of products. Yfantis and Flatman (1991) and Ananda, Singh, and Flatman (1993) looked at the problem of the product of three normal means and provided confidence intervals. This paper is an extension of the work done in Ananda, Singh, and Flatman (1993) and shows that their procedures can be extended for the proposed problem.

In this paper, it will be shown how to incorporate prior knowledge and present data to construct confidence interval for Q . We consider several types of prior families and give their Bayesian estimators and Bayesian credible regions (Bayesian version of confidence intervals). The Bayesian tools are efficient whenever the prior knowledge about these unknown parameters are available, as is the case in many practical situations. Also, we consider generalized Bayesian estimators and confidence intervals for Q based on the noninformative prior (assuming no prior knowledge about the unknown parameters). Simulated examples are given to demonstrate the procedures and lastly a real example is given.

CHAPTER 2

NO PRIOR INFORMATION

First, let us consider the case with no prior information. As in Ananda, Singh, and Flatman (1993), let us consider independent improper noninformative prior

$$\pi(\theta_{ij}, \sigma_{ij}) = \frac{1}{\sigma_{ij}}, \quad 0 < \sigma_{ij} < \infty, -\infty < \theta_{ij} < \infty \quad (2)$$

on each set of parameters $(\theta_{ij}, \sigma_{ij}), i = 1, 2; j = 1, 2, \dots, r_i$. Then the joint noninformative prior would be their product. The marginal posterior distribution of θ_{ij} given (\bar{x}_{ij}, s_{ij}) is a t-distribution and is given by (Ananda, Singh, and Flatman (1993))

$$\theta_{ij} / (\bar{x}_{ij}, s_{ij}) \sim T(n_{ij} - 1, \bar{x}_{ij}, s_{ij}^2 / n_{ij}). \quad (3)$$

Here $P \sim T(\alpha, \mu, \sigma^2)$ denotes that the variable P has a t-distribution with degrees of freedom α , location parameter μ and scale parameter σ , i.e. $(P - \mu) / \sigma$ has a student t-distribution with degrees of freedom α .

The posterior distribution of $Q = (\theta_{11}\theta_{12} \cdots \theta_{1r_1}) / (\theta_{21}\theta_{22} \cdots \theta_{2r_2})$ given the data is the products and ratio of the above t-distributions. The posterior distribution of $Q = (\theta_{11}\theta_{12} \cdots \theta_{1r_1}) / (\theta_{21}\theta_{22} \cdots \theta_{2r_2})$ given the data can be numerically tabulated by generating a large number (say N) of data points from this distribution. This can be done by randomly generating N data points from each of the t-distributions given in (3) and by

calculating $Q = (\theta_{11}\theta_{12} \cdots \theta_{1r_1})/(\theta_{21}\theta_{22} \cdots \theta_{2r_2})$. These N values, say Q_1, Q_2, \dots, Q_N are coming from the posterior distribution of $Q = (\theta_{11}\theta_{12} \cdots \theta_{1r_1})/(\theta_{21}\theta_{22} \cdots \theta_{2r_2})$. Then the Bayes estimate of Q is the mean of this posterior distribution and it can be estimated by

$$\sum_{i=1}^N Q_i / N.$$

The $100(1-\alpha)\%$ confidence interval (credible region) for Q is (Q_L, Q_U) where $\Pr(Q_L < Q < Q_U) = 1-\alpha$. The lower and upper confidence limits can be estimated using simulation. Then the $100\alpha/2$ percentile of the above simulated sample is an estimate for the lower confidence limit and $100(1-\alpha/2)$ percentile is an estimate for the upper confidence limit. Notice here that N must be very large in order to get an accurate estimate by simulation. Furthermore notice that even though the posterior density of θ_{ij} given (\bar{x}_{ij}, s_{ij}) is a symmetric density, the posterior distribution of Q is no longer a symmetric density and therefore the credible region calculated this way may not produce the shortest possible credible region.

In order to create a credible set, an S-plus program will generate 20,000 observations from a $t(\alpha)$ distribution for each θ_{ij} . The 20,000 $t(\alpha)$ observations will then be converted to $t(\alpha, \mu, \sigma^2)$.

To test this, a program was written in S-Plus (See Appendix) which randomly generates 6 $t(n_{ij} - 1, \bar{x}_{ij}, s_{ij}^2/n_{ij})$ distributions of 20,000 observations. In this paper all examples will be of the form $Q = \frac{\theta_{11}\theta_{12}\theta_{13}\theta_{14}}{\theta_{21}\theta_{22}}$. The methods however can easily be extended to many different forms of Q .

The following three examples are given. In each case a histograms of Q and θ_{11} for the 20,000 simulated observations are given with lines showing the 5th and 95th percentiles, the median, and the Actual Value. Note the histograms show that the distribution of Q is not a symmetric distribution while the distributions of the θ_{11} are symmetric t-distributions.

Example 1: In this first example we have the following parameters:

Sample Sizes: $n_{11} = n_{12} = \dots = n_{22} = 40$

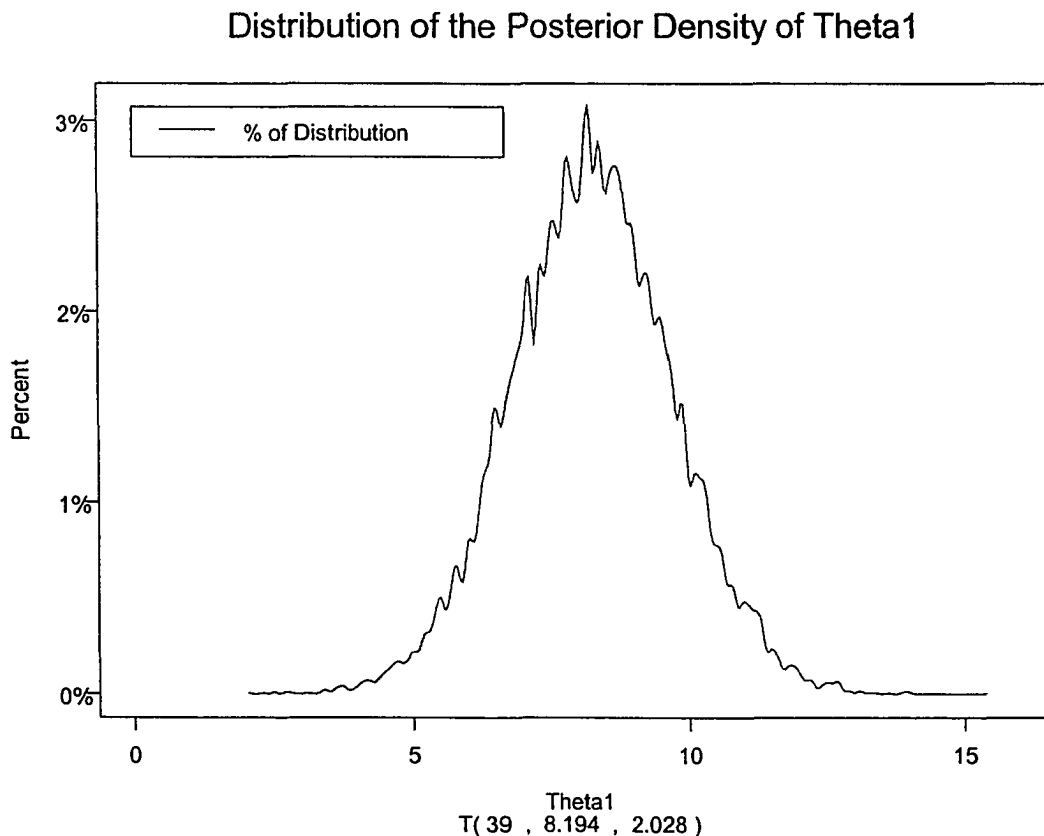
Actual Population values: $\theta_{11} = 10.0, \theta_{12} = 0.4, \theta_{13} = 6.5, \theta_{14} = 8.0, \theta_{21} = 2.5, \theta_{22} = 3.2$

Standard Deviations: $\sigma_{11} = 10.0, \sigma_{12} = 5.0, \sigma_{13} = 3.0, \sigma_{14} = 3.0, \sigma_{21} = 2.0, \sigma_{22} = 2.0$

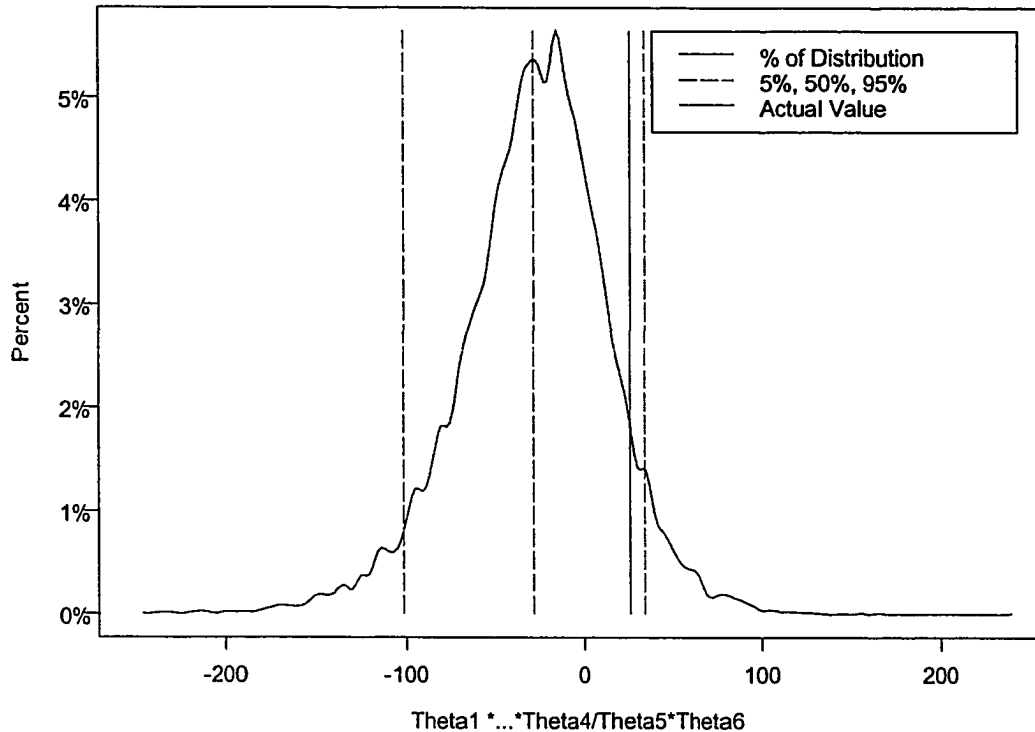
Generated Means: $\bar{x}_{11} = 8.194, \bar{x}_{12} = -0.562, \bar{x}_{13} = 5.627, \bar{x}_{14} = 7.459, \bar{x}_{21} = 1.996, \bar{x}_{22} = 3.347$

Generated Standard Deviations:

$s_{11} = 9.006, s_{12} = 4.595, s_{13} = 3.233, s_{14} = 3.029, s_{21} = 1.562, s_{22} = 2.075$



Distribution of Posterior Density



The Actual Value of $Q = 26.0$

A 90% Credible Set for Q is given by $(-101.568, 33.977)$ with the median being -27.707 .

The first graph above shows the generated distribution for θ_{11} , which is a Student's T distribution with 39 degrees of freedom, a location parameter of 8.194, and scale

parameter of 2.028. The variance of the T distribution is $\frac{s_{11}^2}{n_{11}} = \frac{9.006^2}{40} = 2.028$. The

second graph shows the distribution of Q . To create the graph a histogram with 100 buckets is calculated. Then by drawing a simple spline through the midpoints of the histogram buckets the distribution above is created. The actual value of 26 is shown as the solid vertical line the 90% credible set and the median are shown as the dashed lines.

Example 2: In the second example we have the following parameters:

Sample Sizes: $n_{11} = 20, n_{12} = 20, n_{13} = 30, n_{14} = 30, n_{21} = 40, n_{22} = 40$

Actual Population values: $\theta_{11} = 5.0, \theta_{12} = 1.0, \theta_{13} = 4.0, \theta_{14} = 3.2, \theta_{21} = 5.0, \theta_{22} = 7.0$

Standard Deviations: $\sigma_{11} = 4.0, \sigma_{12} = 4.0, \sigma_{13} = 3.0, \sigma_{14} = 3.0, \sigma_{21} = 2.0, \sigma_{22} = 2.0$

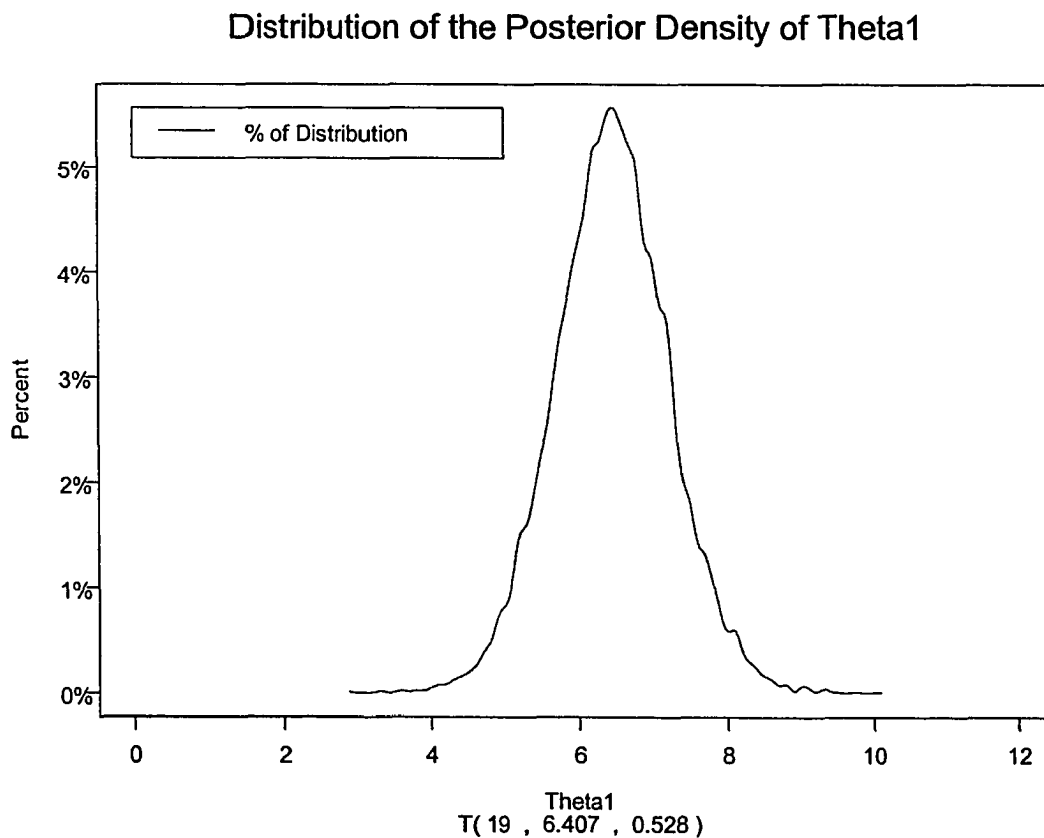
Generated Means: $\bar{x}_{11} = 6.407, \bar{x}_{12} = -0.175, \bar{x}_{13} = 4.658, \bar{x}_{14} = 2.965, \bar{x}_{21} = 5.033, \bar{x}_{22} = 6.239$

Generated Standard Deviations:

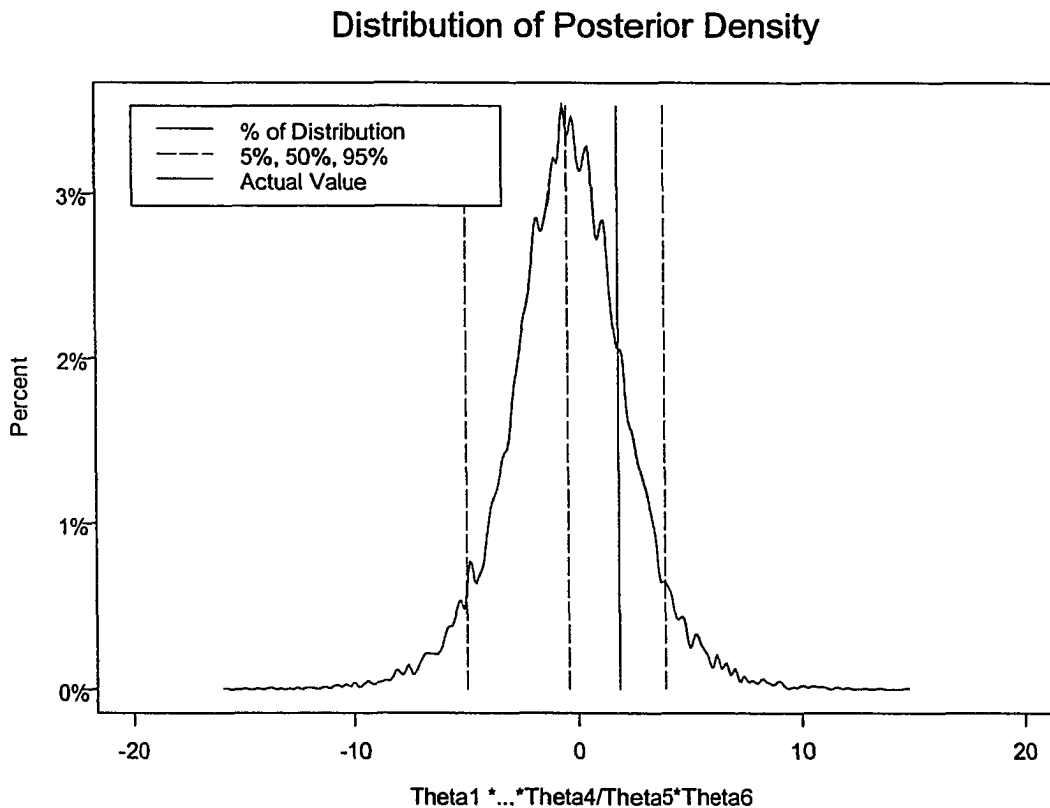
$s_{11} = 3.249, s_{12} = 3.886, s_{13} = 3.069, s_{14} = 3.277, s_{21} = 1.759, s_{22} = 2.162$

The Actual Value of $Q = 1.829$

A 90% Credible Set is given by $(-4.949, 3.866)$ with the median being -0.449 .



Again the graph of the generated distribution for θ_{11} is shown above. It is a T distribution with 19 degrees of freedom, location parameter of 3.957 and scale parameter of 0.528. The variance of this T distribution is $\frac{s_{11}^2}{n_{11}} = \frac{3.249^2}{20} = 0.528$. The following graph shows the distribution of Q for this example. This time the actual value is 1.829 is shown as the solid vertical line the 90% credible set and the median are again shown as the dashed lines.



Example 3: In this example smaller sample sizes will be used the parameters are as follows:

Sample Sizes: $n_{11} = 10, n_{12} = 10, n_{13} = 15, n_{14} = 15, n_{21} = 20, n_{22} = 20$

Actual Population values: $\theta_{11} = 2.0, \theta_{12} = 2.0, \theta_{13} = 3.0, \theta_{14} = 3.0, \theta_{21} = 4.0, \theta_{22} = 4.0$

Standard Deviations: $\sigma_{11} = 1.5, \sigma_{12} = 2.0, \sigma_{13} = 2.0, \sigma_{14} = 3.0, \sigma_{21} = 2.5, \sigma_{22} = 3.5$

Generated Means: $\bar{x}_{11} = 2.444, \bar{x}_{12} = 1.359, \bar{x}_{13} = 3.541, \bar{x}_{14} = 2.017, \bar{x}_{21} = 4.227, \bar{x}_{22} = 4.469$

Generated Standard Deviations:

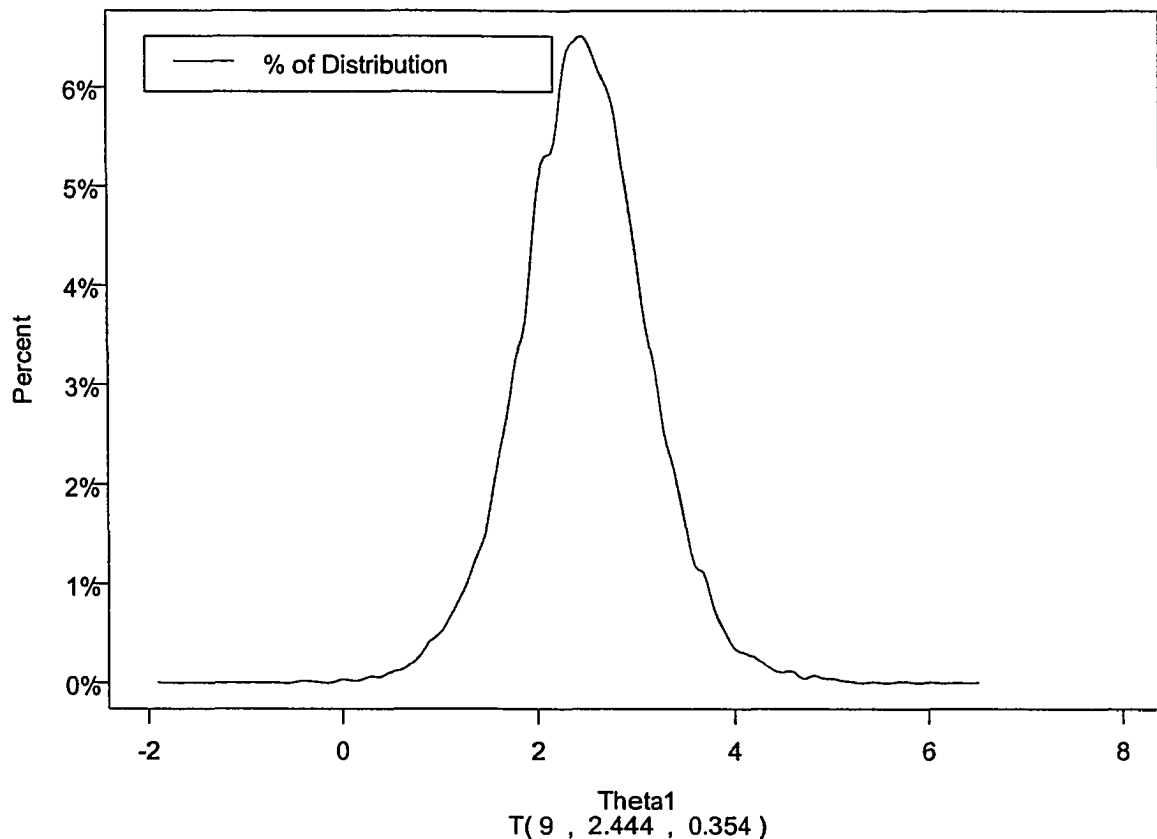
$s_{11} = 1.881, s_{12} = 1.64, s_{13} = 2.127, s_{14} = 2.958, s_{21} = 3.474, s_{22} = 3.143$

The Actual Value of $Q = 2.25$

A 90% Credible Set is given by (0.141, 3.375) with the median being 1.095.

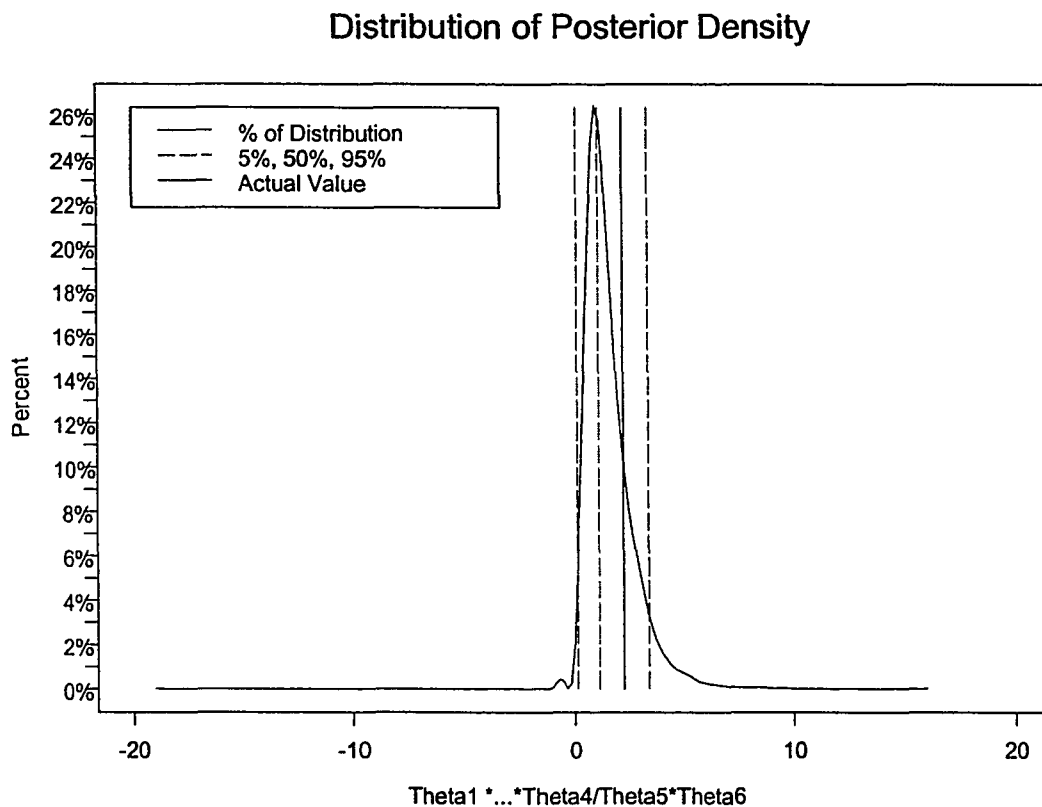
As in the previous examples the graph of θ_{11} is shown below. It is a T distribution this time with 9 degrees of freedom, location parameter of 2.444 and scale parameter of 0.354. The scale parameter of this T distribution calculated as in the previous examples

Distribution of the Posterior Density of Theta1



as $\frac{s_{11}^2}{n_{11}} = \frac{1.881^2}{10} = 0.321$. Next the graph of the Q distribution is shown. For this example

the actual value of 2.25 is shown as the solid vertical line as before the 90% credible set and the median are shown as the dashed lines.



This example shows the extreme skewness that sometimes arises in these problems. Many examples were run in preparing this thesis and often Q had extreme outliers. This occurs when θ_{21} or θ_{22} , the parameters in the denominator of Q are close to zero. When the generated values are selected some of them are going to be very close to zero. Since θ_{21} and θ_{22} are divisors they cause the value of Q to be very large in absolute value.

Many of the examples run were much more skewed than this example. However the credible sets still captured the actual value as expected.

CHAPTER 3

PRIOR INFORMATION

Now we will assume we have prior information about θ_{ij} and σ_{ij}^2 . A joint prior density for θ_{ij} and σ_{ij}^2 from a conjugate family will be chosen. If $\pi_{ij}(\theta_{ij} / \sigma_{ij}^2)$ is a normal $N(\mu_{ij}, \tau_{ij}\sigma_{ij}^2)$ density and $\pi_{ij}(\sigma_{ij}^2)$ is an inverted gamma $IG(\alpha_{ij}, \beta_{ij})$ density, then $\pi_{ij}(\theta_{ij}, \sigma_{ij}^2) = \pi_{ij}(\theta_{ij} / \sigma_{ij}^2)\pi_{ij}(\sigma_{ij}^2)$ is a joint prior density is from a conjugate family of the distribution of X_{ij} . We will assume the values of $\mu_{ij}, \tau_{ij}, \alpha_{ij}$, and β_{ij} are known.

Berger (1985) gives the result that under the above assumptions the marginal posterior distribution for θ_{ij} given (\bar{x}_{ij}, s_{ij}) is given by:

$$\pi(\theta_{ij} | \bar{x}_{ij}, s_{ij}) \sim T(2\alpha_{ij} + n_{ij} - 1, u(x)_{ij}, [(\tau_{ij}^{-1} + n_{ij})(\alpha_{ij} + (n_{ij} - 1)/2)\beta'_{ij}]^{-1}) \quad (5)$$

where

$$u(x)_{ij} = \frac{(\mu_{ij} + n_{ij}\tau_{ij}\bar{x}_{ij})}{(n_{ij}\tau_{ij} + 1)}$$

and

$$\beta'_{ij} = \left[\beta_{ij}^{-1} + \frac{(n_{ij} - 1)s_{ij}^2}{2} + \frac{n_{ij}(\bar{x}_{ij} - \mu_{ij})^2}{(2 + 2n_{ij}\tau_{ij})} \right]^{-1}$$

The product of four of the above t-distributions divided by the product of two of the above t-distributions will give the joint distribution. The Bayes estimate of

$$Q = \frac{\theta_{11}\theta_{12}\theta_{13}\theta_{14}}{\theta_{21}\theta_{22}} \text{ is given by}$$

$$\hat{Q} = \frac{\prod_{j=1}^4 \frac{(\mu_{1j} + n_{1j}\tau_{1j}\bar{x}_{1j})}{(n_{1j}\tau_{1j} + 1)}}{\prod_{j=1}^2 \frac{(\mu_{2j} + n_{2j}\tau_{2j}\bar{x}_{2j})}{(n_{2j}\tau_{2j} + 1)}}$$

As in method 1 the posterior distribution of Q will be simulated. 20,000 observations from a $t(2\alpha_{ij} + n_{ij} - 1)$ will be generated for each θ_{ij} . Then, as in method 1, these will be multiplied times the variance and added to the mean to give a t-distribution as in equation (5). Program 2 in the appendix of this thesis shows the program.

Example 4 The following parameters are used.

Sample Sizes: $n_{11} = 8, n_{12} = 10, n_{13} = 12, n_{14} = 14, n_{21} = 8, n_{22} = 10$

$$\alpha_{11} = 1.5, \alpha_{12} = 1.2, \alpha_{13} = 2.0, \alpha_{14} = 1.7, \alpha_{21} = 2.0, \alpha_{22} = 2.2$$

$$\beta_{11} = 0.7, \beta_{12} = 0.8, \beta_{13} = 1.0, \beta_{14} = 0.6, \beta_{21} = 0.5, \beta_{22} = 0.9$$

The α_{ij} and β_{ij} were used as parameters for the Inverse Gamma distribution to generate the following sigma values:

$$\sigma_{11} = 0.7715, \sigma_{12} = 1.0241, \sigma_{13} = 0.9007, \sigma_{14} = 1.6299, \sigma_{21} = 2.0724, \sigma_{22} = 1.205$$

The sigma are squared and multiplied times the known τ_{ij} values to use as the variances and these along with the known means μ_{ij} are used as the normal parameters to generate

θ_{ij}

$$\tau_{11} = 0.25, \tau_{12} = 0.15, \tau_{13} = 0.25, \tau_{14} = 0.75, \tau_{21} = 1.00, \tau_{22} = 1.50$$

$$\mu_{11} = 5, \mu_{12} = 10, \mu_{13} = 5, \mu_{14} = 15, \mu_{21} = 20, \mu_{22} = 30$$

$$\theta_{11} = 3.8336, \theta_{12} = 10.1342, \theta_{13} = 4.7435, \theta_{14} = 13.3002, \theta_{21} = 21.4444, \theta_{22} = 30.0226$$

Generated Means:

$$\bar{x}_{11} = 4.7259, \bar{x}_{12} = 10.5563, \bar{x}_{13} = 4.5581, \bar{x}_{14} = 14.0251, \bar{x}_{21} = 21.0643, \bar{x}_{22} = 30.0433$$

Generated Standard Deviations:

$$s_{11} = 0.799, s_{12} = 1.3304, s_{13} = 0.7707, s_{14} = 1.7577, s_{21} = 1.5889, s_{22} = 0.926$$

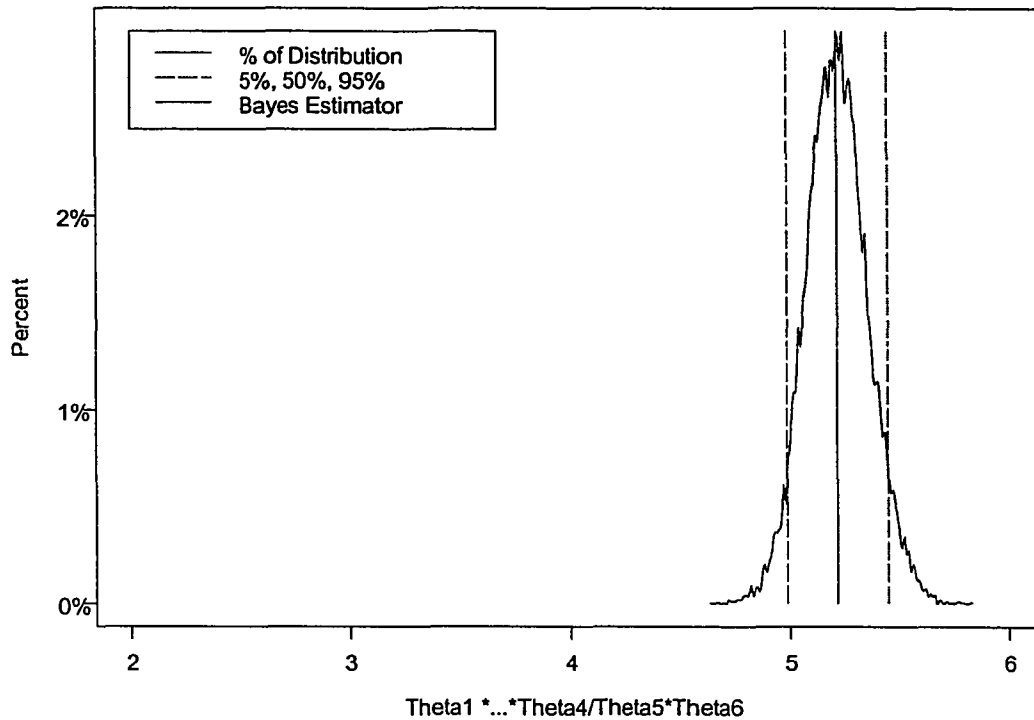
Using the above generated values along with the parameter values a distribution of 20,000 observations from the t-distribution show in equation 5 was obtained. We get the following results.

The Actual Value of Q is 5.038.

The Bayesian Estimator is 5.2107.

A 90% Credible Set is given by (4.9859, 5.4487) with the median being 5.2113. For comparison purposes the 90% Credible Set generated by method 1 is given by (4.843, 7.162) with a median of 5.933. Notice the range of the 90 % Credible Set from method 2 is shorter than the set given by method 1 and the median is closer to the actual value in method 2 than in method 1. The graph below is of Q , the posterior distribution generated by method 2 in example 4. The method 2 graphs show the bayesian estimator as a solid line and the 90% credible region and median as dashed lines. Usually the bayesian estimator of Q is so close to the median that the dashed line showing the median is covered by the solid line for the actual value.

Distribution of Posterior Density



Example 5 In this example all the parameters are the same as in example 4 with the exception of n_{ij} .

Sample Sizes: $n_{11} = n_{12} = \dots = n_{22} = 40$

Known α_{ij} and β_{ij} values:

$$\alpha_{11} = 1.5, \alpha_{12} = 1.2, \alpha_{13} = 2.0, \alpha_{14} = 1.7, \alpha_{21} = 2.0, \alpha_{22} = 2.2$$

$$\beta_{11} = 0.7, \beta_{12} = 0.8, \beta_{13} = 1.0, \beta_{14} = 0.6, \beta_{21} = 0.5, \beta_{22} = 0.9$$

Generated sigma values:

$$\sigma_{11} = 2.2807, \sigma_{12} = 5.7185, \sigma_{13} = 1.3594, \sigma_{14} = 0.4869, \sigma_{21} = 1.7887, \sigma_{22} = 0.4886$$

The known τ_{ij} and μ_{ij} values are:

$$\tau_{11} = 0.25, \tau_{12} = 0.15, \tau_{13} = 0.25, \tau_{14} = 0.75, \tau_{21} = 1.00, \tau_{22} = 1.50$$

$$\mu_{11} = 5, \mu_{12} = 10, \mu_{13} = 5, \mu_{14} = 15, \mu_{21} = 20, \mu_{22} = 30$$

Generated Theta values:

$$\theta_{11} = 5.1004, \theta_{12} = 10.2802, \theta_{13} = 5.388, \theta_{14} = 14.7992, \theta_{21} = 22.7142, \theta_{22} = 29.4074$$

Generated Means:

$$\bar{x}_{11} = 5.3876, \bar{x}_{12} = 11.1318, \bar{x}_{13} = 5.8749, \bar{x}_{14} = 14.8504, \bar{x}_{21} = 22.122, \bar{x}_{22} = 29.5309$$

Generated Standard Deviations:

$$s_{11} = 1.8644, s_{12} = 5.8536, s_{13} = 1.5137, s_{14} = 0.4407, s_{21} = 1.9172, s_{22} = 0.4494$$

From the above parameters and generated values 20,000 observations from the t-distribution shown in equation 5 were generated.

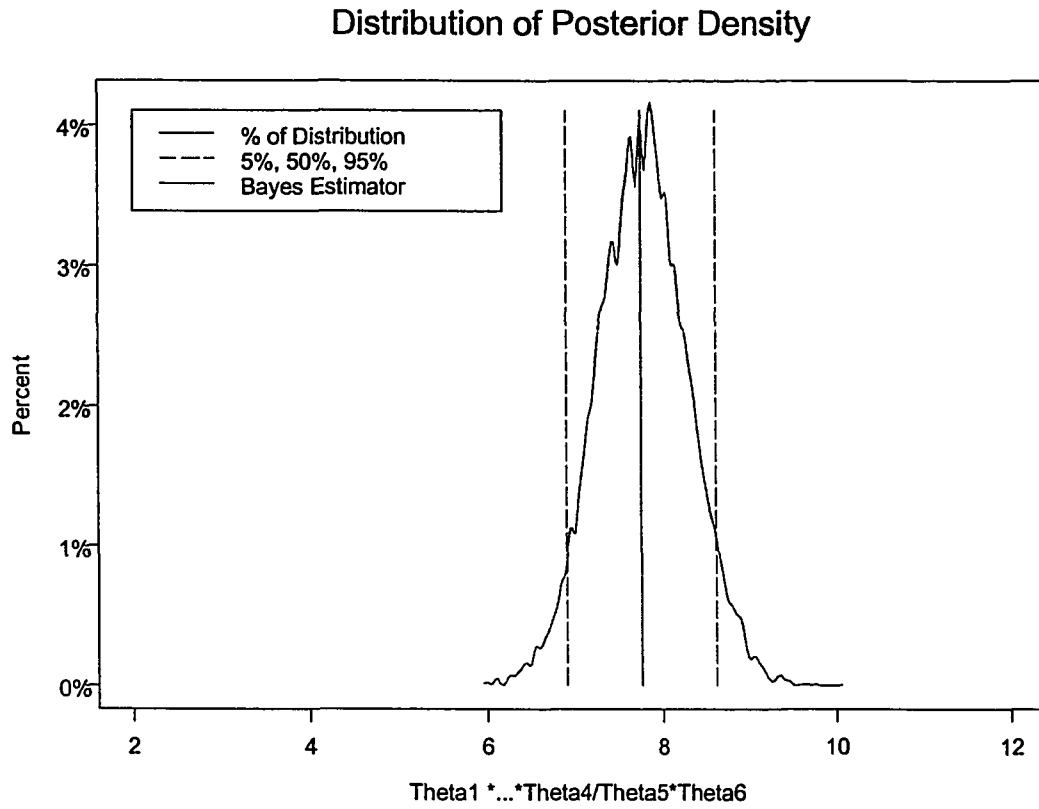
The Actual Value of $Q = 8.009$

The Bayesian Estimator is 7.7538

A 90% Credible Set is given by (6.9103, 8.6121) with the median being 7.7567.

For comparison purposes the 90% Credible Set generated by method 1 is given by (7.157, 10.178) with a median of 8.611. Again the range of the 90 % Credible Set from method 2 is shorter than the set given by method 1 and the median is closer to the actual value in method 2 than in method 1. The graph below is of Q , the posterior distribution generated by method 2 in example 5. In this graph we can see with the larger n_{ij} the smaller the variance of the posterior distribution.

In the previous two examples we see that when prior information is available, which is often the case, we get much better estimates of the actual value and the range of the credible set of the posterior distribution is smaller.



CHAPTER 4

SIGMA KNOWN OR LARGE SAMPLE SIZES

In this method we will look at the case when the population variances $\sigma_{11}^2, \sigma_{12}^2, \dots, \sigma_{22}^2$ are either known or the sample sizes are large enough that the sample standard deviations s_{ij} are good approximations for σ_{ij} . In addition we will assume a normal prior distribution $N(\mu_{ij}, \tau_{ij}^2)$ for each of the θ_{ij} . For this example we will assume the values of μ_{ij} and τ_{ij} are known.

Under these assumptions it can be shown that the marginal posterior distribution of θ_{ij} given (\bar{x}_{ij}, s_{ij}) is

$$\pi(\theta_{ij} | \bar{x}_{ij}, s_{ij}) \sim N \left(\frac{\frac{n_{ij}\bar{x}_{ij}}{\sigma_{ij}^2} + \frac{\mu_{ij}}{\tau_{ij}^2}}{\frac{n_{ij}}{\sigma_{ij}^2} + \frac{1}{\tau_{ij}^2}}, \frac{1}{\frac{n_{ij}}{\sigma_{ij}^2} + \frac{1}{\tau_{ij}^2}} \right) \quad (6)$$

Using this information the Bayes estimate of Q is given by

$$\hat{Q} = \prod_{j=1}^4 \frac{n_{1j}\bar{x}_{1j}/\sigma_{1j}^2 + \mu_{1j}/\tau_{1j}^2}{n_{1j}/\sigma_{1j}^2 + 1/\tau_{1j}^2} \bigg/ \prod_{j=1}^2 \frac{n_{2j}\bar{x}_{2j}/\sigma_{2j}^2 + \mu_{2j}/\tau_{2j}^2}{n_{2j}/\sigma_{2j}^2 + 1/\tau_{2j}^2} \quad (7)$$

Equation 6 is given by Berger (1985), equation 7 follows by the independence of the x_{ij} and from the independence of the different priors.

A program that simulates 20,000 observations is used to demonstrate method 3.

The program is shown in the appendix to this thesis. The program creates, just like the

programs for method 1 and method 2, a 90% credible set along with the median of the posterior distribution. Three more examples are shown.

Example 6 In this example we will use small n_{ij} but assume σ_{ij} is known. The example is run at the same time using s_{ij} so the results can be compared. The following are the parameters used in this example.

$$n_{11} = 14, n_{12} = 16, n_{13} = 18, n_{14} = 20, n_{21} = 12, n_{22} = 14$$

$$\mu_{11} = 10.0, \mu_{12} = 20.0, \mu_{13} = 5.0, \mu_{14} = 15.0, \mu_{21} = 5.0, \mu_{22} = 10.0$$

$$\tau_{11} = 1.0, \tau_{12} = 2.0, \tau_{13} = 1.5, \tau_{14} = 3.0, \tau_{21} = 1.0, \tau_{22} = 1.5$$

From the above known parameter values the theta values below are generated.

$$\theta_{11} = 8.9766, \theta_{12} = 20.3466, \theta_{13} = 6.8982, \theta_{14} = 15.2506, \theta_{21} = 4.7625, \theta_{22} = 13.7896$$

The known sigma values are:

$$\sigma_{11} = 3.0, \sigma_{12} = 5.0, \sigma_{13} = 2.0, \sigma_{14} = 3.0, \sigma_{21} = 2.0, \sigma_{22} = 3.0$$

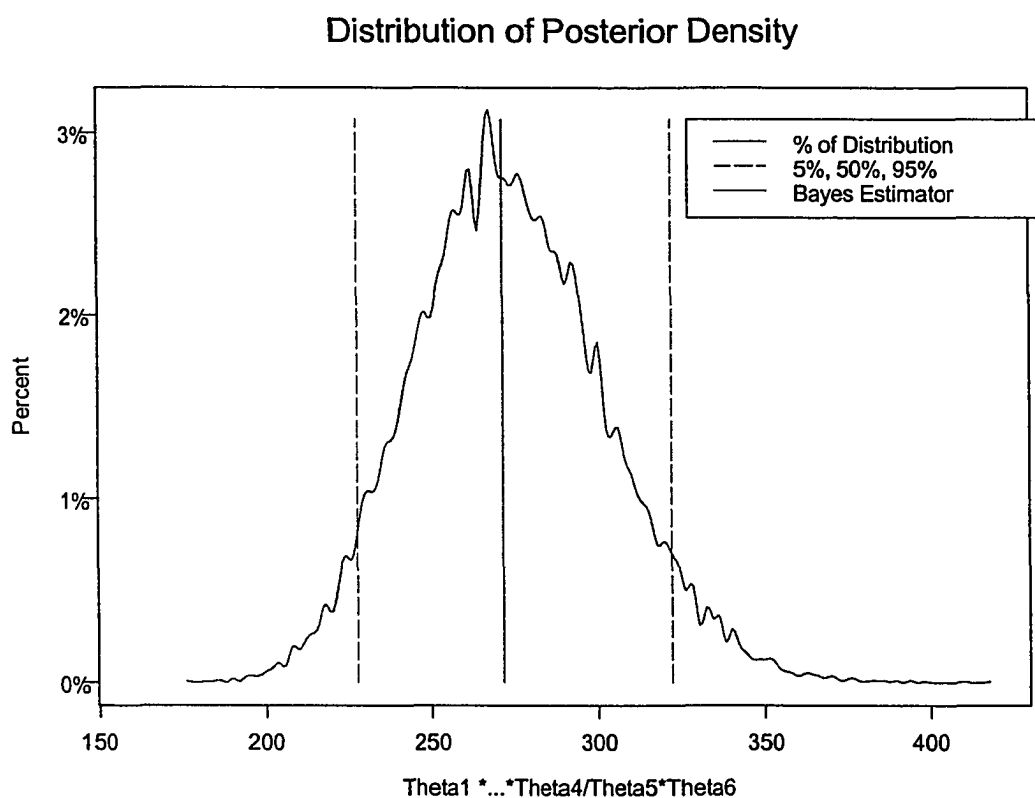
The generated values are:

$$\bar{x}_{11} = 8.3187, \bar{x}_{12} = 18.0886, \bar{x}_{13} = 6.7883, \bar{x}_{14} = 16.0471, \bar{x}_{21} = 5.4392, \bar{x}_{22} = 12.9106$$

$$s_{11} = 1.6528, s_{12} = 4.3777, s_{13} = 2.095, s_{14} = 3.6775, s_{21} = 1.6886, s_{22} = 3.5552$$

Using method 3 the 90% credible set for Q using the known values for sigma is (227.4018, 321.9837) with a median value of 271.2864. Using this method the Bayes Estimator is 271.2141. Using the generated s_{ij} the 90% credible set for Q is (221.0221, 305.7398) with a median value of 259.7052. Using the generated s_{ij} the Bayes Estimator is 260.0945. The credible set using the generated s_{ij} is smaller than the one created using the known sigma values, however, when we compare the two Bayes Estimators to the actual value of 292.576 we see that the estimator using the known value is much closer.

For comparison purposes the same example was run using in method 1 (no prior information). Method 1 generated a 90% credible set of (172.112, 415.803) with a median of 260.618. Notice that the range of the credible set for method 3 is smaller than the credible set of method 1. This is the expected result since we knew more information about the parameters in method 3. Below is the graph of this example using the known values for the standard deviations.



Example 7: We will again assume σ_{ij} are known. The only change from example 6 is the n_{ij} .

In this example we will use larger sample sizes:

$$n_{11} = 30, n_{12} = 25, n_{13} = 35, n_{14} = 40, n_{21} = 45, n_{22} = 30$$

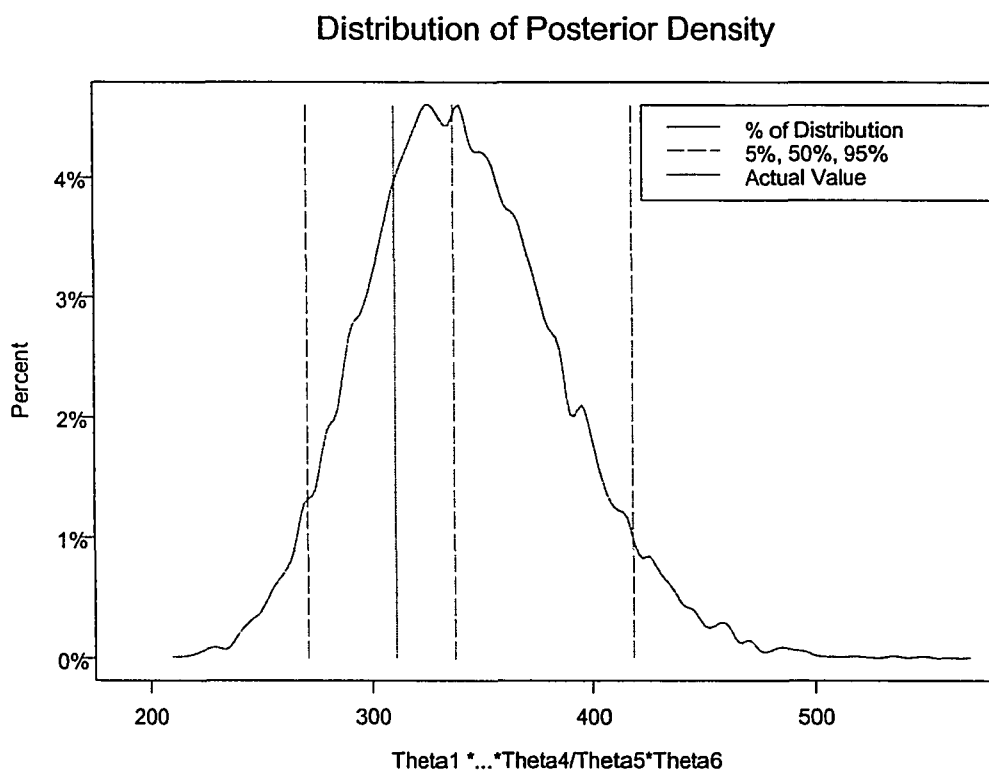
The generated values are:

$$\bar{x}_{11} = 10.5115, \bar{x}_{12} = 17.4488, \bar{x}_{13} = 8.1794, \bar{x}_{14} = 14.6028, \bar{x}_{21} = 6.0552, \bar{x}_{22} = 12.3033$$

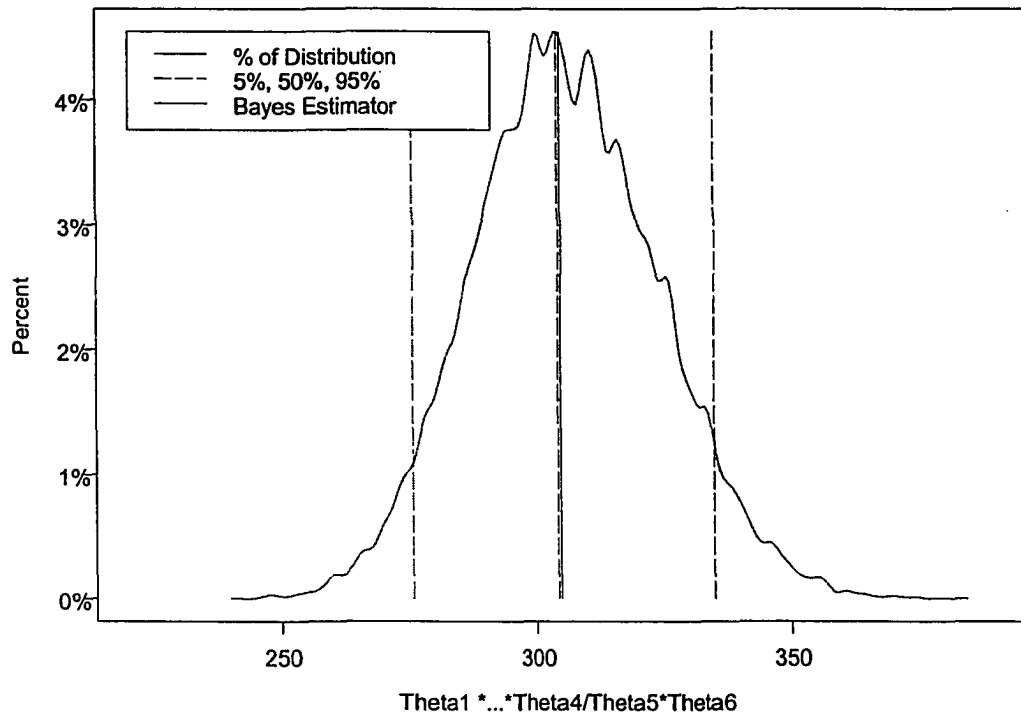
$$\theta_{11} = 10.6335, \theta_{12} = 16.2606, \theta_{13} = 8.5624, \theta_{14} = 14.3554, \theta_{21} = 6.2248, \theta_{22} = 10.9837$$

Using method 3 the 90% credible set for Q is (275.8139, 8787, 335.142) with a median value of 304.3312. The Bayes Estimator is 304.7853 and the actual value is 310.85.

Using in method 1 (no prior information) generated a 90% credible set of (271.187, 417.394) with a median of 337.028. As expected the range of the credible set for method 3 is smaller than the credible set of method 1. The first graph that follows shows the results of method 1 the second is the distribution of Q after running method 3.



Distribution of Posterior Density



Example 8: Here we assume σ_{ij} are unknown. The generated s_{ij} will be used instead of σ_{ij} . Otherwise example 8 will use all the same parameters as example 7.

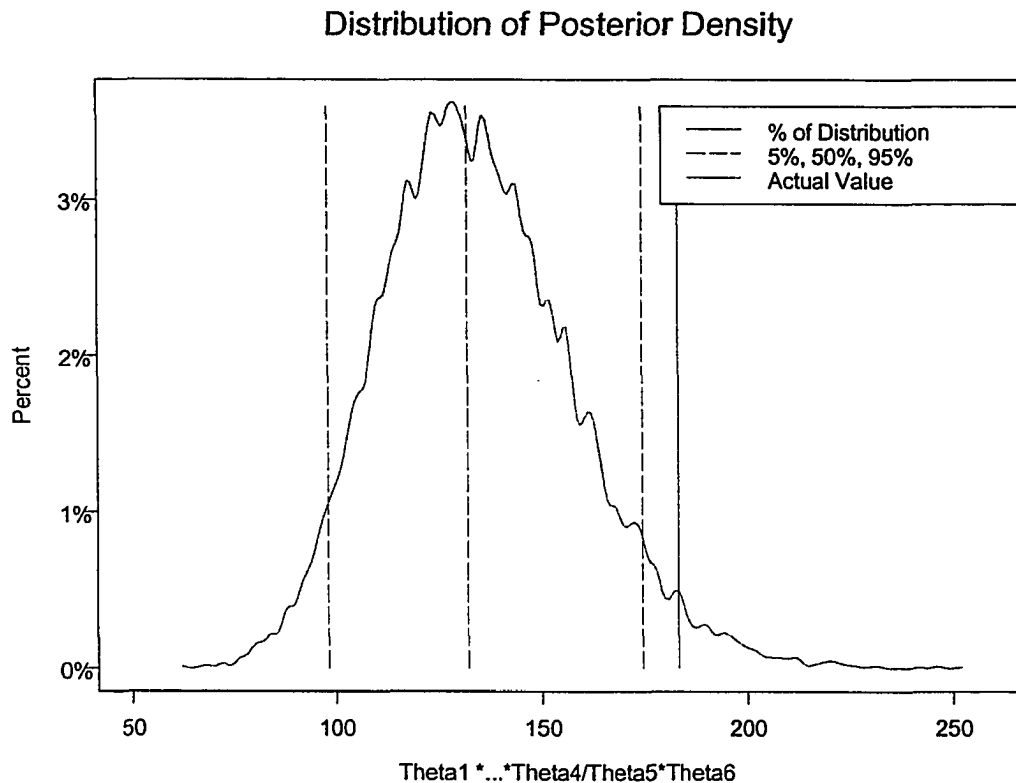
The generated values are:

$$\bar{x}_{11} = 10.9517, \bar{x}_{12} = 16.4027, \bar{x}_{13} = 4.3852, \bar{x}_{14} = 12.8007, \bar{x}_{21} = 4.5354, \bar{x}_{22} = 12.515$$

$$s_{11} = 3.7288, s_{12} = 5.1602, s_{13} = 2.2102, s_{14} = 2.9437, s_{21} = 2.0519, s_{22} = 3.2594$$

$$\theta_{11} = 11.198, \theta_{12} = 16.4875, \theta_{13} = 4.3948, \theta_{14} = 12.6949, \theta_{21} = 4.5907, \theta_{22} = 12.2483$$

Using method 3 the 90% credible set for Q is (164.7927, 210.1946) with a median value of 186.1213. The Bayes Estimator is 186.4547 with an actual value of 183.193. Using in method 1 (no prior information) generated a 90% credible set of (98.155, 174.399) with a median of 132.077. As expected the range of the credible set for method 3 is smaller than the credible set of method 1 and the median value from method 3 is closer to the



actual value than the median from method 1. Again first is shown the graph from from method 1 then the graph using method 3.

CHAPTER 5

EPA EXAMPLE AND CONCLUSION

In the EPA Exposure Assessment document [1] describes many ways to assess exposure to harmful chemicals. One of these is the ingestion of chemicals in surface water while swimming. The equation for assessing the chemical intake is given by

$$\text{Intake}(mg / kg - \text{day}) = \frac{CW \times CR \times ET \times EF \times ED}{BW \times AT}$$

Where:

CW = Chemical Concentration in Water (mg/liter)

CR = Contact Rate (liters/hour)

ET = Exposure Time (hours/event)

EF = Exposure Frequency (events/year)

ED = Exposure Duration (years)

BW = Body Weight (kg)

AT = Average Time (period over which exposure is averaged – days)

Means for some of these values are given in the EPA document. However the mean for chemical concentration and body weight variables are arbitrary values. The standard deviations are also arbitrary since they were not provided in the EPA documentation. The following table shows the values for this example.

Variable	Mean	Standard Deviation
CW	3	0.2
CR	50	5
ET	1.5	5
EF	7	1.5
ED	70	5
BW	70	10
AT	ED*365	ED*365

First lets look at Method 1 with sample sizes equal to 40 as in example 1. For this example there are five variables in the numerator of Q and two in the denominator. The programs where easily adapted to this problem.

Generated Means:

$$\bar{x}_{11} = 2.924, \bar{x}_{12} = 50.106, \bar{x}_{13} = 1.507, \bar{x}_{14} = 7.028, \bar{x}_{15} = 70.823, \bar{x}_{21} = 70.162, \bar{x}_{22} = 25,606.915$$

Generated Standard Deviations:

$$s_{11} = 0.211, s_{12} = 4.134, s_{13} = 0.439, s_{14} = 1.393, s_{15} = 4.526, s_{21} = 4.744, s_{22} = 1762.222$$

A 90% Credible Set is given by (0.055, 0.067) with the median being 0.061. The actual value is 0.062

Using Method 3 where we assume θ_{ij} have normal priors and the σ_{ij} are either known or the sample sizes are large enough to use the sample standard deviations. In this case we assume the σ_{ij} are known. The μ_{ij} and σ_{ij} are from the table above in addition the parameters below were used.

$$n_{11} = 15, n_{12} = 16, n_{13} = 17, n_{14} = 13, n_{15} = 16, n_{21} = 14, n_{22} = 12$$

$$\tau_{11} = 1.0, \tau_{12} = 3.0, \tau_{13} = 0.2, \tau_{14} = 0.5, \tau_{15} = 3.0, \tau_{21} = 5.0, \tau_{22} = 1095.0$$

Generating 20,000 sample observations gave generated values of

$$\bar{x}_{11} = 3.4545, \bar{x}_{12} = 54.0174, \bar{x}_{13} = 1.5671, \bar{x}_{14} = 7.0434, \bar{x}_{15} = 67.4677, \bar{x}_{21} = 68.3389, \bar{x}_{22} = 26089.2348$$

$$s_{11} = 0.2, s_{12} = 5.0, s_{13} = 0.5, s_{14} = 1.5, s_{15} = 5.0, s_{21} = 5.0, s_{22} = 1825$$

$$\theta_{11} = 3.3944, \theta_{12} = 52.7117, \theta_{13} = 1.4766, \theta_{14} = 7.586, \theta_{15} = 67.5756, \theta_{21} = 68.7514, \theta_{22} = 26012.822$$

Which results in a 90% confidence interval of (0.0738, 0.0786).

CONCLUSION

This thesis discussed three methods of deriving confidence intervals for exposure risk assessment. The situation where these methods will be useful would be any one of the form $Q = \frac{\theta_{11}\theta_{12}\cdots\theta_{1r_1}}{\theta_{11}\theta_{12}\cdots\theta_{1r_2}}$. The types of situations where these methods would arise would be in risk analysis and in environmental exposure analysis. Specifically, the EPA [1] documents many situations where these methods may be used yet in the EPA document no good way of finding Q is put forth. The Bayesian methods shown in this thesis will give better results especially when prior knowledge of the parameters is available.

APPENDIX

Splus program for Method 1

```

function(theta, sigma, num, iter, q)
{
#####
#
# theta is the vector of location parameters      #
# sigma is the vector of shape parameters        #
# num is the vector of sample sizes              #
# x is the vector of sample means                #
# s is the vector of sample standard deviations  #
# rg is the vector of iter randomly generated   #
# data points                                    #
# iter is the number of iterations               #
# q is a vector of quantiles                     #
# post is the posterior distribution              #
#
#####
# initialize variables
  rg <- array(c(rep(0, iter * 6)), dim = c(iter, 6))
  s <- c(rep(0, 6))
  x <- c(rep(0, 6))
  v <- c(rep(0, 6))
  actual <- (theta[1] * theta[2] * theta[3] * theta[4]) / (theta[5] *
theta[6])
  for(i in 1:6) {
    norm <- rnorm(num[i], theta[i], sigma[i])
    x[i] <- mean(norm)
    s[i] <- stdev(norm)
    v[i] <- sqrt(var(norm)/num[i])
    rg[, i] <- rt(iter, num[i] - 1) * v[i] + x[i]
  }
# end for #
  post <- (rg[, 1] * rg[, 2] * rg[, 3] * rg[, 4]) / ((rg[, 5] * rg[,
6]))
  ci <- round(quantile(sort(post), q), 3)
  cat("The lower limit, median, upper limit are: ", "\n")
  cat(ci, "\n")
  cat("The actual value is :", round(actual, 3), "\n")
  for(i in 1:6) {
    cat("x", i, " is: ", round(x[i], 3))
    cat(" s", i, " is: ", round(s[i], 3), "\n")
  }
# end for #
#####
# Graph the Prior Distribution #
#####

```

```

prior.vec <- as.vector(rg[, 1])
num.prior <- length(prior.vec)
range <- max(prior.vec) - min(prior.vec)
prior.hist <- hist(prior.vec, nclass = 100, plot = F)
prior.pct <- prior.hist$counts/num.prior
prior.breaks <- as.vector(prior.hist$breaks[prior.hist$breaks >
min(
  prior.vec)])
prior.min <- min(prior.vec)
prior.spline <- spline(prior.breaks, prior.pct, periodic = F,
boundary = 0,
  xmin = min(prior.breaks), xmax = max(prior.breaks))
prior.spline.plot <- cbind(prior.spline$x[prior.spline$x >=
prior.min],
  prior.spline$y[prior.spline$x >= prior.min])
x11 <- (as.integer(min(prior.vec)/2) - 1) * 2
if(prior.min < 2 && prior.min > 0) {
  x11 <- 0
}
xul <- (as.integer(max(prior.vec)/2) + 1) * 2
if(max(prior.vec) < 0 && max(prior.vec) > -2) {
  xul <- 0
}
graphsheat()
plot(prior.spline.plot, type = "l", xlab = " Theta1 ", ylab =
"Percent",
  yaxt = "n", xlim = c(x11, xul, lty = 1, col = 1))
ymax <- max(prior.pct)
loc <- round(x[1], 3)
shp <- round(v[1]^2, 3)
subtitle <- paste("T(", num[1] - 1, " , ", loc, " , ", shp, ")")
title(main = "Distribution of the Prior Density of Theta1", sub =
subtitle)
ylabls <- seq(0, as.integer(ymax * 100)/100, by = 0.01)
axis(2, at = ylabls, labels = paste(format(ylabls * 100),
rep("%", length(
  ylabls))), sep = "")
lines(prior.spline.plot, col = 3, lty = 1)
legend(x11, ymax, "% of Distribution", col = 3, lty = 1)
#####
# Graph the Posterior Distribution #
#####
post.vec <- as.vector(post)
num.post <- length(post.vec)
range <- max(post.vec) - min(post.vec)
post.hist <- hist(post.vec, nclass = 100, plot = F)
post.pct <- post.hist$counts/num.post
post.breaks <- as.vector(post.hist$breaks[post.hist$breaks >
min(post.vec)])
post.min <- min(post.vec)
post.spline <- spline(post.breaks, post.pct, periodic = F,
boundary = 0,
  xmin = min(post.breaks), xmax = max(post.breaks))
post.spline.plot <- cbind(post.spline$x[post.spline$x >=
post.min],
  post.spline$y[post.spline$x >= post.min])
x11 <- ((as.integer(min(post.vec)/10) - 1) * 10)

```

```

if(post.min < 10 && post.min > 0) {
  xll <- 0
}
xul <- ((as.integer(max(post.vec)/10) + 1) * 10)
if(max(post.vec) < 0 && max(post.vec) > -10) {
  xul <- 0
}
graphsheat()
plot(post.spline.plot, type = "l", xlab =
  " Theta1 *...*Theta4/Theta5*Theta6", ylab = "Percent", yaxt
= "n",
  xlim = c(xll, xul, lty = 1, col = 1))
ymax <- max(post.pct)
title(main = "Distribution of Posterior Density")
ylabls <- seq(0, as.integer(ymax * 100)/100, by = 0.01)
axis(2, at = ylabls, labels = paste(format(ylabls * 100),
rep("%", length(
  ylabls))), sep = "")
lines(post.spline.plot, col = 3, lty = 1)
for(i in 1:length(ci)) {
  lines(c(ci[i], ci[i]), c(0, ymax), col = 6, lty = 4)
}
lines(c(actual, actual), c(0, ymax), col = 4, lty = 1)
bottom <- ci[1] - post.min
top <- max(post.vec) - ci[3]
if(bottom >= top) {
  legplot <- xll
}
else if(bottom < top) {
  legplot <- ci[3] + 5
}
legend(legplot, ymax, c("% of Distribution", "5%", "50%", "95%",
"Actual Value"),
  col = c(3, 6, 4), lty = c(1, 4, 1))
}

```

Splus program for Method 2

```

function(alpha, beta, mu, tao, num, iter, q)
{
#####
#
# alpha and beta are the parameters of the #
# prior inverse gamma function #
# mu and tao are the parameters of the prior #
# normal function #
# theta is the vector of location parameters #
# sigma is the vector of shape parameters #
# num is the vector of sample sizes #
# x is the vector of sample means #
# s is the vector of sample standard deviations #
# rg is the vector of iter randomly generated #
# data points #

```

```

# q is a vector of desired quantiles usually #
#      5% 50% and 95% #
#####
# initialize variables
  rg <- array(c(rep(0, iter * 6)), dim = c(iter, 6))
  s <- c(rep(0, 6))
  x <- c(rep(0, 6))
  u <- c(rep(0, 6))
  sd.post <- c(rep(0, 6))
  df <- c(rep(0, 6))
  sigma <- c(rep(0, 6))
  theta <- c(rep(0, 6)) # start program
  for(i in 1:6) {
    sigma[i] <- 1/rgamma(1, alpha[i], 1/beta[i])
    theta[i] <- rnorm(1, mu[i], tao[i] * sigma[i])
    norm <- rnorm(num[i], theta[i], sigma[i])
    x[i] <- mean(norm)
    s[i] <- stdev(norm)
    beta.prime <- (1/(beta[i]) + ((num[i] - 1) * s[i]^2)/2) +
((num[i] * (
      x[i] - mu[i])^2)/(2 + 2 * num[i] * tao[i]))^(-1)
    df[i] <- 2 * alpha[i] + num[i] - 1
    u[i] <- (mu[i] + num[i] * tao[i] * x[i])/(num[i] * tao[i] +
1)
    sd.post[i] <- 1/((tao[i]^(-1) + num[i]) * (alpha[i] +
(num[i] - 1)/2
      ) * beta.prime)
    rg[, i] <- rt(iter, df[i]) * sd.post[i] + u[i]
  }
# end for #
  bayes.estimator <- (u[1] * u[2] * u[3] * u[4])/(u[5] * u[6])
  post <- (rg[, 1] * rg[, 2] * rg[, 3] * rg[, 4])/(rg[, 5] * rg[,
6])
  ci <- round(quantile(post, q), 4)
  cat("The Bayes Estimator is: ", round(bayes.estimator, 4), "\n")
  cat("The lower limit, median, upper limit are: ", "\n")
  cat(ci, "\n")
  for(i in 1:6) {
    cat("x", i, " is: ", round(x[i], 4))
    cat(" s", i, " is: ", round(s[i], 4))
    cat(" theta", i, " is: ", round(theta[i], 4))
    cat(" sigma", i, " is: ", round(sigma[i], 4), "\n")
  }
# end for #
# Comparison with Method 1 #
  method1(x, s, num, iter, q)
#####
# Graph the Prior Distribution #
#####
  prior.vec <- as.vector(rg[, 1])
  num.prior <- length(prior.vec)
  range <- max(prior.vec) - min(prior.vec)
  prior.hist <- hist(prior.vec, nclass = 100, plot = F)
  prior.pct <- prior.hist$counts/num.prior
  prior.breaks <- as.vector(prior.hist$breaks[prior.hist$breaks >
min(
  prior.vec)])

```



```

prior.min <- min(prior.vec)
prior.max <- max(prior.vec)
prior.spline <- spline(prior.breaks, prior.pct, periodic = F,
boundary = 0,
      xmin = min(prior.breaks), xmax = max(prior.breaks))
prior.spline.plot <- cbind(prior.spline$x[prior.spline$x >=
prior.min],
      prior.spline$y[prior.spline$x >= prior.min])
d <- 10
if(prior.max - prior.min <= 20) {
  d <- 2
}
xll <- (as.integer(min(prior.vec)/d) - 1) * d
if(prior.min < d && prior.min > 0) {
  xll <- 0
}
xul <- (as.integer(max(prior.vec)/d) + 1) * d
if(max(prior.vec) < 0 && max(prior.vec) > - d) {
  xul <- 0
}
graphsheat()
plot(prior.spline.plot, type = "l", xlab = " Theta1 ", ylab =
"Percent",
      yaxt = "n", xlim = c(xll, xul, lty = 1, col = 1))
ymax <- max(prior.pct)
loc <- round(u[1], 3)
shp <- round(sd.post[1]^2, 3)
subtitle <- paste("T(", df[1], " , ", loc, " , ", shp, ")")
title(main = "Distribution of the Prior Density of Theta1", sub =
subtitle)
ylab = seq(0, as.integer(ymax * 100)/100, by = 0.01)
axis(2, at = ylab, labels = paste(format(ylab * 100),
rep("%", length(
      ylab))), sep = "")
lines(prior.spline.plot, col = 3, lty = 1)
legend(xll, ymax, "% of Distribution", col = 3, lty = 1)
#####
# Graph the Posterior Distribution #
#####
post.vec <- as.vector(post)
num.post <- length(post.vec)
range <- max(post.vec) - min(post.vec)
post.hist <- hist(post.vec, nclass = 100, plot = F)
post.pct <- post.hist$counts/num.post
post.breaks <- as.vector(post.hist$breaks [post.hist$breaks >
min(post.vec)])
post.min <- min(post.vec)
post.max <- max(post.vec)
post.spline <- spline(post.breaks, post.pct, periodic = F,
boundary = 0,
      xmin = min(post.breaks), xmax = max(post.breaks))
post.spline.plot <- cbind(post.spline$x[post.spline$x >=
post.min],
      post.spline$y[post.spline$x >= post.min])
denom <- 10
if(post.max - post.min <= 20) {
  denom <- 2
}

```

```

}
xll <- ((as.integer(min(post.vec)/denom) - 1) * denom)
if(post.min < denom && post.min > 0) {
  xll <- 0
}
xul <- ((as.integer(max(post.vec)/denom) + 1) * denom)
if(max(post.vec) < 0 && max(post.vec) > - denom) {
  xul <- 0
}
graphsheat()
plot(post.spline.plot, type = "l", xlab =
      " Theta1 *...*Theta4/Theta5*Theta6", ylab = "Percent", yaxt
= "n",
      xlim = c(xll, xul, lty = 1, col = 1))
ymax <- max(post.pct)
title(main = "Distribution of Posterior Density")
ylabls <- seq(0, as.integer(ymax * 100)/100, by = 0.01)
axis(2, at = ylabls, labels = paste(format(ylabls * 100),
rep("%", length(
  ylabls))), sep = "")
lines(post.spline.plot, col = 3, lty = 1)
for(i in 1:length(ci)) {
  lines(c(ci[i], ci[i]), c(0, ymax), col = 6, lty = 4)
}
lines(c(bayes.estimator, bayes.estimator), c(0, ymax), col = 4,
lty = 1)
bottom <- ci[1] - xll
top <- max(post.vec) - ci[3]
if(bottom >= top) {
  legplot <- xll
}
else if(bottom < top) {
  legplot <- ci[3] + 5
}
legend(legplot, ymax, c("% of Distribution", "5%, 50%, 95%",
"Bayes Estimator"), col = c(3, 6, 4), lty = c(1, 4, 1))
}

```

Plus program for Method 3

```

function(mu, tao, sigma, num, iter, q, flag)
{
#####
#                                     #
# theta is the vector of location parameters #
# sigma is the vector of shape parameters #
# num is the vector of sample sizes #
# x is the vector of sample means #
# s is the vector of sample standard deviations #
# rg is the vector of iter randomly generated #
# data points #
# q is a vector of desired quantiles usually #
# 5% 50% 95% #
}

```

```
#####
# initialize variables
  rg <- array(c(rep(0, iter * 6)), dim = c(iter, 6))
  s <- c(rep(0, 6))
  x <- c(rep(0, 6))
  u <- c(rep(0, 6))
  s.post <- c(rep(0, 6))
  theta <- c(rep(0, 6))
  for(i in 1:6) {
    theta[i] <- rnorm(1, mu[i], tao[i]^2)
    norm <- rnorm(num[i], theta[i], sigma[i])
    x[i] <- mean(norm)
    #####
#   If flag is T assume sigma known      #
#   If flag is F use standard deviation  #
#   of the generated distribution        #
#####
    if(flag)
      s[i] <- sigma[i]
    else s[i] <- stdev(norm)
    u[i] <- (((num[i] * x[i])/s[i]^2) + (mu[i]/tao[i]^2))/((
      num[i]/s[i]^2) + (1/tao[i]^2))
    s.post[i] <- 1/((num[i]/s[i]^2) + (1/tao[i]^2))
    rg[, i] <- rnorm(iter, u[i], s.post[i])
  }
# end for #
bayes.estimator <- (u[1] * u[2] * u[3] * u[4])/(u[5] * u[6])
post <- (rg[, 1] * rg[, 2] * rg[, 3] * rg[, 4])/(rg[, 5] * rg[,
  6])
ci <- round(quantile(post, q), 4)
cat("The Bayes Estimator is: ", round(bayes.estimator, 4), "\n"
)
cat("The lower limit, median, upper limit are: ", "\n")
cat(ci, "\n")
for(i in 1:6) {
  cat("x", i, " is: ", round(x[i], 4))
  cat(" s", i, " is: ", round(s[i], 4))
  cat(" theta", i, " is: ", round(theta[i], 4), "\n")
}
# end for #
#####
#   Graph the Prior Distribution      #
#####
prior.vec <- as.vector(rg[, 1])
num.prior <- length(prior.vec)
range <- max(prior.vec) - min(prior.vec)
prior.hist <- hist(prior.vec, nclass = 100, plot = F)
prior.pct <- prior.hist$counts/num.prior
prior.breaks <- as.vector(prior.hist$breaks[prior.hist$breaks >
  min(prior.vec)])
prior.min <- min(prior.vec)
prior.max <- max(prior.vec)
prior.spline <- spline(prior.breaks, prior.pct, periodic = F,
  boundary = 0, xmin = min(prior.breaks), xmax = max(
  prior.breaks))
prior.spline.plot <- cbind(prior.spline$x[prior.spline$x >=
  prior.min], prior.spline$y[prior.spline$x >= prior.min])
```

```

    )
d <- 10
if(prior.max - prior.min <= 20) {
  d <- 2
}
xll <- (as.integer(min(prior.vec)/d) - 1) * d
if(prior.min < d && prior.min > 0) {
  xll <- 0
}
xul <- (as.integer(max(prior.vec)/d) + 1) * d
if(max(prior.vec) < 0 && max(prior.vec) > - d) {
  xul <- 0
}
graphsheat()
plot(prior.spline.plot, type = "l", xlab = " Theta1 ", ylab =
      "Percent", yaxt = "n", xlim = c(xll, xul, lty = 1, col
      = 1))
ymax <- max(prior.pct)
loc <- round(u[1], 3)
shp <- round(s.post[1], 3)
subtitle <- paste("N(", loc, " , ", shp, ")")
title(main = "Distribution of the first Prior Density", sub =
      subtitle)
ylabes <- seq(0, as.integer(ymax * 100)/100, by = 0.01)
axis(2, at = ylabes, labels = paste(format(ylabes * 100), rep(
      "%", length(ylabes))), sep = "")
lines(prior.spline.plot, col = 3, lty = 1)
legend(xll, ymax, "% of Distribution", col = 3, lty = 1)
#####
# Graph the Posterior Distribution #
#####
post.vec <- as.vector(post)
num.post <- length(post.vec)
range <- max(post.vec) - min(post.vec)
post.hist <- hist(post.vec, nclass = 100, plot = F)
post.pct <- post.hist$counts/num.post
post.breaks <- as.vector(post.hist$breaks [post.hist$breaks >
      min(post.vec)])
post.min <- min(post.vec)
post.max <- max(post.vec)
post.spline <- spline(post.breaks, post.pct, periodic = F,
      boundary = 0, xmin = min(post.breaks), xmax = max(
      post.breaks))
post.spline.plot <- cbind(post.spline$x[post.spline$x >=
      post.min], post.spline$y[post.spline$x >= post.min])
denom <- 10
if(post.max - post.min <= 20) {
  denom <- 2
}
xll <- ((as.integer(min(post.vec)/denom) - 1) * denom)
if(post.min < denom && post.min > 0) {
  xll <- 0
}
xul <- ((as.integer(max(post.vec)/denom) + 1) * denom)
if(max(post.vec) < 0 && max(post.vec) > - denom) {
  xul <- 0
}

```

```

graphsheet()
plot(post.spline.plot, type = "l", xlab =
     " Theta1 *...*Theta4/Theta5*Theta6", ylab = "Percent",
     yaxt = "n", xlim = c(xll, xul, lty = 1, col = 1))
ymax <- max(post.pct)
title(main = "Distribution of Posterior Density")
ylabls <- seq(0, as.integer(ymax * 100)/100, by = 0.01)
axis(2, at = ylabls, labels = paste(format(ylabls * 100), rep(
     "%", length(ylabls))), sep = "")
lines(post.spline.plot, col = 3, lty = 1)
for(i in 1:length(ci)) {
     lines(c(ci[i], ci[i]), c(0, ymax), col = 6, lty = 4)
}
lines(c(bayes.estimator, bayes.estimator), c(0, ymax), col = 4,
     lty = 1)
bottom <- ci[1] - xll
top <- max(post.vec) - ci[3]
if(bottom >= top) {
     legplot <- xll
}
else if(bottom < top) {
     legplot <- ci[3] + 5
}
legend(legplot, ymax, c("% of Distribution", "5%", "50%", "95%",
     "Bayes Estimator"), col = c(3, 6, 4), lty = c(1, 4, 1))
# Comparison with Method 1 #
method1(theta, s, num, iter, q)
}

```

BIBLIOGRAPHY

- [1] United States Environmental Protection Agency, EPA 540189-002, Chap 6, Volume 1. (1989).
- [2] Ananda, M. M. A., Singh, A. K., and Flatman, G. T. (1993), Bayesian confidence intervals for the product of three normal means. *Analytica Chimica Acta* **277** (1993) 503-510.
- [3] James Berger, *Statistical Decision Theory*, Springer-Verlag, New York (1985).
- [4] S. Hogg and A. Craig, *Introduction to Mathematical Statistics*, Macmillan, New York (1978).

VITA

**Graduate College
University of Nevada, Las Vegas**

Mark S Kirkham

Home Address:

**404 Heritage Circle N
Burnsville, Minnesota 55337**

Degrees:

**Bachelor of Science, Mathematics, 1992
California State University, Fullerton**

Thesis Title:

Confidence Intervals for Exposure Risk Assessment

Thesis Examination Committee:

**Chairperson, Dr. Malwane Ananda, Ph. D.
Committee Member, Dr. Ashok Singh, Ph. D.
Committee Member, Dr. Angel Muleshkov, Ph. D.
Graduate Faculty Representative, Dr. Laxmi Gewali, Ph. D.**