1-1-1981

Analysis of the one-way random effects model under slight departures from normality.

Bo-Jeong Kim
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation
Kim, Bo-Jeong, "Analysis of the one-way random effects model under slight departures from normality." (1981). Electronic Theses and Dissertations. 6125.
https://scholar.uwindsor.ca/etd/6125

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000 ext. 3208.
NOTE TO USERS

This reproduction is the best copy available.
ANALYSIS OF THE ONE-WAY RANDOM EFFECTS MODEL
UNDER SLIGHT DEPARTURES FROM NORMALITY

by

Bo-Jeong Kim

A Dissertation
Submitted to the Faculty of Graduate Studies
Through the Department of Mathematics
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
at The University of Windsor

Windsor, Ontario, Canada

1981
ABSTRACT

ANALYSIS OF THE ONE-WAY RANDOM-EFFECTS MODEL
UNDER SLIGHT DEPARTURES FROM NORMALITY

Bo-Jeong Kim

In this work the posterior distributions of the variance components in the analysis of variance in the one-way random-effects model are developed. The distributions of the effects are assumed to be known to belong to the class of exponential power functions.

Approximations and asymptotic functions of the posterior distributions are also evolved in order to provide somewhat simplified probability distributions with which to work.

- o -

iii
DEDICATION

This work is dedicated
To my Mother
and
To my Wife
For their Patience and unfailing
Encouragement and Support
During my Studies leading
To this Dissertation
ACKNOWLEDGEMENTS

The author wishes to acknowledge his indebtedness to the many people in the Department of Mathematics at the University of Windsor; faculty and students who have contributed to the success and enjoyment of the term of studies leading to this dissertation.

First, to Professor N. Shklov, who supervised these studies, sincere thanks for his supervision and advice since the author's arrival in the Department. Without this guidance these studies would not have attained fruition.

Appreciation for their courtesy in serving as members of the Examining Committee is also due to

Dr. B. M. Hill,
Dr. E. Kreyszig,
Dr. A. Raouf, and
Dr. D. S. Tracy.

The late Fr. D. Faught, Head of the Department of Mathematics at the time of the author's arrival at the University, is to be remembered for his kindness and help. Also thanks to Dr. F. W. Lemire, who followed Fr. Faught as Head of the Department. And lastly, thanks are due to the present Head of the Department, Dr. H. R. Atkinson, whose understanding and consideration materially aided the later phases of these studies.
Dr. E. Kreyszig is to be thanked for his encouragement and friendly advice during the more difficult period of the studies. Dr. D. S. Tracy, likewise, has been kind enough to take time to discuss both general and particular topics in the field of statistics.

From the University of the author's undergraduate years, Jeonbug (Chompuk) National University in Jeonju, Korea, Professors Jong-Geun Park and Dae-Shik Chun have followed the progress of their former student with interest and unfailing expectations. Their encouragement and kind wishes are appreciated and acknowledged with thanks.

The first two, and most difficult years that the author spent in Canada, were made easier and much more enjoyable by Dr. and Mrs. C. B. Crummey and family of Toronto who helped with the learning of the language and adjusting to Canadian customs. This friendship continues unabated to the present and the news of Dr. Crummey having suffered a stroke in March 1981 gives deep sorrow to all who know him.

Finally, to Dr. J. H. Toop, who as a student shared an office for several years, heartfelt gratitude and thanks for continuing encouragement and support in completing these studies when these became onerous. Also for his assistance over the years in teaching many of the finer points of the English language and western customs.

The Department of Mathematics is also to be thanked for financial aid.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Illustrations</td>
<td>ix</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. The formulation of the problem</td>
<td>5</td>
</tr>
<tr>
<td>3. The likelihood function $L(\mu, \sigma^2, \sigma^2_\alpha)$</td>
<td>16</td>
</tr>
<tr>
<td>4. The posterior inference on $\sigma^2$</td>
<td>29</td>
</tr>
<tr>
<td>5. The posterior inference on $\sigma^2_\alpha$</td>
<td>48</td>
</tr>
<tr>
<td>6. The posterior inference on $\sigma^2_\alpha / \sigma^2$</td>
<td>69</td>
</tr>
<tr>
<td>7. Conclusion</td>
<td>80</td>
</tr>
<tr>
<td>References (in order of citation)</td>
<td>88</td>
</tr>
<tr>
<td>References and Author Index</td>
<td></td>
</tr>
<tr>
<td>(Author alphabetically)</td>
<td>93</td>
</tr>
<tr>
<td>Bibliography</td>
<td>95</td>
</tr>
<tr>
<td>Vita Auctoris</td>
<td>102</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
LIST OF TABLES

TABLE

1. UPPER 100α PERCENT POINTS OF THE EXPONENTIAL POWER DISTRIBUTION FOR VARIOUS VALUES OF β IN UNITS OF STANDARD DEVIATION σ .................................... 14

2. RELATIONSHIP BETWEEN (β, y2) ........................................... 15

---

viii

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
LIST OF ILLUSTRATIONS

FIGURE

1. EXPONENTIAL POWER DISTRIBUTION WITH COMMON STANDARD DEVIATION FOR VARIOUS VALUES OF $\theta$ .......................... 13
Chapter 1

INTRODUCTION

The development of the analysis of variance dates from the pioneering work of Sir Ronald Aylmer Fisher (1890-1962), done mainly in the 1920s. Bayesian methods, by contrast, date back to the (posthumous) publication of Bayes Theorem in 1763-4 [1]. The popularity of Bayesian methods changes with time, varying inversely with the difficulty of applying contemporary methods.

Analysis of variance, to be reliably applied, presupposes rather stringent assumptions upon the statistical distributions of the sample groups and upon the population. These assumptions about the samples and population are seldom, if ever, met in applications. In fact, the distribution may be known to differ from the assumptions but, for lack of better methods, the assumptions have to be accepted. As a result, Bayesian approaches are one means used to investigate the effects which variations in the assumptions have upon results. It is found that considerable departures from the assumptions commonly cause small changes in the results and that these changes are fairly insensitive to these departures.

In the present work some effects of departures from normality requirements of samples and population will be investigated and the resulting changes in the analysis of variance will be exhibited.
1.1 THE ANALYSIS OF VARIANCE

The t-test is used to compare two population means for the significance of their differences. In practice, samples from the two populations are used and the null hypothesis is tested by using the distribution of the sample means. Pairwise comparisons of more than two populations using t-tests involves too much computation to be efficient and, moreover, raises to a prohibitive level the probability of Type I errors (i.e. rejection of a true hypothesis).

The analysis of variance extends this method to compare mean and variance estimates from individual samples with those from the samples combined, which are used to represent the population. The assumptions used for the analysis of variance are:

1) each sample group is normally distributed,
2) the variances for each are equal, and
3) the samples are independent.

Unbiased estimates of variance of the population are subjected to the F-test to determine if the differences are significant.

1.2 BAYES' THEOREM

Bayes' Theorem assumes that a number of mutually exclusive random events, $C_i$, (i.e. "causes"), having known a priori probabilities, $P(C_i)$, may be followed by an event, $E$, which has known conditional probabilities, $P(E \mid C_i)$, for the given causes.
If the event has been observed, it is desired to find the \textit{posteriori} probabilities, $P(C_i \mid E)$. This is given by the well-known relation

$$P(C_i \mid E) = \frac{P(C_i) P(E \mid C_i)}{\sum_{k} P(C_k) P(E \mid C_k)}.$$  

The most common difficulty arising in practice is that the $P(C_i)$ are not known \textit{a priori}. Nonetheless, Bayesian analysis is useful for making inferences about hypotheses, particularly when alternate methods of analysis are not available, or are too cumbersome to apply.

To use Bayesian methods, it is first necessary to make assumptions about the $P(C_i)$. These prior probabilities, $P(C_i)$, are subjectively chosen by the experimenter and represent \textit{a priori} beliefs about the distribution of the $P(C_i)$, usually based upon previous experience or personal preference. This 'subjective prior probability' is in contrast to 'objective probability' which may be interpreted in terms of frequency functions.

When these given subjective prior probability distributions are combined with the likelihood function, the resulting posterior distribution shows how the prior beliefs are modified by information coming from the data.

When the likelihood function obtained from the data looks \textit{sharp}, then it commonly happens that prior information permits the $P(C_i)$ to be plausibly estimated and, in general, the results are relatively insensitive to changes in the assumptions made about these $P(C_i)$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
1.3 THE SCOPE OF THIS WORK

An outline of the problem is presented in Chapter 2, where some previous work and methods are displayed for comparison. Chapter 3 develops the likelihood function; and chapters 4, 5, and 6 show the application of Bayesian methods in the analysis of variance, using relaxed assumptions upon the random effect.

The classical analysis of variance assumptions are relaxed by presenting individual observations, \( y_{ij} \), as being composed of the sum of a location parameter, \( \mu \), a random effect, \( a_i \), associated with the \( i \)-th group, and a random error, \( e_{ij} \), of the individual observation. That is:

\[
y_{ij} = \mu + a_i + e_{ij}
\]

The only restrictions upon the \( a_i \) and \( e_{ij} \) are that we know their variances and that their expectations are each zero. For analysis, it is assumed that these \( a_i \) and \( e_{ij} \) have a specific kind of non-normal distribution which exhibits these characteristics. The effects of departures from normality upon the distributions of certain parameters are found.
Chapter 2

THE FORMULATION OF THE PROBLEM

We shall examine the analysis of variance in the one-way random-effects model, that is:

\[ y_{ij} = \mu + a_i + e_{ij} \quad (2.1) \]

\[ (i = 1, 2, \ldots, I; \]
\[ j = 1, 2, \ldots, J), \]

where \( y_{ij} \) is the \( j^{th} \) observation in the \( i^{th} \) group, \( \mu \) is a location parameter, \( a_i \) is the random effect associated with the \( i^{th} \) group, and \( e_{ij} \) is the error in the \( (i,j)^{th} \) observation. We shall assume that the \( a_i \) are distributed independently of the \( e_{ij} \) and that:

\[ E(a_i) = 0, \quad \text{for all } i \]
\[ E(e_{ij}) = 0, \quad \text{for all } i, j \]
\[ \text{variance } (a_i) = \sigma_a^2 \quad \text{for all } i \]
\[ \text{variance } (e_{ij}) = \sigma_e^2 \quad \text{for all } i, j \]
\[ \text{covariance } (a_i, a_{i'}) = 0 \quad \text{for all } i \neq i' \]
\[ \text{covariance } (a_i, e_{ij}) = 0 \quad \text{for all } i, j. \]

We have, therefore:

\[ E(y_{ij} - \mu) = \sigma_e^2 + \sigma_a^2. \]

The parameters \( \sigma_e^2 \) and \( \sigma_a^2 \) are called "variance components" and the problem of estimating them has been attacked by many authors -- see, for example, Bross (1950) [2], Bulmer (1957) [3], Bush and Anderson (1963) [4], Crump (1946) [5] and (1951) [6],
Daniels (1939) [7], Fisher (1935) [8], Green (1954) [9], Haley (1963) [10] and many others. In most of these works the problem was analysed from a sampling-theory point of view. Two major difficulties arose and, in most of the above works, were left basically unresolved. One was the "negative estimated variance" problem. That is, using (2.1) and the assumption that the $a_i$ and the $e_{ij}$ are independent among themselves, the unbiased estimate of $\sigma^2_a$, viz.:

$$\hat{\sigma_a^2} = \frac{SSB / (I-1) - SSW / (J-1)}{J},$$

with

$$SSW = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \bar{y}_i)^2,$$

$$SSB = \sum_{i=1}^{I} J (\bar{y}_i - \bar{y})^2,$$

$$\bar{y}_i = \sum_{j=1}^{J} y_{ij} / J,$$

$$\bar{y} = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij} / IJ,$$

can clearly take on negative values. Attempts have been made to restrict the values of $\sigma_a^2$ to a positive range (see, for example, Herbach (1959) [11] and Thompson (1962) [12]). In the work by Thompson, the author uses a "restricted" maximum likelihood principle and the result is only slightly different from the traditional approach using the full maximum likelihood. However, this approach has the effect of destroying the unbiasedness property and of very much complicating the distributions upon which one makes inferences.
The second difficulty arising in the traditional approach is the sensitivity to departures from the underlying assumptions. Most writers presume that the error, $e_{ij}$, and the random-effects component, $a_i$, are normally distributed. However, Scheffe (1959) [13], has shown that non-normality, particularly in the $a_i$ and to a lesser degree in the $e_{ij}$, has a serious effect on the distributions of the criteria which one uses to make inferences about the parameters in the one-way model.

In an attempt to solve these (and other) problems, the Bayesian approach has been used by several recent authors. Tiao and Tan (1965) [14] assumed the normality of errors and random effects and used a non-informative prior probability distribution to develop posterior distributions of the variance components. Hill (1965) [15] used the same assumptions but went much further by using several prior probability distributions (of which the non-informative case is only one) in his development. Hill gave much more insight in his explanation of the case when $\alpha_k$ is negative.

The approaches used to analyse the second problem are many and varied. However, the underlying idea of most of these is the replacement of the normal distribution by a family of non-normal distributions or by an approximation to a distribution which is more general than the normal. In addition, most of the works place a heavy emphasis on the third and fourth moments as measures of non-normality. For example, E. S. Pearson (1928) [16] and (1929) [17] has studied the effect of universal
excess and skewness of a variable related to Student's t.

R. C. Geary (1936) [18] obtained the expression for the distribution of t in samples drawn from a slightly asymmetrical population. A. K. Gayen (1949) [19] and (1950a) [20] used the Edgeworth Series to develop the distribution of t and the variance ratio in random samples of any size drawn from a non-normal universe. An attempt to combine the Bayesian approach with a non-normal population was made by Box and Tiao (1964) [21]. In that work the authors used the following non-normal family of distributions to measure the effects of non-normality on the posterior distributions of the parameter involved in the test statistics:

$$f(y; \theta, \sigma, \beta) = k \exp \left[ -\frac{1}{2} \left( \frac{y-\theta}{\sigma} \right)^2 \right], \quad -\infty < y < \infty, \quad (2.3)$$

where,

$$k = \left\{ \left[ 1 + \frac{3}{2} (1+\beta) \right] \frac{1}{\sigma^2} \right\}^{-1}, \quad \text{and}$$

$$-\infty < \theta < \infty, \quad 0 < \sigma < \infty, \quad -1 < \beta < 1.$$ 

Here \( \beta \) is considered a "measure" of non-normality.

Although this approach has many limitations, it does illustrate one alternative for studying the effects of slight departures from normality in the distribution of the various parameters of an assumed model.
In this work we analyse the effect of non-normality by the Bayesian method of analysis. We limit our population to the class of exponential power distributions (2.3) in place of the normal distribution, but use the Edgeworth Series.

Our purpose is to develop the posterior distributions of the parameter in question and, if possible, their approximations or asymptotic expansions, since the distributions themselves will be quite complex and difficult to use. We study particularly the effects of non-normal values on kurtosis. Moreover, we shall also investigate the posterior distributions in the special case when \( \sigma^2 \) is negative.

Since it will play such an important role, a discussion of the class of exponential power distributions, (2.3), and the Edgeworth Series seems appropriate at this time.

Now the standardized normal distribution may be written:

\[
p(x) = k \exp \left[ -\frac{1}{2} |x|^q \right] \quad \text{with } q = 2.
\]

By allowing \( q \) to take values other than 2 we obtain what may be called the "class of exponential power distributions", These distributions were considered by Diananda (1949) [22], Box (1953b) [23], and Turner (1960) [24], with \( q = 2/(1+\beta) \). They can be written in the general form:

\[
p(y | \theta, \varphi, \beta) = k \varphi^{-1} \exp \left[ -\frac{1}{2} \frac{|y-\theta|^q}{\varphi^q} \right], \quad -\infty < y < \infty,
\]

where
$$k^{-1} = \left[ \frac{1 + \frac{1 + \beta}{2}}{2} \right]^{\frac{1+\beta}{2}}$$

and $0 > \beta > 0$, $-\infty < \theta < \infty$, $-1 < \beta < 1$.

In (2.4), $\theta$ is a location parameter and $\beta$ is a scale parameter.

It can be readily shown that:

$$E(y) = \theta,$$  \hspace{2cm} (2.5)

$$\text{var}(y) = \sigma^2$$

$$= 2^{1+\beta} \left\{ \frac{\Gamma_{\frac{1}{2}}(1+\theta)}{\Gamma_{\frac{3}{2}}(1+\theta)} \right\} \frac{\sigma^2}{\theta^2}.$$  

We may alternatively express (2.4) as:

$$P(y|\theta, \sigma^2, \beta) = \omega(\beta) \sigma^{-1} \exp \left\{ -\sigma \left[ \frac{y - \theta}{\sigma^2} \right] \right\} \left[ \frac{1}{\Gamma(\beta)} \right]^{\frac{1}{\beta}}$$

$$\omega(\beta) = \left\{ \frac{\Gamma_{\frac{3}{2}}(1+\theta)}{\Gamma_{\frac{1}{2}}(1+\theta)} \right\}^{\frac{1}{\beta}},$$

$$\theta > 0, \quad -\infty < \theta < \infty, \quad -1 < \beta < 1.$$  

The parameters $\theta$ and $\sigma^2$ are then the mean and the standard deviation of the population, respectively. The parameter $\beta$ can be regarded as a measure of kurtosis, indicating the extent of the "non-normality" of the population. In particular, when $\beta = 0$, the distribution is normal. When $\beta = 1$, the distribution
is the double exponential:

\[ P(y|\theta, \sigma, \beta = 1) = \frac{1}{\sqrt{2} \sigma} \exp \left[-\frac{1}{2} \frac{|y-\theta|}{\sigma}\right]. \quad -\infty < y < \infty \tag{2.7} \]

Finally, when \( \beta \) tends to -1, it can be shown that the distribution tends to the rectangular distribution:

\[
\lim_{\beta \to -1} P(y|\theta, \sigma, \beta) = \frac{1}{2\sqrt{3} \sigma}; \quad \theta - \sqrt{3} \sigma < y < \theta + \sqrt{3} \sigma. \tag{2.8}
\]

Fig. 1 shows the exponential power distribution for various values of \( \beta \). The distributions shown have common mean and standard deviation. We see that for \( \beta > 0 \) the distributions are leptokurtic, and for \( \beta < 0 \) they are platykurtic.

To further illustrate the effect of \( \beta \) on the shape of the distribution, Table 1 gives the upper 100\% percent points in units of \( \sigma \) for various choices of \( \beta \) with \( \theta \) assumed zero. Except for \( \beta \) equal to zero and one, and the limiting case \( \beta \to -1 \), the percentage points in the table were calculated by numerical integration on a computer (Tiao and Lund (1970) [25]).

In (2.6) we have employed the non-normality measure which makes the double exponential and the rectangular distribution equally discrepant from the normal. However, we might have used, for example, the familiar kurtosis measure

\[ \gamma_4 = k_4/k_2^2 \]

for the class of exponential power distributions, where \( k_n \) is the \( n^{th} \) cumulant of the distribution.

It is readily shown that:
\[
\gamma_2 = \left\{ \left[ \frac{\xi}{2} (1 + \xi) \right] \right\} \left\{ \left[ \frac{\xi}{2} (1 + \xi) \right] \right\}^2 - 3.
\] 

Table 2 gives the values of \(\gamma_2\) for the various values of \(\xi\). The value of \(\gamma_2\), for the double exponential distribution appears as 3 and, for the rectangular distribution, as -1.2.

H. Cramér (1928) [29] has shown that the Edgeworth Series provides an asymptotic expansion of the probability distribution in powers of \(n^{\frac{1}{2}}\), with a remainder term of the same order as the first term neglected. The terms of order \(n^{\frac{1}{2}}\) contain the moments \(\mu_3, \mu_4, \ldots, \mu_{4+2}\). In this work we shall not go beyond the third and fourth moments. In order to simplify the notation, we introduce \(Y_1\) and \(Y_2\), where:

\[Y_1 = \mu_3/\sigma^3\quad\text{and}\quad Y_2 = \mu_4/\sigma^4 - 3.\]

Thus the Edgeworth Series is:

\[
\phi(x) = \phi(x) - \left[ \frac{\gamma_1}{3!} \right] \phi^{(3)}(x) + \left[ \frac{\gamma_2}{4!} \right] \phi^{(4)}(x) + \left[ \frac{10 \gamma^2_2}{6!} \right] \phi^{(6)}(x),
\]

where \(\phi(x) = \frac{1}{\sqrt{2\pi}} \exp \left[ -x^2/2\sigma^2 \right]\),

and \(\phi^{(v)}(x)\) is the \(v\)th derivative of \(\phi(x)\).
Figure 1

Fig. 1. Exponential Power Distributions With Common
Standard Deviation For Various Values of $\beta$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\alpha$</th>
<th>0.25</th>
<th>0.10</th>
<th>0.05</th>
<th>0.025</th>
<th>0.01</th>
<th>0.005</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.00</td>
<td></td>
<td>0.87</td>
<td>1.39</td>
<td>1.56</td>
<td>1.65</td>
<td>1.70</td>
<td>1.71</td>
<td>1.73</td>
</tr>
<tr>
<td>-0.75</td>
<td></td>
<td>0.84</td>
<td>1.36</td>
<td>1.57</td>
<td>1.71</td>
<td>1.84</td>
<td>1.91</td>
<td>2.05</td>
</tr>
<tr>
<td>-0.50</td>
<td></td>
<td>0.80</td>
<td>1.35</td>
<td>1.61</td>
<td>1.81</td>
<td>2.03</td>
<td>2.16</td>
<td>2.41</td>
</tr>
<tr>
<td>-0.25</td>
<td></td>
<td>0.73</td>
<td>1.31</td>
<td>1.63</td>
<td>1.89</td>
<td>2.18</td>
<td>2.37</td>
<td>2.75</td>
</tr>
<tr>
<td>0.00</td>
<td></td>
<td>0.67</td>
<td>1.28</td>
<td>1.64</td>
<td>1.96</td>
<td>2.33</td>
<td>2.58</td>
<td>3.09</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td>0.62</td>
<td>1.25</td>
<td>1.65</td>
<td>2.02</td>
<td>2.46</td>
<td>2.77</td>
<td>3.43</td>
</tr>
<tr>
<td>0.50</td>
<td></td>
<td>0.58</td>
<td>1.22</td>
<td>1.65</td>
<td>2.06</td>
<td>2.58</td>
<td>2.94</td>
<td>3.76</td>
</tr>
<tr>
<td>0.75</td>
<td></td>
<td>0.53</td>
<td>1.18</td>
<td>1.64</td>
<td>2.09</td>
<td>2.68</td>
<td>3.10</td>
<td>4.08</td>
</tr>
<tr>
<td>1.00</td>
<td></td>
<td>0.49</td>
<td>1.14</td>
<td>1.63</td>
<td>2.12</td>
<td>2.77</td>
<td>3.28</td>
<td>4.39</td>
</tr>
</tbody>
</table>

Table 1. Upper 100\(\alpha\) Percent Points Of The Exponential Power Distribution For Various Values Of $\beta$ In Units Of Standard Deviation $\sigma$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 2

<table>
<thead>
<tr>
<th>θ</th>
<th>( \gamma_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.00</td>
<td>-1.2</td>
</tr>
<tr>
<td>-0.75</td>
<td>-1.07</td>
</tr>
<tr>
<td>-0.50</td>
<td>-0.81</td>
</tr>
<tr>
<td>-0.25</td>
<td>-0.45</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.25</td>
<td>0.55</td>
</tr>
<tr>
<td>0.50</td>
<td>1.22</td>
</tr>
<tr>
<td>0.75</td>
<td>2.03</td>
</tr>
<tr>
<td>1.00</td>
<td>3.00</td>
</tr>
</tbody>
</table>

Table 2. Relationship Between \((\theta, \gamma_2)\).
Chapter 3

THE LIKELIHOOD FUNCTION \( L(\mu, \sigma^2, \sigma_a^2) \)

Before the posterior density functions are found the likelihood function \( L(\mu, \sigma^2, \sigma_a^2) \) will be cast into a form in which the non-normality is collected into a single term, \( D \), where its effects may be exhibited and studied conveniently.

To do this, the distribution of \( a_1 \), \( f(a_1) \), is expanded in the Edgeworth series in which the parameter of symmetry, \( Y_1 \), is zero because of the fact that we are using the class of exponential power distributions. The non-zero terms are then used in the joint likelihood function \( L(\mu, \sigma^2, \sigma_a^2) \) which will then be further simplified.

It is assumed that the prior opinion about \( \mu \) is diffuse and that this is effectively independent of \( (\sigma^2, \sigma_a^2) \). That is,

\[
P(\mu, \sigma^2, \sigma_a^2) = p(\mu) p(\sigma^2, \sigma_a^2)
\]  

(3.1)
where, because \( \mu \) is diffuse, \( p(\mu) \) is merely a multiplicative constant.

The \( \mu \) is therefore integrated out of \( D \) to yield the term \( E \) containing all the non-normality information and it is found that only a few terms of \( E \) are useful for the analysis of the posterior density function.
3.1 THE LIKELIHOOD FUNCTION

Using the Edgeworth series the distribution of \( a_i \) is approximated by:

\[
f(a_i) \approx \left\{ 1 + \frac{Y^2}{4!} \left[ \left( \frac{a_i}{\sigma_a} \right)^4 - 6 \left( \frac{a_i}{\sigma_a} \right)^2 + 3 \right] \right\} \frac{1}{\sqrt{2\pi} \sigma_a} \exp \left[ - \frac{a_i^2}{2\sigma_a^2} \right].
\]

Using this expansion, the joint likelihood function, \( L(\mu, \sigma^2, \sigma_a^2) \) becomes:

\[
L = \left( \sigma_a \sqrt{2\pi} \right)^{-1} \exp \left[ - \frac{1}{2\sigma_a^2} \sum_i \sum_j (y_{ij} - \mu - a_i)^2 \right] \prod_{i=1}^{I} \left\{ 1 + \frac{Y^2}{4!} \left[ \left( \frac{a_i}{\sigma_a} \right)^4 - 6 \left( \frac{a_i}{\sigma_a} \right)^2 + 3 \right] \right\} da_i \ldots da_I.
\]

which may be more conveniently expressed as:

\[
L = \left( \sigma_a \sqrt{2\pi} \right)^{-1} \exp \left[ - \frac{1}{2\sigma_a^2} \sum_i \sum_j (y_{ij} - \mu - a_i)^2 \right] \prod_{i=1}^{I} H(a_i) da_i,
\]

where

\[
H(a_i) = 1 + \frac{Y^2}{4!} \left[ \left( \frac{a_i}{\sigma_a} \right)^4 - 6 \left( \frac{a_i}{\sigma_a} \right)^2 + 3 \right].
\]

This \( H(a_i) \) is chosen to gather terms for integration at a later stage.
Simplifying, we obtain:

\[
L = (\sigma_1 \sqrt{2\pi})^{-I} (\sigma_2 \sqrt{2\pi})^{-1} \exp \left[ - \frac{SSW}{2\sigma_1^2} - \frac{SSB}{2(\sigma_2^2 + J\sigma_2^4)} - \frac{IJ(\mu - \bar{y})^2}{2(\sigma_2^2 + J\sigma_2^4)} \right].
\]

\[
\prod_{i=1}^{I} \int_{a_i} H(a_i) \exp \left[ - \frac{(a_i - P_i \bar{y})^2}{2Q} \right] da_i, \quad (3.5)
\]

where

\[
SSW = \sum_{i} \sum_{j} (y_{ij} - \bar{y}_i)^2,
\]

\[
SSB = \sum_{i} \sum_{j} (\bar{y}_i - \bar{y})^2,
\]

\[
P_i = \frac{J\sigma_2^4}{\sigma_2^2 + J\sigma_2^4} (\bar{y}_i - \mu),
\]

\[
Q = \frac{\sigma_2^2 \sigma_2^4}{\sigma_2^2 + J\sigma_2^4}.
\]

Proceeding to evaluate the last factor we find:

\[
\prod_{i=1}^{I} \int_{a_i} H(a_i) \exp \left[ - \frac{(a_i - P_i)^2}{2Q} \right] da_i
\]

\[
= \left\{ \left[ 1 + \frac{a_i}{4} \left( \frac{a_i}{\sigma_2^4} - \frac{a_i^4}{\sigma_2^4} + 3 \right) \right] \exp \left[ - \frac{(a_i - P_i)^2}{2Q} \right] \right\} \prod_{i=1}^{I} da_i
\]

\[
= \sqrt{2\pi Q} \left\{ 1 + \frac{a_i}{2} \left[ \frac{3J\sigma_2^4}{(\sigma_2^2 + J\sigma_2^4)^2} - \frac{6JP_i^4}{(\sigma_2^2 + J\sigma_2^4)^2} + \frac{P_i^4}{\sigma_2^4} \right] \right\}
\]

Therefore the likelihood function \( L(\mu, \sigma_2^2, \sigma_2^4) \) becomes:
\[
L = (\sigma \sqrt{2\pi})^{-\frac{1}{2}} \left( \sigma^2 + J_{\alpha}^2 \right)^{-\frac{1}{2}} \exp \left[ \frac{SSW}{2\sigma^2} - \frac{SSB}{2(\sigma^2 + J_{\alpha}^2)} - \frac{IJ(\mu - \bar{y})^2}{2(\sigma^2 + J_{\alpha}^2)} \right].
\]

\[
\alpha (\sigma^2) \left( \sigma^2 + J_{\alpha}^2 \right)^{-\frac{1}{2}} \exp \left[ \frac{SSW}{2\sigma^2} - \frac{SSB}{2(\sigma^2 + J_{\alpha}^2)} - \frac{IJ(\mu - \bar{y})^2}{2(\sigma^2 + J_{\alpha}^2)} \right] \cdot D
\]

where

\[
D = \prod_{i=1}^{n} \left\{ 1 + \frac{Y_i}{4!} \left[ \frac{3 J^2 \sigma_a^4}{(\sigma^2 + J_{\alpha}^2)^2} - \frac{6 J R^2}{(\sigma^2 + J_{\alpha}^2)^2} + \frac{I^2}{\sigma_a^4} \right] \right\}. \tag{3.7}
\]

This \( D \) is chosen to segregate the non-normality information from the data.

Now, to evaluate the factor \( D \) we proceed as follows:

\[
D = \prod_{i=1}^{n} \left\{ 1 + \frac{Y_i}{4!} \left[ \frac{3 J^2 \sigma_a^4}{(\sigma^2 + J_{\alpha}^2)^2} - \frac{6 J R^2}{(\sigma^2 + J_{\alpha}^2)^2} + \frac{I^2}{\sigma_a^4} \right] \right\}
\]

\[
= \prod_{i=1}^{n} \left\{ 1 + \frac{Y_i}{8(\sigma^2 + J_{\alpha}^2)} \left[ \frac{3 J^2 \sigma_a^4}{(\sigma^2 + J_{\alpha}^2)^2} - \frac{6 J R^2}{(\sigma^2 + J_{\alpha}^2)^2} + \frac{I^2}{\sigma_a^4} \right] \right\}
\]

\[
= \prod_{i=1}^{n} \left[ b_0 + b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 \right], \tag{3.8}
\]

where

\[
b_0 = 1 + \frac{Y_i}{8(\sigma^2 + J_{\alpha}^2)} \]

\[
b_1 = \frac{-Y_i}{4(\sigma^2 + J_{\alpha}^2)} \]

\[
b_2 = \frac{Y_i}{24(\sigma^2 + J_{\alpha}^2)} \]

Again by choosing a convenient substitution, this may now be written:
\[ D = \prod_{i=1}^{\infty} (b_i + c_i) \]

\[ = b_0 + b_0 \sum_{i} \frac{r}{i} \sum_{c_i} a_i c_i + b_0 \sum_{i} \frac{r^2}{i} \sum_{c_i} \sum_{c_{i,j}} a_{i,j} a_{i,j} + \cdots + \frac{r}{\xi} c_i , \]

where

\[ c_i = b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 , \]

or, in terms of the previous notation:

\[ = b_0 + b_0 \sum_{i} \left[ b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 \right] - \]

\[ + b_0 \sum_{i} \left[ b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 \right] \cdot \left[ b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 \right] - \]

\[ + b_0 \sum_{i} \left[ b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 \right] \cdot \left[ b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 \right] \cdot \]

\[ \cdot \left[ b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 \right] + \cdots \quad (3.9) \]

\[ + \prod_{i} \left[ b_1 (\bar{y}_i - \mu)^2 + b_2 (\bar{y}_i - \mu)^4 \right] . \]

If now we express \( D \) in more detail for use in future work, we obtain, using the fact that: \( \bar{y}_i - \mu = (\bar{y} - \mu) + (\bar{y}_i - \bar{y}) \),

\[ D = b_0 + b_0 \sum_{A_{h}} b_{A_h} \left\{ \sum_{S_{e}} \sum_{S_{j}} \sum_{S_{i}} \sum_{S_{i,j}} \sum_{S_{i,j,k}} \sum_{S_{i,j,k,l}} \left[ \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \right] \right\} - \]

\[ + b_0 \sum_{A_{h}} b_{A_h} \left\{ \sum_{S_{e}} \sum_{S_{j}} \sum_{S_{i}} \sum_{S_{i,j}} \sum_{S_{i,j,k}} \sum_{S_{i,j,k,l}} \left[ \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \right] \right\} \]

\[ + b_0 \sum_{A_{h}} b_{A_h} \left\{ \sum_{S_{e}} \sum_{S_{j}} \sum_{S_{i}} \sum_{S_{i,j}} \sum_{S_{i,j,k}} \sum_{S_{i,j,k,l}} \left[ \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \right] \right\} \]

\[ + \cdots + \]

\[ + b_0 \sum_{A_{h}} b_{A_h} \ldots b_{A_l} \left\{ \sum_{S_{e}} \sum_{S_{j}} \sum_{S_{i}} \sum_{S_{i,j}} \sum_{S_{i,j,k}} \sum_{S_{i,j,k,l}} \left[ \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \sum_{S_{i,j,k,l}} \right] \right\} \]

\[ (3.10) \]
To simplify the above expression we use a convention. Let us denote

\[ G_{2A_1,2A_2}(s) = \sum \binom{2A_1}{s_1} \binom{2A_2}{s_2} \left[ \sum_{i_1 i_2 \ldots i_k} (\bar{y}_{i_1} - \bar{y})^i (\bar{y}_{i_2} - \bar{y})^j \right], \]

the summation being taken over all possible \((s_1, s_2)\) such that \(s = s_1 + s_2\), \(0 \leq s_1 \leq 2A_1\), \(0 \leq s_2 \leq 2A_2\).

Now, in general:

\[ G_{2A_1,2A_2,\ldots,2A_c}(s) = \sum \binom{2A_1}{s_1} \binom{2A_2}{s_2} \ldots \binom{2A_c}{s_c} \left[ \sum_{i_1 i_2 \ldots i_c} (\bar{y}_{i_1} - \bar{y})^i (\bar{y}_{i_2} - \bar{y})^j \ldots (\bar{y}_{i_c} - \bar{y})^k \right], \quad (3.11) \]

the summation again being taken over all possible combinations of \((s_1, s_2, \ldots, s_c)\) such that \(s_1 + s_2 + \ldots + s_c = s\), \(0 \leq s_i \leq 2A_i\), \(\ldots\), \(0 \leq s_c \leq 2A_c\).

Then

\[ \sum_{S=0}^{2A_1} \sum_{S_2=0}^{2A_2} \binom{2A_1}{s_1} \binom{2A_2}{s_2} \left[ \sum_{i_1 i_2 \ldots i_k} (\bar{y}_{i_1} - \bar{y})^i (\bar{y}_{i_2} - \bar{y})^j \right] \]

\[ = \sum_{S=0}^{2A_1+2A_2} \sum \binom{2A_1}{s_1} \binom{2A_2}{s_2} \left[ \sum_{i_1 i_2 \ldots i_k} (\bar{y}_{i_1} - \bar{y})^i (\bar{y}_{i_2} - \bar{y})^j \right] \]

\[ = \sum_{S=0}^{2A_1+2A_2} G_{2A_1,2A_2}(s), \quad (3.12) \]

where \(S = s_1 + s_2\).

That is to say, the powers of \((\bar{y} - \mu)\) are now segregated into the \(G\) terms.

Using the above convention we may now write \(D\) as:
\[ D = b^I_0 + b_{(4)}^I \left\{ b_1 \left[ (\bar{\tau} - \mu)^2 G_{24}(0) + (\bar{\tau} - \mu) G_{24}(1) + G_{24}(2) \right] \right. \]
\[ + b_2 \left[ (\bar{\tau} - \mu)^3 G_{24}(0) + (\bar{\tau} - \mu)^2 G_{24}(1) + (\bar{\tau} - \mu) G_{24}(2) + (\bar{\tau} - \mu) G_{24}(3) + G_{24}(4) \right] \]
\[ + b_3 \left[ (\bar{\tau} - \mu)^4 G_{24}(0) + (\bar{\tau} - \mu)^3 G_{24}(1) + (\bar{\tau} - \mu)^2 G_{24}(2) + (\bar{\tau} - \mu) G_{24}(3) + G_{24}(4) \right] \]
\[ \left. + b_4 \left[ (\bar{\tau} - \mu)^5 G_{24}(0) + (\bar{\tau} - \mu)^4 G_{24}(1) + (\bar{\tau} - \mu)^3 G_{24}(2) + (\bar{\tau} - \mu)^2 G_{24}(3) + (\bar{\tau} - \mu) G_{24}(4) \right] \right\} \]
\[ + b_5 \left[ (\bar{\tau} - \mu)^6 G_{24}(0) + (\bar{\tau} - \mu)^5 G_{24}(1) + (\bar{\tau} - \mu)^4 G_{24}(2) + (\bar{\tau} - \mu)^3 G_{24}(3) + (\bar{\tau} - \mu)^2 G_{24}(4) + (\bar{\tau} - \mu) G_{24}(5) + G_{24}(6) \right] \]
\[ + b_6 \left[ (\bar{\tau} - \mu)^7 G_{24}(0) + (\bar{\tau} - \mu)^6 G_{24}(1) + (\bar{\tau} - \mu)^5 G_{24}(2) + (\bar{\tau} - \mu)^4 G_{24}(3) + (\bar{\tau} - \mu)^3 G_{24}(4) + (\bar{\tau} - \mu)^2 G_{24}(5) + (\bar{\tau} - \mu) G_{24}(6) \right] \]
\[ + \ldots + b_{n} \left[ (\bar{\tau} - \mu)^n G_{24}(0) + (\bar{\tau} - \mu)^{n-1} G_{24}(1) + \ldots + (\bar{\tau} - \mu)^2 G_{24}(n-2) + (\bar{\tau} - \mu) G_{24}(n-1) + G_{24}(n) \right] \]
\[ + \ldots + \]
\[ + b_{2n} \left[ (\bar{\tau} - \mu)^{2n} G_{24}(0) + (\bar{\tau} - \mu)^{2n-1} G_{24}(1) + \ldots + (\bar{\tau} - \mu)^2 G_{24}(2n-2) + (\bar{\tau} - \mu) G_{24}(2n-1) + G_{24}(2n) \right] \]
\[ + \ldots + \]
\[ + b_{2n+1} \left[ (\bar{\tau} - \mu)^{2n+1} G_{24}(0) + (\bar{\tau} - \mu)^{2n} G_{24}(1) + \ldots + (\bar{\tau} - \mu)^2 G_{24}(2n) + (\bar{\tau} - \mu) G_{24}(2n+1) + G_{24}(2n+2) \right] \]
\[
(3.13)
\]
If we collect terms involving \((\overline{y}-\mu)^2\) and simplify the coefficients of these terms, we obtain for that term:

\[
(\overline{y}-\mu)^2 \left\{ b_0 \left[ b_1 G_2(0) + b_2 G_4(2) \right] + \right.
\]

\[
+ b_2 \left[ b_1 b_2 G_{2,2}(4) + b_2 b_2 G_{2,4}(6) \right] + \right.
\]

\[
+ b_3 \left[ b_1 b_2 G_{3,2}(4) + b_1 b_2 b_1 G_{3,4}(6) + \ldots \right] + 
\]

\[
+ b_4 b_2 G_{4,4}(10) \right\} + \ldots + 
\]

\[
+ b_4 \sum_{\lambda_1} \sum_{\lambda_2} \ldots \sum_{\lambda_4} b_4^L \lambda_1 \lambda_2 \ldots \lambda_4 \left\{ \sum_{i=1}^{p} 2\lambda_i - 2 \right\} \right) 
\]

Now returning to the general expression and collecting the terms containing \((\overline{y}-\mu)^r\) and simplifying, we find for \(D:\)

\[
D = b_0^r + 
\]

\[
+ \sum_{r=2}^{\infty} (\overline{y}-\mu)^r \left\{ \sum_{p=1}^{r-1} b_0^{r-p} \sum_{\lambda_1=1}^{p} \ldots \sum_{\lambda_p=1}^{p} b_0^{\lambda_1} \ldots b_0^{\lambda_p} \right) \right. 
\]

\[
\times G_{\lambda_1,\lambda_2,\ldots,\lambda_p} \left( \sum_{i=1}^{p} 2\lambda_i - r \right) \right), \quad (3.15)
\]

where

\[
G_{\lambda_1,\lambda_2,\ldots,\lambda_p} \left( \sum_{i=1}^{p} 2\lambda_i - r \right) = 1, \text{ if } \sum_{i=1}^{p} 2\lambda_i - r = 0.
\]

(Note that \(G_{\lambda_1}(0) = G_{2\lambda_1,\lambda_2}(0) = \ldots = G_{2\lambda_1,2\lambda_2,\ldots,\lambda_p}(0) = 1\).)
3.2 THE POSTERIOR DISTRIBUTION

We have cast the factor D into the above form in order to obtain a simplified form for the likelihood function \( L(\mu, \sigma^2, \sigma^2_\alpha) \). By denoting a subjectively-chosen prior density as \( p(\mu, \sigma^2, \sigma^2_\alpha) \), the posterior density becomes:

\[
f(\mu, \sigma^2, \sigma^2_\alpha) \propto p(\mu, \sigma^2, \sigma^2_\alpha) \cdot L(\mu, \sigma^2, \sigma^2_\alpha).
\]

Since all of our analysis will be based on the assumption of diffuse prior opinion for \( \mu \) effectively independent of that for \( (\sigma^2, \sigma^2_\alpha) \), we have, roughly,

\[
p(\mu, \sigma^2, \sigma^2_\alpha) = p(\mu) \cdot p(\sigma^2, \sigma^2_\alpha)
\]

\[
= \text{const.} \cdot p(\sigma^2, \sigma^2_\alpha)
\]

\[
\propto p(\sigma^2, \sigma^2_\alpha).
\]

That is to say, the prior density function \( p(\mu, \sigma^2, \sigma^2_\alpha) \) is defined up to a multiplicative constant, \( p(\mu) \), which leaves the posterior density function, \( f(\mu, \sigma^2, \sigma^2_\alpha) \), unchanged. This constant will have no effect upon \( f(\sigma^2, \sigma^2_\alpha) \) since it will cancel upon normalizing the product on the right hand side of \( f \propto p \cdot L \).

The above assumption of diffuse prior opinion for \( \mu \) is not unduly restrictive and leads to the approximate marginal posterior density \( f(\sigma^2, \sigma^2_\alpha) \) by integrating \( \mu \) out of \( f(\mu, \sigma^2, \sigma^2_\alpha) \), thus:
\[
f(\sigma^2, \sigma_a^2) = \int_{-\infty}^{\infty} f(\mu, \sigma^2, \sigma_a^2) \, d\mu
\]

\[
= \int_{-\infty}^{\infty} p(\mu, \sigma^2, \sigma_a^2) \cdot L(\mu, \sigma^2, \sigma_a^2) \, d\mu
\]

\[
\propto p(\sigma^2, \sigma_a^2) \int_{-\infty}^{\infty} (\sigma^2)^{-1/2} (\sigma^2 + J \sigma_a^2)^{-1/2} \exp\left[-\frac{SSW}{2\sigma^2} - \frac{SSB}{2(\sigma^2 + J \sigma_a^2)} - \frac{IJ (\mu-\bar{y})^2}{2(\sigma^2 + J \sigma_a^2)}\right] \cdot D \, d\mu
\]

\[
\propto p(\sigma^2, \sigma_a^2) \int_{-\infty}^{\infty} (\sigma^2)^{-1/2} \exp\left[-\frac{SSB}{2(\sigma^2 + J \sigma_a^2)}\right] \cdot \exp\left[-\frac{SSW}{2\sigma^2}\right] (\sigma^2 + J \sigma_a^2)^{-1/2} \cdot \exp\left[-\frac{IJ (\mu-\bar{y})^2}{2(\sigma^2 + J \sigma_a^2)}\right] \, d\mu .
\]

By use of another convenient substitution this may be written:

\[
f(\sigma^2, \sigma_a^2) \propto p(\sigma^2, \sigma_a^2) (\sigma^2)^{-1/2} \exp\left[-\frac{SSW}{2\sigma^2}\right] \cdot (\sigma^2 + J \sigma_a^2)^{-1/2} \exp\left[-\frac{SSB}{2(\sigma^2 + J \sigma_a^2)}\right] \cdot E ,
\]

where

\[
E = \int_{-\infty}^{\infty} \exp\left[-\frac{IJ (\mu-\bar{y})^2}{2(\sigma^2 + J \sigma_a^2)}\right] \, d\mu .
\]

This \( E \) is the segregated non-normality information obtained by integrating the \( \mu \) out of \( D \).

Now \( D \) is a polynomial in \( (\bar{y}-\mu) \), where the powers range from 0 to 41, and

\[
\text{Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.}
\]
\[
\int_{\mu} \left( \bar{y} - \mu \right)^{r} \cdot \exp \left[ - \frac{IJ (\bar{y} - \mu)^2}{2(\sigma^2 + J \sigma^2)} \right] d\mu \\
= \begin{cases} \\
\sqrt{2\pi} \left( \frac{\sigma^2 + J \sigma^2}{IJ} \right)^{\frac{1}{2}} \cdot 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (r-1) \cdot \left( \frac{\sigma^2 + J \sigma^2}{IJ} \right)^{\frac{r}{2}}, & \text{for } r \text{ even} \\
0, & \text{for } r \text{ odd,} \\
\end{cases}
\]

so that we have:

\[
E = \sqrt{2\pi} \left( \frac{\sigma^2 + J \sigma^2}{IJ} \right)^{\frac{1}{2}} \cdot \left[ b_0^T + \sum_{t=1}^{I} b_0^{(I-t)} \sum_{A_{t \leq 1}} \sum_{A_{t \mid 1}} \sum_{A_{t \uparrow \downarrow}} \sum_{A_{t \uparrow \leftarrow \downarrow}} \ldots \sum_{A_{t \uparrow \leftarrow \downarrow}} \left[ b_{\lambda_1} b_{\lambda_2} \ldots b_{\lambda_t} \cdot G_{\lambda_1}, \lambda_2, \ldots, \lambda_t \left( \sum_{i=1}^{t} 2A_i \right) \right] \right] \\
+ \sum_{n=1}^{\frac{I}{2}} \left[ 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2m-1) \right] \left( \frac{\sigma^2 + J \sigma^2}{IJ} \right)^{\frac{2\lambda}{2}} \sum_{t=2}^{I} \sum_{A_{t \leq 1}} \sum_{A_{t \mid 1}} \sum_{A_{t \uparrow \downarrow}} \sum_{A_{t \uparrow \leftarrow \downarrow}} \ldots \sum_{A_{t \uparrow \leftarrow \downarrow}} \left[ b_{\lambda_1} b_{\lambda_2} \ldots b_{\lambda_t} \cdot G_{\lambda_1}, \lambda_2, \ldots, \lambda_t \left( \sum_{i=1}^{t} 2A_i - 2m \right) \right] \right) \}
\]

Therefore we may write for the likelihood function \( L(\sigma^2, \sigma^2) \):

\[
L \propto (\sigma^2)^{-\frac{I(I+3)/2}{2}} \exp \left[ - \frac{SSW}{2\sigma^2} \right] (\sigma^2 + J \sigma^2)^{-\frac{(I-1)/2}{2}} \exp \left[ - \frac{SSB}{2(\sigma^2 + J \sigma^2)} \right] \cdot F
\]

where

\[
F = \left[ 2\pi \left( \frac{\sigma^2 + J \sigma^2}{IJ} \right) \right]^{-\frac{1}{2}} \cdot E
\]
3.3 THE NON-NORMALITY FACTOR

Examination of the non-normality factor, $E$, in the likelihood function $L(\sigma^2, \alpha^2)$ shows that information about the SSB has been concentrated in the first three terms. Therefore we drop the fourth and subsequent terms since they yield no information for our further analysis of non-normality. The long and complicated calculation and manipulations used in obtaining this result will be omitted for clarity.

After integrating $\mu$ out of $D$ and dropping terms for sufficiently large $J$, $E$ is approximated by:

$$E \approx \int \exp \left[ - \frac{IJ(\bar{y} - \mu)^2}{2(\sigma^2 + J\alpha^2)} \right] \left\{ b_0^2 + b_0^4 \sum_i c_i + b_0^2 \sum_{i<k<\ell} c_i c_{\ell} \right\} d\mu$$

$$= \frac{1}{2\pi} \left( \frac{\sigma^2 + J\alpha^2}{I J} \right)^{\frac{1}{2}} \left( 1 + \frac{\nu^2}{8} \right)^{-\frac{1}{2}} \cdot \left[ \left( 1 + \frac{\nu^2}{8} \right) + \left( 1 + \frac{\nu^2}{8} \right) \left( \frac{\nu^2}{4} \right) \left( \frac{1 - 2I}{2I} \right) + \left( \frac{\nu^2}{4} \right) \left( \frac{3(I - 1)}{2I} \right) \right]$$

$$+ \left( 1 + \frac{\nu^2}{8} \right) \left( \frac{\nu^2}{4} \right) \left( \frac{1 - I}{I} \right) + \left( \frac{\nu^2}{4} \right) \left( \frac{L - 3}{I} \right) \left( \frac{SSB}{\sigma^2 + J\alpha^2} \right)$$

$$+ \left( 1 + \frac{\nu^2}{8} \right) \left( \frac{\nu^2}{4} \right) \left( \frac{1}{6} \right) \left( \frac{SSB}{\sigma^2 + J\alpha^2} \right)^2 \right\}. \quad (3.19)$$

This may be written as the proportion:

$$E \propto (\sigma^2 + J\alpha^2)^{\frac{1}{2}} \left[ 1 + d_1 + d_2 \left( \frac{SSB}{\sigma^2 + J\alpha^2} \right) + d_3 \left( \frac{SSB}{\sigma^2 + J\alpha^2} \right)^2 \right], \quad (3.20)$$

where
\[ d_1 = \frac{\gamma_a}{64 I} \times (5Y_a I - 5Y_a + 8), \]

\[ d_2 = \frac{\gamma_a}{32 I} \times [\gamma_a (I - 5) - 8(I - 1)], \]

\[ d_3 = \frac{\gamma_a (Y_a + 8)}{192}. \]

We are then left with the factor:

\[
\left[ 1 + d_1 + d_2 \left( \frac{SSB}{\sigma^2 + J \alpha_b^2} \right) + d_3 \left( \frac{SSB}{\sigma^2 + J \alpha_b^2} \right)^2 \right]
\]

which contains the information about non-normality.

By using the above information, we may approximate the posterior density function \( f(\sigma^2, \alpha_b^2) \) as:

\[
f(\sigma^2, \alpha_b^2) \propto p(\sigma^2, \alpha_b^2) \times (\sigma^2)^{-\frac{1}{2}} \times \exp \left[ -\frac{SSB}{2\sigma^2} \right] \times \exp \left[ -\frac{SSB}{2(\sigma^2 + J \alpha_b^2)} \right] \times \left[ 1 + d_1 + d_2 \left( \frac{SSB}{\sigma^2 + J \alpha_b^2} \right) + d_3 \left( \frac{SSB}{\sigma^2 + J \alpha_b^2} \right)^2 \right].
\]

\[ (3.21) \]

For our further analysis, the above expression will be used as the posterior density function of \((\sigma^2, \alpha_b^2)\).
Chapter 4

THE POSTERIOR INFERENCE ON $\sigma^2$

The posterior density function of $(\sigma^2, \alpha^2)$ has now been cast into a form which displays the non-normality factor separately. This form is then used to obtain the posterior density functions of $\sigma^2$, $\alpha^2$, and the ratio $\alpha^2 / \sigma^2$, which will be dealt with in this and the following two chapters.

The marginal posterior density function of $\sigma^2$ which we wish to examine is obtained by integrating $\alpha^2$ out of $f(\sigma^2, \alpha^2)$, thus:

$$f(\sigma^2) = \int_0^{\infty} f(\sigma^2, \alpha^2) \, d\alpha^2$$

$$\propto (\sigma^2)^{-\frac{1+q}{2}} \exp \left[ - \frac{SSW}{2\sigma^2} \right] \cdot$$

$$\cdot \int_0^{\infty} p(\sigma^2, \alpha^2) \cdot (\sigma^2 + J \alpha^2)^{-\frac{1}{2}} \exp \left[ - \frac{SSB}{2(\sigma^2 + J \alpha^2)} \right] \cdot$$

$$\cdot \left[ 1 + c_1 + d_2 \left( \frac{SSB}{\sigma^2 + J \alpha^2} \right) + d_3 \left( \frac{SSB}{\sigma^2 + J \alpha^2} \right) \right] \, d\alpha^2.$$  (4.1)

To begin the analysis let us consider the factor of $f(\sigma^2)$

$$p'(\sigma^2) \equiv \int_0^{\infty} p(\sigma^2, \alpha^2) \cdot (\sigma^2 + J \alpha^2)^{-\frac{1}{2}} \exp \left[ - \frac{SSB}{2(\sigma^2 + J \alpha^2)} \right] \cdot$$

$$\cdot \left[ 1 + c_1 + d_2 \left( \frac{SSB}{\sigma^2 + J \alpha^2} \right) + d_3 \left( \frac{SSB}{\sigma^2 + J \alpha^2} \right) \right] \, d\alpha^2.$$  (4.2)

as a partial posterior distribution based only on the data $\mathbf{y}$;
and consider the other factor of $f(\sigma^2)$

$$L(\sigma^2 | SSW) \equiv (\sigma^2) \exp \left\{ -\frac{SSW}{2\sigma^2} \right\} \quad (4.3)$$

as a likelihood function based only on the deviations $(\bar{y}_i - \bar{y}_.)$. The overall marginal posterior density may thus be written:

$$f(\sigma^2) \propto p'(\sigma^2) \cdot L(\sigma^2 | SSW).$$

We will use three different priors for $p'(\sigma^2)$. First we use a stable estimation prior; second we assume the independence of $\sigma^2$ and $\sigma^2\alpha$; and third we will use the non-informative prior used by Tiao and Tan.

4.1 ASSUMPTION OF A STABLE ESTIMATION PRIOR

Inference on $\sigma^2$ is most sensitive to the factor $p'(\sigma^2)$. This sensitivity has been best reported in the work of Hill, from whose paper [15] the following three paragraphs are quoted.

When $p'(\sigma^2)$ can be regarded as gentle relative to $L(\sigma^2 | SSW)$ the principle of stable measurement applies and $f(\sigma^2)$ will be nearly proportional to $L(\sigma^2 | SSW)$.

For example, if $p'(\sigma^2) \propto \sigma^{-2}$ (strictly speaking this is impossible, but when stable estimation occurs, the exact choice of $p'(\sigma^2)$ is immaterial), then:

$$f(\sigma^2) \propto (\sigma^2)^{-(J-1)/2} \exp \left\{ -\frac{SSW}{2\sigma^2} \right\}$$

so that $\sigma^2$ is distributed approximately as $SSW/\chi^2_{J-1}$, where $\chi^2_{V}$ is a random variable having chi-square distribution with $V$ degrees of freedom.
Parenthetically we note that any information from non-normality is not used at all. This result is in harmony with the conventional approach under which SSW/\sigma_\omega^2 has the chi-square distribution with \( I(\J-1) \) degrees of freedom for given \( \sigma_\omega^2 \).

From a Bayesian viewpoint, however, this "stable measurement posterior" is merely one (rather simple) posterior distribution that may arise, and we can not regard it as having any innate significance, except, at best, as a reasonable approximation in some circumstances. Indeed, the factor \( p'(\sigma_\omega^2) \) may not be gentle relative to \( L(\sigma_\omega^2 | \text{SSW}) \) even though sample sizes are large and \( L(\sigma_\omega^2 | \text{SSW}) \) is quite sharp and we must be extremely cautious in using the stable measurement approximation.

4.2 ASSUMPTION OF INDEPENDENCE OF \( \sigma_\omega^2 \) AND \( \sigma_\alpha^2 \)

We shall now assume that \( \sigma_\omega^2 \) and \( \sigma_\alpha^2 \) are independent. Then

\[
p(\sigma_\omega^2, \sigma_\alpha^2) = p_1(\sigma_\omega^2) \cdot p_2(\sigma_\alpha^2).
\]

This assumption is justified if we consider that \( \sigma_\alpha^2 \) is the variation of true means between groups and \( \sigma_\omega^2 \) is the variation of the measurement error of each observation. We then have, under this assumption:

\[
f(\sigma_\omega^2) = \int_0^\infty f(\sigma_\omega^2, \sigma_\alpha^2) d\sigma_\alpha^2
\]

\[
\propto p_1(\sigma_\omega^2) \cdot L(\sigma_\omega^2 | \text{SSW}) \cdot 
\int_0^\infty p_2(\sigma_\alpha^2) \cdot \left(\frac{\sigma_\omega^2 + \J \sigma_\alpha^2}{\sigma_\omega^2 + \J \sigma_\alpha^2}\right)^{-\frac{\J-1}{2}} \exp \left[-\frac{\text{SSB}}{2(\sigma_\omega^2 + \J \sigma_\alpha^2)}\right] \cdot 
\cdot \left[1 + d_1 + d_2 \left(\frac{\text{SSB}}{\sigma_\omega^2 + \J \sigma_\alpha^2}\right) + d_3 \left(\frac{\text{SSB}}{\sigma_\omega^2 + \J \sigma_\alpha^2}\right)^2\right] d\sigma_\alpha^2. \tag{4.5}
\]
If we use

$$p_1(\sigma_1^2) \propto (\sigma_1^2)^{-(\lambda_1/2)+1} \exp \left[ - \frac{c_1}{2\sigma_1^2} \right]$$

(4.6)

so that \( (\sigma_1^2)^{-1} \) has a gamma distribution with parameters \( \lambda_1/2 \) and \( c_1/2 \) chosen subjectively, then:

$$f(\sigma^2) \propto p_1(\sigma^2) \cdot \mathbb{L}(\sigma^2 | \operatorname{SSB}) \int_0^\infty (\sigma_1^2)^{-(\lambda_1/2)} \exp \left[ - \frac{c_1}{2\sigma_1^2} \right] \cdot (h^2)^{-(I-1/2)} \exp \left[ - \frac{SSB}{2h^2} \right] \cdot \left[ 1 + d_1 + d_2 \left( \frac{SSB}{h^2} \right) + d_3 \left( \frac{SSB}{h^2} \right)^2 \right] d\sigma_1^2,$$

where

$$h^2 = \sigma^2 + J\sigma_1^2.$$

If we denote the last integral by \( T \), then:

$$T = \int_0^\infty (\sigma_1^2)^{-(\lambda_1/2)-1} \exp \left[ - \frac{c_1}{2\sigma_1^2} \right] \cdot (h^2)^{-(I-1/2)} \exp \left[ - \frac{SSB}{2h^2} \right] \cdot \left[ 1 + d_1 + d_2 \left( \frac{SSB}{h^2} \right) + d_3 \left( \frac{SSB}{h^2} \right)^2 \right] d\sigma_1^2$$

(4.7)

$$= (1+d_1) \int_0^\infty (\sigma_1^2)^{-(\lambda_1/2)-1} \exp \left[ - \frac{c_1}{2\sigma_1^2} \right] \cdot (h^2)^{-(I-1/2)} \exp \left[ - \frac{SSB}{2h^2} \right] d\sigma_1^2$$

$$+ d_2(\operatorname{SSB}) \int_0^\infty (\sigma_1^2)^{-(\lambda_1/2)-1} \exp \left[ - \frac{c_1}{2\sigma_1^2} \right] \cdot (h^2)^{-(I-1/2)} \exp \left[ - \frac{SSB}{2h^2} \right] d\sigma_1^2$$

$$+ d_3(\operatorname{SSB})^2 \int_0^\infty (\sigma_1^2)^{-(\lambda_1/2)-1} \exp \left[ - \frac{c_1}{2\sigma_1^2} \right] \cdot (h^2)^{-(I-1/2)} \exp \left[ - \frac{SSB}{2h^2} \right] d\sigma_1^2.$$

(4.8)

If we look at the first integral of the last expression in more detail, we see that:
Here we have substituted $y = J\sigma^2$.

Thus the above integral may be considered, at least on the positive half-line, as being proportional to a density function of two independent random variables

$$Z = X - Y = \left\{ \left( \left[ \frac{I-3}{2}, \frac{SSB}{2} \right] \right)^{-1} \right\} - \left( \left[ \frac{\lambda a}{2}, \frac{JCa}{2} \right] \right)^{-1},$$

where $X$ and $Y$ are independent random variables having inverted gamma density functions.

$$f(t) \propto t^{\lambda-1} e^{-\frac{\lambda t}{2}}, \quad t \geq 0,$$

Similarly, the second integral, $I_2$, and the third integral, $I_3$, may be regarded as proportional to the density function of a random variable

$$Z = X - Y = \left( \left[ \frac{I-1}{2}, \frac{SSB}{2} \right] \right)^{-1} - \left( \left[ \frac{\lambda a}{2}, \frac{JCa}{2} \right] \right)^{-1},$$

and

$$Z = X - Y = \left( \left[ \frac{I+1}{2}, \frac{SSB}{2} \right] \right)^{-1} - \left( \left[ \frac{\lambda a}{2}, \frac{JCa}{2} \right] \right)^{-1},$$

respectively.

Then the over-all posterior distribution of $\sigma^2$ may be regarded as a weighting of $p_1(\sigma^2) \cdot L(\sigma^2 \mid SSW)$ by $T$, which is a linear combination of density functions, each of which is the density function of the difference of independent inverted gamma variables.
Various situations may now arise. Thus, $\nu_2$ (the parameter of kurtosis) may be 0; the SSB may be 0; $\nu_2$ may be negative; or $\nu_2$ may be positive. These situations we shall now review.

4.2.1 The Parameter Of Kurtosis Is Zero

In the first case, if $\nu_2$ (the parameter of kurtosis) takes on the value 0, which means that $\nu_i \sim N(0, \sigma^2)$, then all the $d_1 = d_2 = d_3 = 0$. We are then left with only the first integral, namely:

$$f(\sigma^2) \propto p_1(\sigma^2) \cdot L(\sigma^2 \mid \text{SSW}) \cdot \int_{0}^{\infty} (\sigma_\nu^2)^{-\nu_2} \exp \left[ -\frac{c_\nu}{2c_\nu^2} (\sigma^2 + Jc_\nu^2) \right] \exp \left[ -\frac{\text{SSB}}{2(\sigma^2 + Jc_\nu^2)} \right] d\sigma_\nu^2$$

$$= p_1(\sigma^2) \cdot L(\sigma^2 \mid \text{SSW}) \cdot \left[ \left[ \frac{I-3}{2}, \frac{\text{SSB}}{2} \right]^{-1} - \left[ \left[ \frac{\lambda_\nu}{2}, \frac{Jc_\nu}{2} \right]^{-1} \right] \right].$$

Therefore the posterior distribution of $\sigma^2$ may be regarded as a weighting of $p_1(\sigma^2) \cdot L(\sigma^2 \mid \text{SSW})$ by the density of

$$Z = \left[ \left[ \frac{I-3}{2}, \frac{\text{SSB}}{2} \right]^{-1} - \left[ \left[ \frac{\lambda_\nu}{2}, \frac{Jc_\nu}{2} \right]^{-1} \right] \right],$$

the resulting function being normalized to have unit area on the positive half-line.

At this juncture it is interesting to find the mean and the mode of $Z$. The mean is found directly from the inverted gamma variates;

$$\int \left[ \frac{I-3}{2}, \frac{\text{SSB}}{2} \right] \text{ and } \int \left[ \frac{\lambda_\nu}{2}, \frac{Jc_\nu}{2} \right], \text{ thus; }$$
\[ E(2) = E(X - Y) \]
\[ = \mathbb{E}\left\{ \left( \left[ \frac{I-3}{2}, \frac{SSB}{2} \right] \right)^{-1} \right\} - \mathbb{E}\left\{ \left( \left[ \frac{\lambda a}{2}, \frac{J C_a}{2} \right] \right)^{-1} \right\} \]
\[ = \frac{SSB/2}{I-3} - \frac{J C_a/2}{\lambda a - 1} \]
\[ = \frac{SSB}{I - 5} - \frac{J C_a}{\lambda a - 2}. \quad (4.11) \]

However, the mode is found only with some difficulty. The expression of the integral, \( I_1 \), does suggest that, when \( \lambda a/2 - 1 \) is large, we can write:

\[ I_1 = \int_0^\infty (\sigma_a^2 + J \sigma_a^2) \frac{SSB}{2(\sigma_a^2 + J \sigma_a^2)} \exp\left[ -\frac{SSB}{2(\sigma_a^2 + J \sigma_a^2)} \right] \left( \sigma_a^2 \right)^{\lambda a/2 - 1} \exp\left[ -\frac{C_a}{2\sigma_a^2} \right] d\sigma_a^2 \]
\[ \approx (\sigma_a^2 + J \bar{\sigma}_a^2) \frac{SSB}{2(\sigma_a^2 + J \bar{\sigma}_a^2)} \int_0^\infty (\sigma_a^2)^{\lambda a/2 - 1} \exp\left[ -\frac{C_a}{2\sigma_a^2} \right] d\sigma_a^2 \]
\[ = (\sigma_a^2 + J \bar{\sigma}_a^2) \frac{SSB}{2(\sigma_a^2 + J \bar{\sigma}_a^2)} \left( \frac{C_a}{2} \right)^{\lambda a/2} \left[ \frac{\lambda a}{2} \right] \]
\[ \propto (\sigma_a^2 + J \bar{\sigma}_a^2) \frac{SSB}{2(\sigma_a^2 + J \bar{\sigma}_a^2)} \left( \frac{C_a}{2} \right)^{\lambda a/2} \left[ \frac{\lambda a}{2} \right], \quad (4.12) \]

where \( \bar{\sigma}_a^2 \) is the value of \( \sigma_a^2 \) at which \( (\sigma_a^2)^{\lambda a/2 - 1} \exp\left[ -\frac{C_a}{2\sigma_a^2} \right] \) has its maximum value. Also:

\[ \ln p_2(\sigma_a^2) = 0, \quad \text{together with} \quad \frac{\partial \ln p_2(\sigma_a^2)}{\partial \sigma_a^2} = 0 \]

gives us

\[ \bar{\sigma}_a^2 = \frac{c_a}{\lambda a + 2}. \]
Therefore
\[ I_1 \propto \left( \sigma^2 + \frac{J_C a}{\lambda_a + 2} \right)^{-(I-1)/2} \exp \left[ -\frac{\text{SSB}}{2(\sigma^2 + \frac{J_C a}{\lambda_a + 2})} \right]. \]

As an aside, we note that the above shows that the difference of independent inverted gamma variates is approximately distributed as an inverted gamma variate in the interval
\[ \sigma^2 \gg -\frac{J_C a}{\lambda_a + 2}, \]
and \( \ln I_1 = 0 \), together with \( \frac{\partial \ln I_1}{\partial \sigma^2} = 0 \) gives us
\[ \bar{\sigma}^2 = \text{MSB} - \frac{J_C a}{\lambda_a + 2}. \] (4.13)

Therefore the mode of \( Z \) is approximately \( \text{MSB} - \frac{J_C a}{\lambda_a + 2} \).

As we see from the mean and the mode, \( Z \) can clearly be negative.
But only the portion of its density for \( \sigma^2 \gg 0 \) is relevant to the over-all posterior distribution of \( \sigma^2 \).

We have now located the mean of \( Z \) at \( \frac{\text{SSB}}{I-5} = \frac{J_C a}{\lambda_a + 2} \) and the mode of \( Z \) roughly at \( \frac{\text{SSB}}{I-1} - \frac{J_C a}{\lambda_a + 2} = \text{MSB} - \frac{J_C a}{\lambda_a + 2} \), which two quantities are measures of the central tendency of the distribution of \( Z \). The mode of \( Z \) plays an especially important role here. These roles are discussed by Hill [15] in the following words:

When \( \text{MSB} - \frac{J_C a}{(\lambda_a + 2)} \) < 0, then the density function of \( Z \)
will be monotonically decreasing for \( \sigma^2 \gg 0 \) and the effect
of the information supplied by the $\bar{y}_i$ will be to modify the partial posterior $p_i(\sigma^2) \cdot L(\sigma^2 | SSW)$ based upon $y_{ij} - \bar{y}_i$ alone by giving the greatest weight to small $\sigma^2$ and thus tending to move opinions downward.

When $\text{MSB} - \frac{J C_\alpha}{(\lambda_\alpha + 2)}$ is positive, then the density of $Z$ will have only a portion of its left tail below zero and there is a possibility of obtaining an extremely sharp posterior $f(\sigma^2)$.

4.2.2 The SSB Is Zero

In the second case we assume that $SSB = 0$ (although to do so is very unrealistic) in order to sec the effect for very small $SSB$.

The expression $T$ is rewritten as:

$$T \propto \int_0^\infty (\sigma_0^2)^{-\frac{1}{2}} \cdot \exp \left[ -\frac{C_\alpha}{2\sigma_0^2} \right] \cdot \frac{SSB}{2h^2} \cdot \frac{SSB}{2h^2} \cdot d\sigma_0^2$$

$$+ \frac{d_2SSB}{1 + d_1} \int_0^\infty (\sigma_0^2)^{-\frac{1}{2}} \cdot \exp \left[ -\frac{C_\alpha}{2\sigma_0^2} \right] \cdot \frac{SSB}{2h^2} \cdot d\sigma_0^2$$

For $SSB = 0$, the last two terms will be zero, so that:

$$T \propto \int_0^\infty (\sigma_0^2)^{-\frac{1}{2}} \cdot \exp \left[ -\frac{C_\alpha}{2\sigma_0^2} \right] \cdot \frac{SSB}{2h^2} \cdot d\sigma_0^2$$

Again, letting $SSB = 0$ in the exponent yields:

$$T \propto \int_0^\infty (\sigma_0^2)^{-\frac{1}{2}} \cdot \exp \left[ -\frac{C_\alpha}{2\sigma_0^2} \right] \cdot \left(\frac{SSB}{2h^2}\right)^{1/2} \cdot d\sigma_0^2 \cdot (4.14)$$

To integrate, we use the same approximation as before:
The above expression shows that the general picture of \( T \) will be:

\[
T \propto \left( q^2 + \frac{J C_\alpha}{\lambda_\alpha + 2} \right) \left( \frac{C_\alpha}{2} \right)^{-\frac{\lambda_\alpha}{2}}
\]

\[
\propto \left( q^2 + \frac{J C_\alpha}{\lambda_\alpha + 2} \right)^{-\frac{\lambda_\alpha}{2}}
\]

(4.15)

Therefore the effect of \( T \) on the posterior density, \( f(\sigma^2) \), again depends upon the prior density of \( \sigma^\alpha \), reflected by the parameters \( C_\alpha \) and \( \lambda_\alpha \), and the posterior experiment \( J \), which is the sample size in each group.

If \( SSB \) is small or zero, then the larger is \( J \), the flatter \( T \) becomes, and, in turn, the posterior density \( f(\sigma^2) \) is proportional to \( p_1(\sigma^2) \, L(\sigma^2 \mid SSW) \), which is free from \( \sigma^\alpha \).

\[
f(\sigma^2) \propto p_1(\sigma^2) \, L(\sigma^2 \mid SSW)
\]

\[
\propto p_1(\sigma^2) \cdot (\sigma^2)^{-\frac{1}{2}(J-1)} \exp \left[ -\frac{SSW}{2\sigma^2} \right]
\]

(4.16)
This contains, however, the factor $p_1(\sigma^2)$ in addition to the result under the stable estimator prior in (4.4). On the other hand, if $N_\alpha$ is small, or almost zero, then we have

$$p_1(\sigma^2) \cdot (\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{SSB}{2\sigma^2}\right] \cdot T,$$

where $T$ gives the weight on the smaller side of $\sigma^2$.

4.2.3 The Parameter of Kurtosis is Negative

When we express $T$ as a sum of independent inverted gamma variates we obtain

$$T \propto (1 + d_1) \cdot f \left[ \left( \left[ \frac{I-3}{2} \right], \frac{SSB}{2} \right)^{-1} - \left( \left[ \frac{\lambda_0}{2} \right], \frac{J_\alpha}{2} \right)^{-1} \right]$$

$$+ d_2(I-3) \cdot f \left[ \left( \left[ \frac{I-1}{2} \right], \frac{SSB}{2} \right)^{-1} - \left( \left[ \frac{\lambda_0}{2} \right], \frac{J_\alpha}{2} \right)^{-1} \right]$$

$$+ d_3(I-3)(I-1) \cdot f \left[ \left( \left[ \frac{I+1}{2} \right], \frac{SSB}{2} \right)^{-1} - \left( \left[ \frac{\lambda_0}{2} \right], \frac{J_\alpha}{2} \right)^{-1} \right]. \quad (4.17)$$

If we now assume a specific value for the parameter of kurtosis, say $\gamma_2 = -1.2$, which means the $a_i$ are uniform random variates, then we naturally expect a much smaller SSB than in the normal random variate case, since the $a_i$ are more evenly distributed over the shorter interval. This means that we are very likely to have a negative mode.
If we now look at the coefficients of the three terms, which depend upon the value of \( \gamma_2 \) and the number of groups \( I \); we find that for large \( I \) the coefficient of the third term is found to dominate the other two. We may then write

\[
T \propto f \left[ \left( \frac{I+1}{2}, \frac{SSB}{2} \right)^{-1} - \left( \frac{\lambda_a}{2}, \frac{JC_A}{2} \right)^{-1} \right] + c_1 \cdot f \left[ \left( \frac{I-1}{2}, \frac{SSB}{2} \right)^{-1} - \left( \frac{\lambda_a}{2}, \frac{JC_A}{2} \right)^{-1} \right] + c_2 \cdot f \left[ \left( \frac{I-3}{2}, \frac{SSB}{2} \right)^{-1} - \left( \frac{\lambda_a}{2}, \frac{JC_A}{2} \right)^{-1} \right] \quad (4.17.1)
\]

where:

\[
e_1 = \frac{d_1}{d_2}(I-1), \quad \text{and}
\]

\[
e_2 = \frac{(1+d_1)}{d_2}(I-3)(I-1).
\]

If \( I > 10 \), then \( e_1 \) and \( e_2 \) tend to zero, and we may well ignore those two terms to leave only the first term approximation

\[
f\left\{ \left( \frac{I+1}{2}, \frac{SSB}{2} \right)^{-1} - \left( \frac{\lambda_a}{2}, \frac{JC_A}{2} \right)^{-1} \right\}.
\]

In this case \( T \) will be approximately proportional to the density function of the difference of the independent inverted gamma variates, whose mean is

\[
\frac{JC_A}{MSB} = \frac{JC_A}{\lambda_a + 2}
\]

and mode is

\[
\frac{SSB}{I+3} = \frac{JC_A}{\lambda_a + 2}.
\]
The overall picture shown below indicates that the information supplied from the SSB is very flat unless the mode $\frac{JCa}{\lambda a + 2}$ is near to zero. In the latter case, it will give more weight on the smaller side of $\sigma^2$.

\[ \frac{SSB}{I+3} = \frac{JCa}{\lambda a + 2} \]

4.2.4 The Parameter of Kurtosis is Positive

If we now assume a positive value for $\gamma_2$, the parameter of kurtosis, say $\gamma_2 = 3$, we have that the $a_i$ are double exponential random variates. We then naturally expect a greater SSB, since the $a_i$ are more concentrated around the mean and have heavier tails. The mode will then very likely be positive. Actually the mode also depends upon $I$ and $J$ to an extent.
In this case, since we have found that the coefficient of the third term of (4.17) is again dominant over the other two for large I, we may write in this case as before

$$ T \propto r \left\{ \left( \Gamma \left[ \frac{I+1}{2}, \frac{SSB}{2} \right] \right)^{-1} - \left( \Gamma \left[ \frac{\lambda a}{2}, \frac{JC}{2} \right] \right)^{-1} \right\}. $$

The overall picture is shown below and this time we will have a good possibility of obtaining an extremely sharp posterior $f(\sigma^2)$. In other words, the information supplied from SSB is quite valuable, since it had mode and mean both on the positive side of $\sigma^2$.

To summarize the foregoing we may say that for $\gamma_2 = -1.2$, and $\gamma_2 = 3$, our choice of value affects the posterior density function, $f(\sigma^2)$, by locating its mode on the negative or on the positive side of $\sigma^2$. The mode is approximately

$$ MSB - \frac{J C a}{\lambda a + 2} $$

and the mean

$$ SSB - \frac{J C a}{I+3, \lambda a + 2}. $$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
4.3 ASSUMPTION OF NON-INFORMATIVE PRIOR

We shall consider one other form of non-informative prior:

\[ p(\sigma^2, \sigma_\alpha^2) \propto (\sigma^2)^{-1} \cdot (\sigma^2 + J\sigma_\alpha^2)^{-1}, \]  

for \( \sigma^2 > 0, \sigma_\alpha^2 > 0 \), (4.18)

which is also diffuse, or gentle. Tiao and Tan [14] obtained this prior by the method of Jeffreys and called it Jeffreys' non-informative prior. Although the \( J \) in the non-informative prior depends upon the sample size in the experiment and would thus seem inappropriate, this prior is used when the likelihood function is dominant over the prior; in such case the exact choice of prior makes no difference.

We find, for this prior:

\[
\begin{align*}
\log f(\sigma^2) &\propto \int_0^\infty (\sigma^2)^{-1} (\sigma^2 + J\sigma_\alpha^2)^{-1} L(\sigma^2 | SSW)(h^2)^{-\frac{1}{2} h^2} \exp \left[ -\frac{SSB}{2h^2} \right] \, dh^2 \\
&\propto (\sigma^2)^{-1} L(\sigma^2 | SSW) \int_0^\infty (h^2)^{-\frac{1}{2} h^2} \exp \left[ -\frac{SSB}{2h^2} \right] \, dh^2 \\
&\propto (\sigma^2)^{-1} L(\sigma^2 | SSW) \int_0^\infty (h^2)^{-\frac{1}{2} h^2} \exp \left[ -\frac{SSB}{2h^2} \right] \, dh^2 \cdot \left[ 1 + d_1 + d_2 \left( \frac{SSB}{h^2} \right) + d_3 \left( \frac{SSB}{h^2} \right)^2 \right] \, dh^2.
\end{align*}
\]

Changing the variable and \( d\sigma_\alpha^2 = \frac{1}{J} \, dh^2 \), we have:

\[
h = \sigma^2 + J\sigma_\alpha^2, \quad \text{and} \quad d\sigma_\alpha^2 = \frac{1}{J} \, dh^2,
\]

we have:
Here, if we have $\gamma_2 = 0$, the above posterior density is reduced to:

$$f(\sigma^2) \propto (\sigma^2)^{-1} L(\sigma^2 \mid \text{SSW}) \int_0^\infty (\mu^2)^{-(L-1)/2} \exp \left[ -\frac{\text{SSB}}{2\mu^2} \right] d\mu.$$ 

Again, changing the variable under the integral sign to $u = \frac{\text{SSB}}{h^2}$, we have:

$$f(\sigma^2) \propto (\sigma^2)^{-1} L(\sigma^2 \mid \text{SSW}) \int_0^\infty u^{-(L-1)/2} \exp \left[ -\frac{u}{2} \right] du \propto (\sigma^2)^{-1} L(\sigma^2 \mid \text{SSW}) \cdot F_{\chi^2_{L-1}} \left( \frac{\text{SSB}}{\sigma^2} \right),$$

where $F_{\chi^2_{L-1}}(\cdot)$ is the cumulative distribution function of a chi-square random variable with $n$ degrees of freedom.

Comparing this posterior with the earlier stable estimation prior of section 4.1, we notice that the use of $(\sigma^2)^{-1} (\sigma^2 + J\sigma^2 \gamma_2)^{-1}$ as Jeffreys non-informative prior results only in the addition of the factor $F_{\chi^2_{L-1}}(\text{SSB}/\sigma^2)$.

Then the overall posterior function may be considered as a weighting of $(\sigma^2)^{-2} L(\sigma^2 \mid \text{SSW})$, which is also posterior under stable estimating prior, by $F_{\chi^2_{L-1}}(\text{SSB}/\sigma^2)$, which is monotonically decreasing and has its inflexion point at $\text{SSB}/(L+1)$. Therefore the effects of $F_{\chi^2_{L-1}}(\text{SSB}/\sigma^2)$ is to give greater weight to smaller $\sigma^2$. For quite large SSB, the function is very slowly decreasing, and the inflexion point, $SSB/(L+1)$, is located very far to the right, which means that the effect of $F_{\chi^2_{L-1}}(\text{SSB}/\sigma^2)$ on overall posterior function
is not serious. On the other hand, for quite small SSB, the function is very rapidly decreasing and SSB/(I+1) is near to zero, which implies that \( \sigma^2 \) is quite small.

To examine the extreme case, let us assume that SSB = 0.

Then the integral in the expression (4.19) will readily reduce to:

\[
\int_0^\infty (\sigma^2 + J \sigma_X^2)^{-[(I-1)/2]} \, d\sigma_X^2 \propto (\sigma^2)^{-(I-1)/2}.
\]

In this case only the fact that SSB = 0 is used. Information regarding non-normality is not used at all.

The posterior distribution is then:

\[
f(\sigma^2) \propto (\sigma^2)^{-[(I-1)/2]} \cdot L(\sigma^2 | SSW) (\sigma^2)^{-[(I-1)/2]} \\
\propto (\sigma^2)^{\frac{I-1}{2}} \exp \left[ - \frac{SSW}{2 \sigma^2} \right],
\]

which means that a posteriori,

\[
\sigma^2 \sim \frac{SSW}{\chi^2_{I-1}}
\]

as was noted by Tiao and Tan.

As pointed out by Hill [15], this is a disturbing result.

For if I is large, then an SSB near zero would, based upon the \( \bar{y}_1 \) alone, lead to a strong opinion that \( \sigma^2 \) is small. On the other hand, a large SSW would, based upon the \( y_j - \bar{y}_1 \) alone, lead to a strong opinion that \( \sigma^2 \) is large. These two separate sources of information can thus lead to very different opinions about \( \sigma^2 \). Surely, in the extreme case, where SSB is very small and SSW is very large, one is reluctant to accept \( \sigma^2 \sim \frac{SSW}{\chi^2_{I-1}} \) as describing the over-all posteriori opinion about \( \sigma^2 \).
It should be pointed out that it is not simply this result (which after all, was based upon an improper prior) that is in question, but really any way of combining the two divergent sources of information, if we keep to the original model.

In other situations, such as $\gamma_2 \neq 0$, and $\mathrm{SSB} \neq 0$, we have from (4.20)

$$f(\omega^2) \propto (\omega^2)^{-\frac{1}{2}} L(\omega^2 | \mathrm{SSB}) \int_{-\infty}^{\infty} \frac{(I-I^\top)}{\omega^2} \exp \left[ -\frac{\mathrm{SSB}}{2\omega^2} \right] \cdot \left[ 1 + d_1 + d_2 \left( \frac{\mathrm{SSB}}{\omega^2} \right) + d_3 \left( \frac{\mathrm{SSB}}{\omega^2} \right)^2 \right] d\omega^2. \quad (4.24)$$

Again changing the variable under the integral sign to $u = \frac{\mathrm{SSB}}{\omega^2}$:

$$f(\omega^2) \propto (\omega^2)^{-\frac{1}{2}} L(\omega^2 | \mathrm{SSB}) \left\{ (1+d_1) \int_0^{\infty} \frac{u}{\mathrm{SSB}} \left( \frac{\mathrm{SSB}}{u^2} \right)^{-\frac{1}{2}} \exp \left[ -\frac{u}{2} \left( \frac{\mathrm{SSB}}{u^2} \right) \right] du - d_2(\mathrm{SSB}) \int_0^{\infty} \left( \frac{u}{\mathrm{SSB}} \right)^{\frac{1}{2}+t_1} \exp \left[ -\frac{u}{2} \left( \frac{\mathrm{SSB}}{u^2} \right) \right] du + d_3(\mathrm{SSB}) \int_0^{\infty} \left( \frac{u}{\mathrm{SSB}} \right)^{\frac{3}{2}+t_2} \exp \left[ -\frac{u}{2} \left( \frac{\mathrm{SSB}}{u^2} \right) \right] du \right\}$$

$$\propto (\omega^2)^{-\frac{1}{2}} L(\omega^2 | \mathrm{SSB}) \left\{ (1+d_1) \int_0^{u_0} \left( \frac{\mathrm{SSB}}{u^2} \right)^{-\frac{1}{2}} \exp \left[ -\frac{u}{2} \right] du - d_2 \int_0^{u_0} \left( \frac{\mathrm{SSB}}{u^2} \right)^{-\frac{1}{2}} \exp \left[ -\frac{u}{2} \right] du + d_3 \int_0^{u_0} \left( \frac{\mathrm{SSB}}{u^2} \right)^{-\frac{1}{2}} \exp \left[ -\frac{u}{2} \right] du \right\} \quad (4.25)$$
where $F_n(t)$ is the cumulative distribution function of a
chi-square random variable with $n$ degrees of freedom, as before.

For large $I$, the coefficient of the third term again dominates the other two terms, and we therefore have

$$f(\sigma^2) \propto (\sigma^2)^{-I} \cdot L(\sigma^2 | SSW) \cdot F_{(I-1)} \left( \frac{SSB}{\sigma^2} \right)$$

which is only slightly different from the normal case, in that the degrees of freedom become $I + 3$ instead of $I - 1$.

With these observations we end the discussion of the posterior density function of $\sigma^2$ and turn to that of $\alpha^2$. 

- 0 -
Chapter 5

THE POSTERIOR INFERENCE ON $\sigma_a^2$

The marginal posterior density function of $\sigma_a^2$ will now be studied by methods similar to those used in the last chapter. That is, the $\sigma^2$ is integrated out of $f(\sigma^2, \sigma_a^2)$ which, for convenience of integration, is separated into the form:

$$p(\sigma^2, \sigma_a^2) = p(\sigma^2) \cdot p(\sigma_a^2).$$

Various types of assumptions on the prior are considered together with cases of some specific values of the parameters.

5.1 THE FORM OF THE EQUATION

The posterior density for $\sigma_a^2$ is:

$$f(\sigma_a^2) \propto \int_0^\infty p(\sigma^2, \sigma_a^2) \cdot (\sigma^2)^{-\frac{1}{2}(I-1)/2} \cdot (\sigma^2 + J \sigma_a^2)^{-\frac{1}{2}(I-1)/2} \cdot \exp \left[ - \frac{SSW}{2\sigma^2} \right] \cdot \exp \left[ - \frac{SSB}{2(\sigma^2 + J \sigma_a^2)} \right] \cdot \left[ (1 + d_1) + \frac{SSB}{\sigma^2 + J \sigma_a^2} \right] \cdot \left[ \frac{SSB}{\sigma^2 + J \sigma_a^2} \right] \cdot \sigma_a^2 \cdot \frac{SSW}{2h^2} \cdot \left[ (\sigma_a^2)^{-\frac{1}{2}(I-1)/2} \cdot \exp \left[ - \frac{SSB}{2h^2} \right] \cdot \left[ (1 + d_1) + \frac{SSB}{h^2} \right] + d_1 \left( \frac{SSB}{h^2} \right)^2 \right] \cdot \sigma_a^2.$$

(5.1)
As before, we first take

\[ p(\sigma^2, \sigma^2_\alpha) = p_1(\sigma^2) \cdot p_2(\sigma^2_\alpha) \]

so that

\[
\begin{align*}
 f(\sigma^2_\alpha) & \propto p_2(\sigma^2_\alpha) \int_0^\infty p_1(\sigma^2) \cdot L(\sigma^2 \mid \text{SSW}) \cdot (h^2) \cdot \exp \left[ -\frac{\text{SSB}}{2h^2} \right] \cdot \\
 & \quad \cdot \left[ (1 + d_1) + d_2 \left( \frac{\text{SSB}}{h^2} \right) + d_3 \left( \frac{\text{SSB}}{h^2} \right)^2 \right] \, d\sigma^2_\alpha .
\end{align*}
\]

(5.2)

We shall now consider the results of using various forms of \( p_1(\sigma^2) \). Thus, if we use \( p_1(\sigma^2) \) subjectively chosen:

\[ p_1(\sigma^2) \propto (\sigma^2)^{-\lambda/2-1} \exp \left[ -\frac{C}{2\sigma^2} \right] . \]

(5.3)

Then the marginal posterior density function becomes:

\[
\begin{align*}
 f(\sigma^2_\alpha) & \propto p_2(\sigma^2_\alpha) \int_0^\infty p_1(\sigma^2) \cdot L(\sigma^2 \mid \text{SSW}) \cdot (h^2) \cdot \exp \left[ -\frac{\text{SSB}}{2h^2} \right] \cdot \\
 & \quad \cdot \left[ (1 + d_1) + d_2 \left( \frac{\text{SSB}}{h^2} \right) + d_3 \left( \frac{\text{SSB}}{h^2} \right)^2 \right] \, d\sigma^2_\alpha ,
\end{align*}
\]

(5.4)

and the marginal likelihood function is denoted:

\[
\begin{align*}
p'(\sigma^2_\alpha) & = \int_0^\infty (\sigma^2)^{-\lambda/2-1} \exp \left[ -\frac{C}{2\sigma^2} \right] \cdot (\sigma^2)^{-\lambda/2} \exp \left[ -\frac{\text{SSW}}{2\sigma^2} \right] \cdot \\
 & \quad \cdot (h^2)^{-\lambda/2} \exp \left[ -\frac{\text{SSB}}{2h^2} \right] \cdot \left[ (1 + d_1) + d_2 \left( \frac{\text{SSB}}{h^2} \right) + d_3 \left( \frac{\text{SSB}}{h^2} \right)^2 \right] \, d\sigma^2_\alpha .
\end{align*}
\]

(5.5)
If we put \( C = k \lambda \), then \( \lim_{\lambda \to 0} p_1(\sigma^{-2}) = (\sigma^{-1})^{-1} \), and we may represent diffuse prior opinion about \( \sigma^{-2} \) by a choice of \( \lambda \) near zero. By taking this diffuse prior we will have the posterior density function of \( \sigma^{-2} \):

\[
f(\sigma^{-2}) \propto p_2(\sigma^{-2}) \cdot p'(\sigma^{-2})
\]

where

\[
p'(\sigma^{-2}) \propto \int_0^\infty (\sigma^{-2})^{\frac{-1}{2}} \exp \left[ -\frac{SSW}{2\sigma^2} \right] \cdot (h^2)^{\frac{-1}{2}} \exp \left[ -\frac{SSB}{2h^2} \right] \cdot \left[ (1 + d_1) + d_2 \left( \frac{SSB}{h^2} \right) + d_3 \left( \frac{SSB}{h^2} \right)^2 \right] d\sigma^{-2}.
\]

\[ (5.6) \]

5.2 THE NORMAL CASE

If we further assume \( \gamma = 0 \), the above expression reduces to:

\[
p'(\sigma^{-2}) \propto \int_0^\infty (\sigma^{-2})^{\frac{-1}{2}} \exp \left[ -\frac{SSW}{2\sigma^2} \right] \cdot (h^2)^{\frac{-1}{2}} \exp \left[ -\frac{SSB}{2h^2} \right] d\sigma^{-2},
\]

under the assumption \( a_i \sim N(0, \sigma^{-2}) \). This case has been completely examined by Hill [15] whose results are quoted below for comparison with the non-normal case.

Let now \( X = \left[ \begin{array}{c} I-3 \\ SSB \\ 2 \\ 2 \end{array} \right] \) and \( Y = \left[ \begin{array}{c} I(J-1) \\ SSW \\ 2 \\ 2 \end{array} \right] \)

have independent gamma distributions. Then, since \( p'(\sigma^{-2}) \)

is proportional to the density of \( Z = (X - Y)^{-1} J \) on the positive half line, the posterior density \( f(\sigma^{-2}) \propto p_2(\sigma^{-2}) \cdot p'(\sigma^{-2}) \)

may be regarded as a truncation from below at zero of the distribution of \( Z \) together with a weighting by \( p_2(\sigma^{-2}) \).

We shall henceforth regard \( p'(\sigma^{-2}) \) as proportional to the
density of $Z$ on the whole real line, using the fact that $p_x(\sigma_a^2) = 0$ for $\sigma_a^2 < 0$, to yield the required truncation at zero.

It is this $p'(\sigma_a^2)$ that plays the role of a likelihood function (actually a marginal likelihood) in the formation of the marginal posterior density $f(\sigma_a^2)$. Since the distribution of $Z$ depends upon $I$, $J$, SSB, and SSW [and later $d_1$, $d_2$, and $d_3$ when $Y_2 \neq 0$], these four quantities determined by the data are parameters for $p'(\sigma_a^2)$. The important thing to realize is that the form of the posterior $f(\sigma_a^2)$ depends crucially upon where the truncation at zero falls in the function $p'(\sigma_a^2)$, and this, in turn, depends upon the above four parameters.

In particular, since the random variable $Z$ has a unimodal density $p'(\sigma_a^2)$ with mode roughly at $\hat{\sigma}_a^2 = (\text{MSB} - \text{MSW})/J$, it follows that when $\hat{\sigma}_a^2 < 0$ the function $p'(\sigma_a^2)$ will be monotonically decreasing in the interval $\sigma_a^2 \geq 0$, with maximum at $\sigma_a^2 = 0$. Since $f(\sigma_a^2) = 0$ for $\sigma_a^2 < 0$, only this interval is of interest in our investigation.

Thus we see that a negative $\hat{\sigma}_a^2$ leads to a likelihood factor with maximum at $\sigma_a^2 = 0$ and decreasing monotonically in $\sigma_a^2$. The over-all impact of such data is thus to give relatively more weight to small $\sigma_a^2$ in the posterior than in the prior, and this is presumably in accord with some frequentistic interpretations of negative $\hat{\sigma}_a^2$.

However, when we inquire how the degree of sharpness with which $p'(\sigma_a^2)$ is peaked at zero (which determines how strongly the data suggest that $\sigma_a^2$ is small) depends upon the magnitude of the assumed negative $\hat{\sigma}_a^2$, we come to a divergence of view point.
For when \( \hat{c}_{\alpha} = (\text{MSB} - \text{MSW})/J < 0 \), then the larger is SSW (and hence keeping SSB, I, and J fixed, the more negative is \( \hat{c}_{\alpha} \)) and the flatter is the function \( p'(c_{\alpha}) \). In fact, if all other quantities are held fixed

\[
\lim_{SSW \to \infty} f(c_{\alpha}) = p_\alpha(c_{\alpha})
\]

so that the posterior is the same as the prior, and in this case the experiment has been completely uninformative about \( c_{\alpha}^2 \). It is also worth observing that as SSB ↓ 0,

\[
\frac{3 \ln p'(c_{\alpha}^2)}{\sigma c_{\alpha}^2} \bigg|_{0}^{\infty}
\]

decreases to

\[
\frac{I J}{2 \text{MSW}} \cdot \frac{(1 - 1/I)(1 - 1/J)}{(1 - 1/J)} \approx - \frac{I J}{2 \text{MSW}}
\]

when both I and J are large, which indicates that the more negative \( \hat{c}_{\alpha} \) is for fixed SSW, then the stronger is the indication that \( c_{\alpha}^2 \) is small.

5.3 THE NORMAL CASE UNDER THE APPROXIMATION OF INTEGRATION

In the expression (5.7), if we use the same kind of approximation of integration as in the previous chapter, we have:

\[
p'(c_{\alpha}^2) \propto (\sigma_{\alpha}^2 + J \sigma_{\alpha}^2)^{-\frac{1}{2}} \exp \left[ - \frac{\text{SSB}}{2(\sigma_{\alpha}^2 + J \sigma_{\alpha}^2)} \right] \cdot \\
\cdot \int_{0}^{\infty} (\sigma_{\alpha}^2)^{-\frac{1}{2}} \exp \left[ - \frac{\text{SSW}}{2\sigma_{\alpha}^2} \right] d\sigma_{\alpha}^2,
\]

where

\( \tilde{\sigma}_{\alpha}^2 \) is the value of \( \sigma_{\alpha}^2 \) maximizing \( (\sigma_{\alpha}^2)^{-\frac{1}{2}} \exp \left[ - \frac{\text{SSW}}{2\sigma_{\alpha}^2} \right] \).
\[
p'(\sigma^2) \propto \left[ J \sigma^2 + \frac{SSW}{I(J-1)+2} \right]^{-(I-1)/2} \exp \left\{ - \frac{SSB}{2\left[J \sigma^2 + \frac{SSW}{I(J-1)+2}\right]} \right\}, \quad (5.10)
\]

where shows that \( J \sigma^2 \) is distributed as an inverted gamma variate and \( \sigma^2 \) is also distributed in the same way.

The mode of \( \sigma^2 \) is

\[
\text{Mod}(\sigma^2) = \frac{1}{J} \left[ \frac{SSB}{I-1} - \frac{SSW}{I(J-1)+2} \right].
\]

The density of the inverted gamma variate depends on \( I, J, SSB, \) and \( SSW \), and the posterior distribution of \( \sigma^2 \) depends on the location of the point of truncation.

The effect of the point of truncation being located at various places on the graph of \( p'(\sigma^2) \) will now be considered. The typical points chosen are sketched on the graphs of the respective cases now to be considered.

5.3.1 Point of Truncation Is At The Left

In the first place, if the truncation is at the point of origin (that is \( \sigma^2 = 0 \)) then the (marginal) likelihood function is very sharp, in which case we have very good information from the data. Moreover, there is no danger when we use a formal prior for \( p(\sigma^2) \).

This situation arises only when \( \frac{SSW}{I(J-1) + 2} = 0 \), in other words, \( SSW = 0 \), that is \( SST = SSB \). Also we have:
\[ F = \frac{\text{SSB}/(I-1)}{\text{SSW}/I(J-1)} = \frac{\text{MSB}}{\text{MSW}} = \infty \quad \text{when} \quad \text{SSW} \downarrow 0. \]

This shows that when the F-ratio value is quite large, or the null hypothesis is rejected, we have very good posterior information from the data about \( \sigma^2_\alpha \).

Since \( \sigma^2_\alpha = \frac{1}{J} (\text{MSB} - \text{MSW}) \approx \frac{1}{J} \left[ \frac{\text{SSB}}{I-1} - \frac{\text{SSW}}{I(J-1)+2} \right] = \text{Mod} (\sigma^2_\alpha) \),

when \( \sigma^2_\alpha > 0 \), we have the above situation.

In fact, when \( \sigma^2_\alpha \) varies from \( \sigma^2_\alpha \gg 0 \) to \( \sigma^2_\alpha \approx 0 \), the truncation of \( p'(\sigma^2_\alpha) \) falls between the point \( \sigma^2_\alpha = 0 \) (when \( \text{SSW} = 0 \)) and the mode (when \( \text{MSW} = \text{MSB} \)). In all such cases, we have very good posterior information from the data.

The graph of this case is shown below.

5.3.2 Point Of Truncation Is At The Mode

In the second place, if the truncation occurs at the mode, then the (marginal) likelihood function carries much more weight for the smaller \( \sigma^2_\alpha \) and decreasing monotonically with the maximum at \( \sigma^2_\alpha = 0 \).
This will happen only when \( \text{Mod}(\sigma_a^2) = 0 \), which means that
\[
\frac{1}{J} \left[ \frac{\text{SSB}}{I-1} - \frac{\text{SSW}}{I(J-1) + 2} \right] = 0
\]
and, in turn,
\[
\frac{\text{SSB}}{I-1} = \frac{\text{SSW}}{I(J-1) + 2} = \frac{\text{SSW}}{I(J-1)} \cdot \frac{I(J-1)}{I(J-1) + 2}
\]
and the F-ratio is:
\[
F = \frac{\text{MSB}}{\text{MSW}} = \frac{I(J-1)}{I(J-1) + 2} < 1.
\]

Thus \( F \leq 1 \) when \( I \) and \( J \) are large.

In this case, since the variance ratio is around 1, we accept the null hypothesis, which means no variation between groups.

When the truncation is between the mode and the point of inflection, \[ \frac{1}{J} \left\{ 2\text{MSB} - \frac{\text{SSW}}{I(J-1) + 2} \right\} \), we have the same situation as the above, namely that \( p'(\sigma_a^2) \) shows the smaller \( \sigma_a^2 \) from the data.

In other words, when \( \hat{\sigma}_a^2 \) varies from 0 to \( -\frac{\text{SSB}}{J(I-1)} \), it shows us that \( p'(\sigma_a^2) \) has its maximum at \( \sigma_a^2 = 0 \) and decreases monotonically in \( \sigma_a^2 \).

The graph below shows the situation for \( \sigma_a^2 < 0 \).
5.3.3 The Point Of Truncation Is Far To The Right

In the third case, when the point of the truncation is toward the far right-hand end, then the marginal likelihood \( p'(\sigma_\alpha^2) \) is asymptotically close to the abscissa, \( p'(\sigma_\alpha^2) \) is more or less a straight line and the information is flat. In other words, we have no information at all from the data and the posterior density will be proportional to the prior itself, that is \( f(\sigma_\alpha^2) \propto p_2(\sigma_\alpha^2) \).

In this case

\[
\text{Mode}(\sigma_\alpha^2) = \frac{1}{J} \left[ \frac{SSB}{I - 1} - \frac{SSW}{I(J-1) + 2} \right] \ll 0
\]

and \( \text{MSB} - \text{MSW} \ll 0 \)

so that \( F = \frac{\text{MSB}}{\text{MSW}} \ll 1 \).

Furthermore, when \( I, J, \) and \( SSW \) are fixed, the smaller is \( SSB, \) the flatter is \( p'(\sigma_\alpha^2) \) and the larger is \( SSW, \) the flatter is \( p'(\sigma_\alpha^2) \) for \( I, J, \) and \( SSB \) fixed. Actually when we have

\[
\hat{\sigma}_\alpha^2 \ll -\frac{SSB}{J(I - 1)} = -\frac{\text{MSB}}{J} \quad \text{we shall have the above result.}
\]

The following graph shows this case.
5.4 THE CASE WHEN SSB IS NEAR ZERO

Thus far we have considered the normal case and a cognate. We now change the direction of our approach and consider other possibilities. The first of these occurs when SSB is near or at zero. Under this assumption the expression (5.6) yields:

$$p'(\sigma_a^2) \propto \left(\sigma^2\right)^{-2} \exp \left[-\frac{SSW}{2\sigma^2} \right] \exp \left[-\frac{SSB}{2h^2} \right] d\sigma^2.$$

Furthermore

$$p'(\sigma_a^2) \propto \left(\sigma^2\right)^{-2} \exp \left[-\frac{SSW}{2\sigma^2} \right] \exp \left[-\frac{SSB}{2h^2} \right] d\sigma^2,$$

$$\propto \left(J_0^2 + \sigma^2\right)^{-2} \exp \left[-\frac{SSW}{2\sigma^2} \right] \exp \left[-\frac{SSB}{2h^2} \right] d\sigma^2,$$

$$\propto \left[J_0^2 + \frac{SSW}{I(J-1) + 2}\right]^{-2}, \sigma^2,$$

(5.11)

The form of the graph of $p'(\sigma_a^2)$ is shown below.
We see that in this case, we have the same result as before. When SSW gets larger, we have the truncation at the farther right hand of the graph and \( p'(\sigma^2_{\alpha}) \) is flatter. No information from the data is utilized and \( f(\sigma^2_{\alpha}) \propto p(\sigma^2_{\alpha}) \). And again in this situation, the improper prior \( p(\sigma^2_{\alpha}) \) is crucially misleading the experimenter into an improper opinion about \( \sigma^2_{\alpha} \).

On the other hand, if SSW is smaller, then \( p'(\sigma^2_{\alpha}) \) gives us that the smaller \( \sigma^2_{\alpha} \) is, the more dominant is the opinion about \( \sigma^2_{\alpha} \), since \( p'(\sigma^2_{\alpha}) \) has its maximum at \( \sigma^2_{\alpha} \) = 0 and is decreasing monotonically.

The over-all result when SSB = 0 in the equation (5.6) (under non-normality) is the same as if we assume SSB = 0 in the normal case (when \( \gamma_2 = 0 \)).

5.5 THE NON-NORMAL CASE

Now when \( \gamma_2 \neq 0 \) we have the non-normal case. The inference depends on the size of I and the value of \( \gamma_2 \). We now look at a few situations.

When \( \gamma_2 \neq 0 \) and SSB \( \neq 0 \) we may consider \( p'(\sigma^2_{\alpha}) \) as before:

\[
p'(\sigma^2_{\alpha}) \propto (1 + d_1) \cdot f \left\{ \left( \frac{I-3}{2}, \frac{SSB}{2} \right)^{-1} - \left( \frac{I(J-1)}{2}, \frac{SSW}{2} \right)^{-1} \right\}
+ d_2 (I-3) \cdot f \left\{ \left( \frac{I-1}{2}, \frac{SSB}{2} \right)^{-1} - \left( \frac{I(J-1)}{2}, \frac{SSW}{2} \right)^{-1} \right\}
+ d_3 (I-3)(I-1) \cdot f \left\{ \left( \frac{I+1}{2}, \frac{SSB}{2} \right)^{-1} - \left( \frac{I(J-1)}{2}, \frac{SSW}{2} \right)^{-1} \right\}
\]

(5.12)
5.5.1 When $I$ Is Large

In the above expression, for large $I$, the coefficient of the third term is dominant. Then:

$$p'(\sigma^2_a) \asymp f\left\{\left(\left[\frac{I+1}{2}, \frac{SSB}{2}\right]\right)^{-1} - \left(\left[\frac{I(J-1)}{2}, \frac{SSW}{2}\right]\right)^{-1}\right\}$$

$$+ \frac{d_2}{d_3(I-1)} \cdot f\left\{\left(\left[\frac{I-1}{2}, \frac{SSB}{2}\right]\right)^{-1} - \left(\left[\frac{I(J-1)}{2}, \frac{SSW}{2}\right]\right)^{-1}\right\}$$

$$+ \frac{(I-d_1)}{d_3(I-1)(I-3)} \cdot f\left\{\left(\left[\frac{I-3}{2}, \frac{SSB}{2}\right]\right)^{-1} - \left(\left[\frac{I(J-1)}{2}, \frac{SSW}{2}\right]\right)^{-1}\right\}. \quad (5.13)$$

For quite large $I$, we may express this as:

$$p'(\sigma^2_a) \asymp f\left\{\left(\left[\frac{I+1}{2}, \frac{SSB}{2}\right]\right)^{-1} - \left(\left[\frac{I(J-1)}{2}, \frac{SSW}{2}\right]\right)^{-1}\right\}, \quad (5.14)$$

and the interpretation would be the same as before, except that the first parameter of the first variate is increased to

$$\frac{(I+1)}{2} \quad \text{from} \quad \frac{(I-3)}{2}. \quad \cdot$$

5.5.2 Variations In The Value Of $\gamma$. 

If in (5.6) we use the same approximation of the integral (5.10) as before, then:

$$p'(\sigma^2_a) \asymp \int_0^\infty (\sigma^2)^{\frac{-I(I+1)}{2}} \exp\left[-\frac{SSW}{2\sigma^2}\right] (h^2)^{\frac{-I(I+1)}{2}} \exp\left[-\frac{SSB}{2h^2}\right] \cdot$$

$$\cdot \cdot \left[(1+d_1) + d_2\left(\frac{SSB}{h^2}\right) + d_3\left(\frac{SSB}{h^2}\right)^2\right] \, d\sigma^{-1}.$$
\[ \alpha \left\{ (1 + d_1)(J \sigma_a^{-1} + \overline{\sigma}^2)^{-1/2} \exp \left[ - \frac{SSB}{2(J \sigma_a^{-1} + \overline{\sigma}^2)} \right] \right. \\
+ d_2(\text{SSB})(J \sigma_a^{-1} + \overline{\sigma}^2)^{-1/2} \exp \left[ - \frac{SSB}{2(J \sigma_a^{-1} + \overline{\sigma}^2)} \right] \right. \\
+ d_3^2(\text{SSB})(J \sigma_a^{-1} + \overline{\sigma}^2)^{-1/2} \exp \left[ - \frac{SSB}{2(J \sigma_a^{-1} + \overline{\sigma}^2)} \right] \right\} . \\
\int_0^{\infty} (\sigma^2)^{-1/2} \exp \left[ - \frac{SSW}{2\sigma^{-2}} \right] d\sigma , \quad (5.15) \]

where \( \overline{\sigma}^2 \) is the maximizing value of \( \sigma^2 \) in the integrand.

We may then write:

\[ p'(\sigma_a^2) \propto \left\{ (1 + d_1) \left[ J \sigma_a^{-1} + \frac{SSW}{I(J-1) + 2} \right]^{-1/2} \right. \\
+ d_2(\text{SSB}) \left[ J \sigma_a^{-1} + \frac{SSW}{I(J-1) + 2} \right]^{-1/2} \right. \\
+ d_3^2(\text{SSB}) \left[ J \sigma_a^{-1} + \frac{SSW}{I(J-1) + 2} \right]^{-1/2} \right\} . \\
\exp \left\{ - \frac{SSB}{2\left[ J \sigma_a^{-1} + \frac{SSW}{I(J-1) + 2} \right]} \right\} \quad (5.16) \]

\( p'(\sigma_a^2) \) is approximately proportional to the sum of density functions, each of which is an inverted gamma density function, defined over the interval

\[ \sigma_a^{-1} \geq \frac{SSW}{J \left[ I(J-1) + 2 \right]} \approx \frac{MSW}{J} . \]
If we assume $\gamma_2 = 3$, i.e., the $\xi_i$ are double exponential random variables, then we naturally expect a much larger SSR than in the normal random variate case, since the $\xi_i$ are more sharply concentrated around the mean and have heavier tails.

The coefficient of the third term will then be quite dominant over the other terms and $p^\prime(\sigma_a^2)$ will be approximately:

$$p^\prime(\sigma_a^2) \propto (\tilde{\sigma}^2 + J \sigma_a^2)^{-\frac{1}{2}} \exp \left[ -\frac{\text{SSR}}{2(\tilde{\sigma}^2 + J \sigma_a^2)} \right].$$

The mode will then most likely be on the positive side of $\sigma_a^2$. Therefore we will have quite a sharp likelihood function and the general shape of the graph will be:

![Graph](image-url)
If, on the other hand, we assume \( \gamma_2 = 1.2 \), i.e. the \( a_i \) are uniform random variates, then we expect a much smaller SSB than for normal random variates. To study the effect of SSB, let us use

\[
p'(\sigma_a^2) \propto (1 + d_1) \cdot f\left\{ \left( \frac{I-3}{2} \right) \left[ \frac{I}{2} \right] - \left( \frac{I+1}{2} \right) \left[ \frac{I+1}{2} \right] \right\} \\
+ d_2 (I-3) \cdot f\left\{ \left( \frac{I-1}{2} \right) \left[ \frac{I}{2} \right] - \left( \frac{I+1}{2} \right) \left[ \frac{I+1}{2} \right] \right\} \\
+ d_3 (I-3)(I-1) \cdot f\left\{ \left( \frac{I+1}{2} \right) \left[ \frac{I}{2} \right] - \left( \frac{I+1}{2} \right) \left[ \frac{I+1}{2} \right] \right\}
\]

As SSB becomes smaller, the mode in each term becomes smaller or negative, and the mode of \( p' \) will tend to move to the negative side of \( \sigma_a^2 \). The general shape of the graph is then:

![Graph of \( p'(\sigma_a^2) \) vs. \( \sigma_a^2 \)](image-url)
Furthermore, we note that as SSB becomes smaller in this case, the likelihood function will give a much smaller $\sigma_0^2$.

If we also consider $I$ large, then again the coefficient of the third term will dominate the other two. Therefore, for large $I$ we have the approximation:

$$p'(\sigma_0^2) \propto f \left\{ \left( \left[ \frac{I+1}{2}, \frac{SSB}{2} \right] \right) - \left( \left[ \frac{I(J-1)}{2}, \frac{SSW}{2} \right] \right) \right\}$$

which has the mean

$$\frac{1}{J} \left[ \frac{SSB}{I-1} - \frac{SSW}{I(J-1) - 2} \right] = \frac{1}{J} \left[ \frac{MSB - SSW}{I(J-1) - 2} \right] \frac{MSB - MSW}{J}$$

and the mode is approximately

$$\frac{1}{J} \left[ \frac{SSB}{I+2} - \frac{SSW}{I(J-1) + 2} \right].$$

Therefore for the non-normal case the approximate likelihood function of $\sigma_0^2$ is identical to that of the normal case except that the first parameter of the inverted gamma function variate is different; that is, it is increased from $(I-3)/2$ to $(I+1)/2$ regardless of the value of $Y_2$.

Thus the result is the same as that found before we assumed $I$ to be large, and the analysis is the same as for the normal case.
5.6 THE NON-INFORMATIVE PRIOR

Finally, if we use Jeffrey's prior

\[ p(\sigma^{-2}, \sigma_{\alpha}^{-2}) = (\sigma^{-2})^{-1} (\sigma^{-2} + J\sigma_{\alpha}^{-2})^{-1}, \]

then the posterior distribution of \( \sigma_{\alpha}^{-2} \) is:

\[
f(\sigma_{\alpha}^{-2}) \propto \int_0^{\infty} p(\sigma^{-2}, \sigma_{\alpha}^{-2})(\sigma^{-2})^{-1} \exp\left[-\frac{SSB}{2\sigma^{-2}}\right](\sigma^{-2} + J\sigma_{\alpha}^{-2})^{-1} \exp\left[-\frac{SSB}{2h^2}\right] \exp\left[-\frac{SSB}{2\sigma^{-2}}\right] (h^2) \exp\left[-\frac{SSB}{2h^2}\right]
\]

\[
\times \left[1 + d_1 + d_2 \left(\frac{SSB}{\sigma^{-2} + J\sigma_{\alpha}^{-2}}\right) + d_3 \left(\frac{SSB}{\sigma^{-2} + J\sigma_{\alpha}^{-2}}\right)^2\right] d\sigma^{-2}
\]

\[
\alpha \propto \int_0^{\infty} (\sigma^{-2})^{-1} \exp\left[-\frac{SSW}{2\sigma^{-2}}\right](h^2) \exp\left[-\frac{SSB}{2h^2}\right] \exp\left[-\frac{SSB}{2\sigma^{-2}}\right] (h^2) \exp\left[-\frac{SSB}{2h^2}\right] \exp\left[-\frac{SSB}{2\sigma^{-2}}\right] (h^2) \exp\left[-\frac{SSB}{2h^2}\right]
\]

\[
\times \left[1 + \frac{d_2}{1 + d_1} \frac{SSB}{h^2} + \frac{d_3}{1 + d_1} \left(\frac{SSB}{h^2}\right)^2\right] d\sigma^{-2} . \quad (5.17)
\]

If \( \gamma_1 = 0 \), then the above expression becomes simply:

\[
f(\sigma_{\alpha}^{-2}) \propto \int_0^{\infty} (\sigma^{-2})^{-1} \exp\left[-\frac{SSW}{2\sigma^{-2}}\right](h^2) \exp\left[-\frac{SSB}{2h^2}\right] \exp\left[-\frac{SSB}{2\sigma^{-2}}\right] (h^2) \exp\left[-\frac{SSB}{2h^2}\right] \exp\left[-\frac{SSB}{2\sigma^{-2}}\right] (h^2) \exp\left[-\frac{SSB}{2h^2}\right]
\]

\[
\times \left[1 + \frac{d_2}{1 + d_1} \frac{SSB}{h^2} + \frac{d_3}{1 + d_1} \left(\frac{SSB}{h^2}\right)^2\right] d\sigma^{-2} .
\]

In this case the posterior distribution of \( \sigma_{\alpha}^{-2} \) is that of the random variable

\[
J^{-1} \left[\frac{SSB}{\chi_{(1)}^2} - \frac{SSW}{\chi_{(1)}^2}\right], \quad \quad (5.18)
\]

truncated from below at zero. Hill [15] observes

This distribution, censored instead of truncated from below at zero, has also been proposed as a fiducial solution.
If, as in our case, we use the same approximation for the integral as before (5.10)

\[ f(\sigma^2_\alpha) \propto \left( J \sigma^2_\alpha + \frac{SSW}{I(J-1)+2} \right)^{-\frac{(1-\nu)}{2}} \exp \left\{ -\frac{SSB}{2 \left( J \sigma^2_\alpha + \frac{SSW}{I(J-1)+2} \right)} \right\} \]

we may apply Hill's [15] observation

Our viewpoint is that when \( \sigma^2_\alpha < 0 \), and particularly when MSW is large, this distribution (even when truncated, and certainly when censored) is inappropriate as a measure of posterior opinion.

In our case when \( \frac{1}{J} \left[ \frac{SSW}{I(J-1)+2} \right] \simeq \frac{MSW}{J} \) is large, we have no information at all about \( \sigma^2_\alpha \), which again agrees with Hill's [15] remark

For we have seen that in this case the data are extremely uninformative, that \( f(\sigma^2_\alpha) \simeq p_2(\sigma^2_\alpha) \) and, since one is left with his prior density, it is inappropriate to use merely a formal prior such as that of Jeffreys.

From our point of view, such a formal prior is useful only as a convenient, but more or less arbitrary, choice made from a collection of priors all of which essentially yield the likelihood function (normalized in some way) as a posterior distribution.

When, as in the present case with \( \sigma^2_\alpha \ll 0 \), the likelihood factor is flat relative to the prior (the reverse of "stable estimation"), it becomes crucial to carefully assess the prior.
Now if we look at the case when $\gamma \neq 0$, the expression (5.17) gives us

$$
\begin{align*}
\varphi^2 \propto (1 + d_1) \int \left\{ \frac{1}{J} \left[ \frac{\varphi_{(H)}^2}{\chi_{(H)}^2} - \frac{\varphi_{(L)}^2}{\chi_{(L)}^2} \right] \right\} \\
+ d_2 (I-1) \int \left\{ \frac{1}{J} \left[ \frac{\varphi_{(H)}^2}{\chi_{(H)}^2} - \frac{\varphi_{(L)}^2}{\chi_{(L)}^2} \right] \right\} \\
+ d_3 (I-1)(I+1) \int \left\{ \frac{1}{J} \left[ \frac{\varphi_{(H)}^2}{\chi_{(H)}^2} - \frac{\varphi_{(L)}^2}{\chi_{(L)}^2} \right] \right\}.
\end{align*}
$$

(5.19)

when $I$ is large, the coefficient of the third term again dominates the other two. Then we may approximate (5.19) by:

$$
\varphi^2 \propto \int \left\{ \frac{1}{J} \left[ \frac{\varphi_{(H)}^2}{\chi_{(H)}^2} - \frac{\varphi_{(L)}^2}{\chi_{(L)}^2} \right] \right\} 
$$

(5.19.1)

and the analysis will be the same as before when $\gamma = 0$, except that the magnitude of $\varphi^2$ depends on $\gamma$.

Finally, in the expression (5.17), if we use the same approximation for the integral, we have:

$$
\varphi^2 \propto \int_0^\infty (n^2)^{-\frac{\gamma-I+1}{2}} \exp \left[ - \frac{\varphi_{(H)}^2}{2n^2} \right] \left\{ (1 + d_1)(n^2)^{-\frac{\gamma-I+1}{2}} \exp \left[ - \frac{\varphi_{(L)}^2}{2n^2} \right] \\
+ (d_2) (\varphi_{(H)}^2)(n^2)^{-\frac{\gamma-I+1}{2}} \exp \left[ - \frac{\varphi_{(L)}^2}{2n^2} \right] \\
+ (d_3) (\varphi_{(H)}^2)(n^2)^{-\frac{\gamma-I+1}{2}} \exp \left[ - \frac{\varphi_{(L)}^2}{2n^2} \right] \right\} dn^2
$$
\[
\frac{f(\sigma_a^2)}{\propto (1+q_1)^{-2} + \sigma_a^2} \exp \left\{ - \frac{\sigma_a^2}{2(\tilde{\gamma}_2^2 + J \sigma_a^2)} \right\} \]

\[
+ \left( d_2(\text{SSR}) \right) \left( \tilde{\gamma}_2^2 + J \sigma_a^2 \right)^{-2} \exp \left\{ - \frac{\sigma_a^2}{2(\tilde{\gamma}_2^2 + J \sigma_a^2)} \right\}
\]

\[
+ \left( d_3(\text{SSR}) \right) \left( \tilde{\gamma}_2^2 + J \sigma_a^2 \right)^{-3} \exp \left\{ - \frac{\sigma_a^2}{2(\tilde{\gamma}_2^2 + J \sigma_a^2)} \right\}
\]

In comparison with the likelihood function (5.16) obtained in the previous section, the posterior density function under non-informative prior is identical to the likelihood function except for the parameter. Therefore, when \( \gamma_2 = 0 \), this result becomes the likelihood function for the normal case, except for the parameter.

In particular, when \( \gamma_2 \approx 1.2 \), and the SSR is small and the SSW is large: say for SSR \( \approx 0 \), or SSR \( \ll \) SSW, we have

\[
\frac{f(\sigma_a^2)}{\propto \left[ J \sigma_a^2 + \frac{\text{SSW}}{I(J-1) + 2} \right] \exp \left\{ - \frac{\text{SSR}}{2 \left[ J \sigma_a^2 + \frac{\text{SSW}}{I(J-1) + 2} \right]} \right\}}
\]

in which case the mode is negative. This means that the...
posterior density function is asymptotic to a horizontal line parallel to the x-axis, i.e., it is approximately constant. Thus, this posterior density function gives no information at all.
The two preceding chapters have produced results which are useful and interesting in themselves and which are also used to find the F ratio, the important final measure in the analysis of variance. These results are based on the likelihood function discussed in Chapter 3 and on some priors which were chosen for their interest or utility.

The next step in finding the F ratio is the development of the ratio \( \frac{\sigma^2_a}{\sigma^2} \). We now, therefore, derive the posterior inference on this ratio using, as before, various priors.

6.1 THE FORM OF THE EQUATION

We now find the posterior density of \( \frac{\sigma^2_a}{\sigma^2} \) based upon the prior:

\[
p(\mu, \sigma^2, \sigma^2_a) \propto (\sigma^2)^{-\frac{1}{2}} (\sigma^2_a)^{-\frac{1}{2}} \exp \left[ -\frac{c_a}{2\sigma^2_a} \right] .
\]

The posterior will be:

\[
p^*(\sigma^2, \sigma^2_a) \propto p(\sigma^2, \sigma^2_a) \cdot L(\sigma^2, \sigma^2_a)
\]

\[
\propto (\sigma^2)^{-\frac{1}{2}} \exp \left[ -\frac{SSW}{2\sigma^2} \right] (\sigma^2_a)^{-\frac{1}{2}} \exp \left[ -\frac{c_a}{2\sigma^2_a} \right] .
\]

\[
\cdot (\sigma^2 + J\sigma^2_a)^{-\frac{1}{2}} \exp \left[ -\frac{SSB}{2(\sigma^2 + J\sigma^2_a)} \right] .
\]

\[
\cdot \left[ 1 + d_1 + d_2 \left( \frac{SSB}{\sigma^2 + J\sigma^2_a} \right) + d^2_3 \left( \frac{SSB}{\sigma^2 + J\sigma^2_a} \right)^2 \right] .
\]

(6.2)
To find the posterior density function of $T^2 = \sigma^2 / \sigma_0^2$, we make the following change of variables:

$$
\begin{align*}
\left\{ \begin{array}{c}
\hat{\sigma}^2 = \sigma^2 \\
T = \sigma^2 / \sigma_0^2
\end{array} \right. \quad \Rightarrow \quad \left\{ \begin{array}{c}
\hat{\sigma}^2 = \sigma^2 \\
\sigma_0^2 = \sigma^2 T
\end{array} \right.,
\end{align*}
$$

$$
g^n(\sigma^2, T) \propto p^n(\sigma^2, \sigma_0^2 = \sigma^2 T) |J|$

$$
\exp \left[ -\frac{\lambda}{2} \right] \exp \left[ \frac{\lambda \sigma^2}{2} T^2 \right] \exp \left[ -\frac{\lambda \sigma_0^2}{2} T^2 \right] \ldots$$

$$
\left[ 1 + d_1 + d_2 \right] \left[ \frac{SSB}{\sigma_0^2 + J \sigma^2 T} \right] + d_3 \left[ \frac{SSB}{\sigma_0^2 + J \sigma^2 T} \right]^2
$$

6.2 THE MARGINAL DENSITY OF $T^2$

The marginal density of $T^2$ is then given by:

$$
g^n(T^2) = \int_0^\infty g^n(\sigma^2, T^2) \, d\sigma^2$$
\[ g^n(T^2) \propto \frac{(T^2)}{(SSW + c_a / T^2 + SSB/(1 + J T^2))^{(N-1) + \lambda_a / 2}} \cdot (1 + J) \cdot T^2 \geq 0\]

\[ \lambda_a \}

\[ d_2(SSB) (\lambda_a + N - 1) \frac{(T^2)}{(SSW + c_a / T^2 + SSB/(1 + J T^2))^{(N-1) + \lambda_a / 2}} \cdot \frac{1}{J} \cdot (T^2 + 1/J) \cdot (1 - (i - 1)^{1/2} - 1)\]

\[ d_3(SSB) (\lambda_a + N - 1) (\lambda_a + N + 1) \frac{(T^2)}{(SSW + c_a / T^2 + SSB/(1 + J T^2))^{(N+3) + \lambda_a / 2}} \cdot \frac{1}{J^2} \cdot (T^2 + 1/J) \cdot (1 - (i - 1)^{1/2} - 2)\]
6.3 THE CASE OF $\lambda_\alpha = 0, c_\alpha = 0$

If we had chosen $\lambda_\alpha = 0, c_\alpha = 0$, corresponding to the prior $(\sigma_\alpha^2)^{-1}$, then:

$$g^w(\tau^2) \propto (1+d_1)(\tau^2)^{-1/2}(\tau^2 + 1/J)^{-1/2}/\left( SSB + \frac{SSB}{1+J\tau^2}\right)^{(N-1)/2}$$

$$+ \frac{d_2(SSB)(N-1)}{J} (\tau^2)^{-1}(\tau^2 + 1/J)^{-1/2}/\left( SSW + \frac{SSB}{1+J\tau^2}\right)^{(N+1)/2}$$

$$+ \frac{d_3(SSB)^2(N+1)(N-1)}{J} (\tau^2)^{-1}(\tau^2 + 1/J)^{-1/2}/\left( SSW + \frac{SSB}{1+J\tau^2}\right)^{(N+3)/2}$$

$$\propto (1+d_1)(\tau^2)^{i_j=1/2}/\left[ SSB + SSW(1+J\tau^2)\right]^{(i_j-1)/2}$$

$$+ d_2(SSB)(i_j-1)(\tau^2)^{i_j-1/2}/\left[ SSB + SSW(1+J\tau^2)\right]^{(i_j-1)/2 + 1}$$

$$+ d_3(SSB)^2(i_j+1)(i_j-1)(\tau^2)^{i_j-1/2}/\left[ SSB + SSW(1+J\tau^2)\right]^{(i_j-1)/2 + 2}$$

In this case, we have the some difficulty at $\tau^2$ near zero. So to ease the difficulty and for convenience, we approximate

$$(\tau^2)^{-1} \text{ by } (\tau^2 + 1/J)^{-1}.$$ 

We then obtain:
\[ g^*(T^2) \propto \left( 1 + d_1 \right) \frac{(1 + Jt^2)^{I(J-1)/2 - 1}}{\left[ SSB + SSW(1 + Jt^2) \right]^{(I(J-1))/2}} \]

\[ + \frac{d_2 (SSB)(IJ-1)(1 + Jt^2)^{I(J-1)/2 - 1}}{\left[ SSB + SSW(1 + Jt^2) \right]^{(I(J+1))/2}} \]

\[ + d_3 (SSB^2) \left( IJ - 1 \right)(IJ+1)(1 + Jt^2)^{I(J-1)/2 - 1} \]

\[ \frac{SSW}{SSB} \left( 1 + Jt^2 \right)^{I(J-1)/2 - 1} \]

\[ \times \frac{d_2 (IJ-1) \left( SSW \right)}{SSB^2} \left( 1 + Jt^2 \right)^{I(J-1)/2 - 1} \]

\[ \frac{SSW}{SSB} \left( 1 + Jt^2 \right)^{I(J+1)/2} \]

\[ + \frac{d_3 (IJ-1)(IJ+1) \left( SSW \right)}{SSB^2} \left( 1 + Jt^2 \right)^{I(J+3)/2} \]

\[ \times \frac{SSW}{SSB} \left( 1 + Jt^2 \right)^{I(J+3)/2} \]

Actually, this result would also be obtained if we had used the Jeffreys prior from the beginning. To see this, we now repeat the derivation using Jeffreys prior.

Thus we have, using Jeffreys prior:

\[ g^*(T^2) \propto \left( 1 + d_1 \right) \frac{(1 + Jt^2)^{I(J-1)/2 - 1}}{\left[ SSB + SSW(1 + Jt^2) \right]^{(I(J-1))/2}} \]

\[ + \frac{d_2 (SSB)(IJ-1)(1 + Jt^2)^{I(J-1)/2 - 1}}{\left[ SSB + SSW(1 + Jt^2) \right]^{(I(J+1))/2}} \]

\[ + d_3 (SSB^2) \left( IJ - 1 \right)(IJ+1)(1 + Jt^2)^{I(J-1)/2 - 1} \]

\[ \frac{SSW}{SSB} \left( 1 + Jt^2 \right)^{I(J-1)/2 - 1} \]

\[ \times \frac{d_2 (IJ-1) \left( SSW \right)}{SSB^2} \left( 1 + Jt^2 \right)^{I(J-1)/2 - 1} \]

\[ \frac{SSW}{SSB} \left( 1 + Jt^2 \right)^{I(J+1)/2} \]

\[ + \frac{d_3 (IJ-1)(IJ+1) \left( SSW \right)}{SSB^2} \left( 1 + Jt^2 \right)^{I(J+3)/2} \]

\[ \times \frac{SSW}{SSB} \left( 1 + Jt^2 \right)^{I(J+3)/2} \]
Now, for the moment, let us consider the posterior density only when $\gamma_\perp = 0$ (that is, the normal case). Then the expression reduces as follows:

\[
g^*(\tau^2) \propto \frac{\left( \frac{SSW}{SSB} \right)^{2(J-1)/2} \chi^2_{(I-J)/2}}{\left\{ 1 + \left[ \frac{SSW}{SSB} \left( 1 + 1/\tau^2 \right) \right]^{(I-J)/2} \right\}^{(I-J)/2}} \tag{6.8}
\]

Hill [15] has treated this fully, therefore his results will be quoted for comparison with the non-normal case.

\[
\frac{SSW}{SSB} \left( 1 + 1/\tau^2 \right) \sim \frac{\chi^2_{(J-I)/2}}{\chi^2_{(I-I)/2}}
\]

truncated from below at $SSW/SSB$, or

\[
(1 + 1/\tau^2) \sim \frac{MSB}{MSW} F_{I(I-I),J-I}, \text{ with } F
\]

truncated from below at $MSW/MSB$, in some but not complete harmony with tradition. Here $F_{I(I-I),J-I}$ denotes a random variable having the truncated F distribution with the indicated number of degrees of freedom.

Now let us consider the situation where $MSB \leq MSW$. To gain insight, we shall make the approximation

\[
(1 + 1/\tau^2) \sim \frac{MSB}{MSW} F_{I(I-I),J-I}\text{F}
\]

with F truncated from below at $MSW/MSB$, and find that

\[
Pr \left\{ \frac{MSW}{MSB} (1+J \tau_t) \leq F \leq \frac{MSW}{MSB} (1+J \tau_t) \right\} = \frac{Pr \{ t_o \leq \tau^2 \leq t_0 \} \frac{MSW}{MSB} (1+J \tau_t) \leq F \leq \frac{MSW}{MSB} (1+J \tau_t) \}}{Pr \{ \frac{MSW}{MSB} \leq F \leq \infty \}} \tag{6.9}
\]
where here $F$ has an ordinary $F$ distribution with the above degrees of freedom.

Since the $F$ has a simple mode at

$$f^* = \frac{[I(J-1)-2](I-1)}{I(J-1)[(I-1)+2]} = \frac{1 - 2/[I(J-1)]}{1 + 2/(I-1)} < 1,$$

and the mode approaches to 1 when both degrees of freedom grow large.

If $MSB \leq MSW$ ($1 \leq MSW/MSB$), then the truncation is at the right-hand side of the mode and the density function of $F^*$ is monotonically decreasing. Therefore the probability

$$\Pr\{0 \leq t \leq t_1\}$$

for the interval of given length will be maximal when $t = 0$, and

$$\Pr\{0 \leq t \leq t_1\} = \frac{Pr\{MSW \leq F \leq \frac{MSW}{MSB} (1 + J t_1)\}}{Pr\{MSW \leq F \leq \infty\}}.$$

Again we quote Hill's [15] results for comparison.

Now, depending on the degrees of freedom, unity (1.00) may be either more or less than the 50th percentile of $F^*$. However, in the present case, with degrees of freedom of the numerator larger than the degrees of freedom of the denominator, unity is always less than the 50th percentile, the latter approaching unity as both degrees of freedom grow large:

When $\frac{MSW}{MSB} \approx 1$ (that is $\frac{MSW}{MSB} \downarrow 1$), $\Pr\{0 \leq t \leq t_1\}$ may be quite large for even small $t_1$.

For example, if both degrees of freedom are large, then:

$$\Pr\{0 \leq t \leq t_1\} \approx 2Pr\{1 \leq F \leq 1 + J t_1\}$$

and if $t_1 = J^{-1}$ this is $2 \cdot Pr\{1 \leq F^* \leq 2\}$ which will be very near to unity.
But when the ratio MSW/MSB is really large, the truncation at MSW/MSB from below occurs at the far right-hand side, and the distribution of $F^*$ approaches a horizontal straight line. In the limit then, all values of $T^2$ become equally probable and no information about $T^2$ is gained. It means that the posterior distribution of $T^2$ becomes diffuse when $MSW/MSB \rightarrow \infty$. This unacceptable result is originally due to the Jeffrey's prior. Therefore when $MSB \ll MSW$, we must carefully choose the prior. We have also seen that the convenient Jeffrey's type of non-informative prior leads us to non-informative posterior opinion.

6.5 THE NON-NORMAL CASE

Now, when $Y_2 \neq 0$, from (6.7) we have:

\[
g^\text{w}(T^2) \propto \frac{(1 + d_1)}{B\left[\frac{1}{2} I (J-1), \frac{1}{2} (I-1)\right]} \cdot \frac{\left[\frac{SSW}{SSB} (1 + J T^2)\right]^{(J-I)/2-1}}{\left[\frac{SSW}{SSB} (1 + J T^2)\right]^{(IJ-I)/2}}
\]

\[
+ \frac{d_2 (IJ-1)}{B\left[\frac{1}{2} I (J-1), \frac{1}{2} (I-1)\right]} \cdot \frac{B\left[\frac{1}{2} I (J-1), \frac{1}{2} (I+1)\right]}{B\left[\frac{1}{2} I (J-1), \frac{1}{2} (I+1)\right]} \cdot \frac{\left[\frac{SSW}{SSB} (1 + J T^2)\right]^{(J-I)/2-1}}{\left[\frac{SSW}{SSB} (1 + J T^2)\right]^{(IJ-I)/2}}
\]

\[
+ \frac{d_3 (IJ-1)(IJ+1)}{B\left[\frac{1}{2} I (J-1), \frac{1}{2} (I+3)\right]} \cdot \frac{B\left[\frac{1}{2} I (J-1), \frac{1}{2} (I+1)\right]}{B\left[\frac{1}{2} I (J-1), \frac{1}{2} (I-1)\right]} \cdot \frac{\left[\frac{SSW}{SSB} (1 + J T^2)\right]^{(J-I)/2-1}}{\left[\frac{SSW}{SSB} (1 + J T^2)\right]^{(IJ-I)/2}}
\]
which states that \[ \frac{SSS}{SSB} \left( 1 + J\tau^2 \right) \sim \chi^2_{(I-1)} \] in the first term, \[ \frac{SSS}{SSB} \left( 1 + J\tau^2 \right) \sim \chi^2_{(I+1)} \] in the second term, and in the third term \[ \frac{SSS}{SSB} \left( 1 + J\tau^2 \right) \sim \chi^2_{(I+3)} \]. Therefore we may express as

\[ g^m(\tau^2) \propto (1 + d_1) \cdot \frac{\chi^2_{(I-1)}}{\chi^2_{(I-1)}} \left[ \frac{SSS}{SSB} \left( 1 + J\tau^2 \right) \right] \]

\[ + d_2 (IJ-1) \frac{B[\frac{1}{2}I(J-1), \frac{1}{2}(I+1)]}{B[\frac{1}{2}I(J-1), \frac{1}{2}(I-1)]} \cdot \frac{\chi^2_{(I+1)}}{\chi^2_{(I+1)}} \left[ \frac{SSS}{SSB} \left( 1 + J\tau^2 \right) \right] \]

\[ + d_3 (IJ-1)(I+1) \frac{B[\frac{1}{2}I(J-1), \frac{1}{2}(I+3)]}{B[\frac{1}{2}I(J-1), \frac{1}{2}(I-1)]} \cdot \frac{\chi^2_{(I+3)}}{\chi^2_{(I+3)}} \left[ \frac{SSS}{SSB} \left( 1 + J\tau^2 \right) \right] . \]  (6.11)

And as in the previous case of \( \gamma_2 = 0 \), we may express \( g^m(\tau^2) \), analogously, in the density function of a F-random variate:

\[ g^m(\tau^2) \propto (1 + d_1) \cdot \frac{\chi^2_{(I-1)}}{\chi^2_{(I-1)}} \left[ \frac{MSW}{MSB} \right] \]

\[ + d_2 (I+1) \frac{\chi^2_{(I-1)}}{\chi^2_{(I-1)}} \left[ \frac{MSW}{SSB/(I-1)} \right] \]

\[ + d_3 (I+1)(I+3) \frac{\chi^2_{(I+3)}}{\chi^2_{(I+3)}} \left[ \frac{MSW}{SSB/(I+3)} \right] , \]  (6.12)

which means that in the first, second, and third term,

\[ \frac{MSW}{MSB} \left( 1 + J\tau^2 \right), \quad \frac{MSW}{SSB/(I-1)} \left( 1 + J\tau^2 \right), \quad \text{and} \quad \frac{MSW}{SSB/(I+3)} \left( 1 + J\tau^2 \right) \] are distributed as \( F_{I,(I-1),(1+3)} \), \( F_{I,(I+1),(1+3)} \), and \( F_{I,(I+3),(1+3)} \), respectively, and
and being truncated from below at \( \frac{\text{MSW}}{\text{MSB}} \), \( \frac{\text{MSW}(I+1)}{\text{SSB}} \), and \( \frac{\text{MSW}(I+3)}{\text{SSB}} \) in that order.

By applying the previous information of truncation and mode of the F-distribution function to our posterior, \( g^n(\tau) \), let us consider the same situation \( \frac{\text{MSW}}{\text{MSB}} \leq \frac{\text{MSW}}{\text{MSB}} \).

When \( \frac{\text{MSW}}{\text{MSB}} \approx 1 \) (although \( \frac{\text{MSW}}{\text{MSB}} \geq 1 \)), then the three F-distribution functions have their truncations at the right-hand side of their mode near to 1, since we have the relationship as

\[
\frac{\text{MSW}(I+3)}{\text{SSB}} > \frac{\text{MSW}(I+1)}{\text{SSB}} > \frac{\text{MSW}(I-1)}{\text{SSB}} = \frac{\text{MSW}}{\text{MSB}} \geq 1 \geq f_{\text{mod}}.
\]

Therefore each of these three density functions of F has its maximum at the point of truncation and is monotonically decreasing.

In the other extreme case, when \( \frac{\text{MSW}}{\text{MSB}} \) goes to \( \infty \), then, as in the normal case, the posterior, \( g^n(\tau) \), is non-informative (flat). Since each F-distribution function is straight

\[
g^n(\tau) \propto (1+d_1) f_{\text{F}(I+1),I} \left( \frac{\text{MSW}}{\text{MSB}} (1+J\tau^2) \right) + d_2(I+1) f_{\text{F}(I),I} \left( \frac{\text{MSW}(I+1)}{\text{SSB}} (1+J\tau^2) \right) + d_3(I+1)(I+3) f_{\text{F}(I+3),I} \left( \frac{\text{MSW}(I+3)}{\text{SSB}} (1+J\tau^2) \right)
\]

\[
\propto (1+d_1) + d_2(I+1) + d_3(I+1)(I+3)
\]

\( \propto 1 \).
The information about non-normality, particularly $\gamma_2$, is not used at all.

Such a non-informative posterior is imbedded by using Jeffrey's non-informative prior or by adopting the approximation $(\tau')^{-1}$ by $(\tau' + 1/J)^{-1}$.

If we again assume that $\gamma_2 = -1.2$, then SSB is smaller as was shown in the previous chapters. In this case $\text{MSW}/\text{MSB} \to \infty$. We would also have the same situation when $\hat{\sigma}_2^2 \ll 0$. Thus, all these cases would yield the non-informative posterior.

On the other hand, if $\gamma_2 = 3$, we expect larger SSB, and $\text{MSW}/\text{MSB} \ll 1$. In this case we have truncation at, or to the left of, the mode, which will give us good sharp posterior information about $\tau^2$.

Lastly, if we assume $I$ to be large, the coefficient of the third term of the expansion dominates the other two terms. In this case, the posterior density function of $\tau^2$ may be approximated by

$$f_{\text{MSW}(I+3)}\left[\frac{\text{MSW}(I+3)}{S\text{SB}(1 + J\tau^2)}\right]$$

and the analysis becomes the same as the normal case except for the second degree of freedom being $(I+3)$ instead of $(I-1)$.

In this chapter we have found for $\gamma_2 = 0$ the posterior density function of the ratio $\tau^2 = \hat{\sigma}_2^2 / \sigma^2$ and considered the extreme cases where $\text{MSB} \ll \text{MSW}$ and $\text{MSW}/\text{MSB} \approx 1$. These results will be considered together with those of the preceding chapters in the next chapter.
Chapter 7

CONCLUSION

7.1 OVERVIEW OF THE PROBLEM

The classical analysis of variance requires that either the data conform to stringent requirements on the population and samples, or that the experimenter be left with some degree of doubt as to the validity of the analysis. Various approaches to study the problem of analysis of variance when the data do not conform to these requirements have been taken, but the Bayesian approach offers convenience both in the analysis and in the freedom to choose the form of the distribution which is assumed for the data. These assumptions on the form of the distribution may be made on the basis of subjective probability (e.g. what the experimenter believes appropriate to the case) or objective probability (i.e. some actual knowledge of the distribution of the data).

One of the requirements in the classical analysis of variance is that the random effects of the model be normally distributed. In this dissertation we investigated certain impacts on the analysis when the random effects are distributed not normally but according to the class of exponential power functions. We also investigated the meaning of a negative estimate, $\hat{\sigma}_e^2$, of the variance of the random effects, as explained in section 7.2.
7.2 THE METHOD USED

In the present analysis it has been assumed that the variate has the form:

$$y_{ij} = \mu + a_i + e_{ij}$$

where $$y_{ij}$$ is the $$j^{th}$$ observation within the $$i^{th}$$ group, 
$$\mu$$ is a location parameter, 
$$a_i$$ is the random effect operating upon the $$i^{th}$$ group, 
and $$e_{ij}$$ is the random error of the $$j^{th}$$ observation within the $$i^{th}$$ group.

Some previous work using this form of the variate has been directed at the distribution of $$\sigma^2$$, the variance of the random error, $$e_{ij}$$ (related to the MSW), and $$\sigma^2_a$$, the variance of the random effects, $$a_i$$ (related to both MSW and MSB). Now $$\hat{\sigma}^2_a$$, the unbiased estimate of $$\sigma^2_a$$, is calculated using both MSW and MSB and, accordingly, may take on negative values. Intuitively it seems that a variance should never be negative and various investigators have tried to interpret the negative case, but usually end up by merely ignoring it as spurious and confining themselves to the positive values.

In this study the Bayesian approach was used. In order to do so it was necessary to assume a prior function. So we subjectively chose a prior, $$p(\mu, \sigma^2, \sigma^2_a)$$, in which $$\mu$$ is diffuse, and assume further that $$\mu$$ is effectively independent of $$(\sigma^2, \sigma^2_a)$$. Various arbitrary assumptions were further made on the distribution of the
prior. The effects of these assumptions on the posterior distribution were examined for several specific forms of the prior.

The posterior distribution is obtained by multiplying the prior by the likelihood function. The likelihood function, \( L(\mu, \sigma^2, \sigma^2_\alpha) \), was obtained by using the Edgeworth series, rearranged so as to collect SSW and SSB into separate terms. It turned out that only three terms contained these quantities so that further terms could be neglected for the purpose of the analysis.

Posterior distributions were obtained for \( \sigma^2 \), \( \sigma^2_\alpha \), and \( \tau_\alpha = \sigma^2_\alpha / \sigma^2 \). These were examined in detail.

7.3 THE RESULTS

7.3.1 The Variance, \( \sigma^2 \), of The Random Error, \( e_i \)

In the classical analysis of variance the unbiased estimate of \( \sigma^2 \) is not affected by the SSB. However, when using the Bayesian approach the density function of \( \sigma^2 \) is found to be considerably affected by the SSB as well as by the SSW. To investigate this effect the series expansion of the density function of \( \sigma^2 \) is examined for the effect of SSB and \( \gamma_2 \), the parameter of kurtosis of the distribution of the \( a_i \).

When the likelihood function is sharp we use \( (\sigma^2)^{-1} \) as the prior, that is, a stable estimating prior. It is then found that

\[ \sigma^2 \] is distributed as \( \chi^2_{SSW} \), the same as in the case of the
sampling approach.

If we assume further that $\sigma^2$ and $\sigma^2_a$ are independently distributed with no restrictions on the distribution of $\sigma^2$ but with $(\sigma^2_a)^{-1}$ distributed as a gamma distribution with parameters $\lambda_a/2$ and $G_a/2$, which are chosen subjectively, we consider four possibilities; $SSB = 0$, $\gamma_1 = 0$, $\gamma_2 < 0$, and $\gamma_2 > 0$.

If $SSB = 0$, then the information about the non-normality of the distribution of the $a_i$ is not used, i.e. the $a_i$ may have any distribution. If the kurtosis parameter, $\gamma_2$, of the distribution of the $a_i$ is zero, the result reduces to that found for a normal distribution of the $a_i$. When $\gamma_2 < 0$, the third term of the posterior density function of $\sigma^2$ is dominant and there is a good possibility of obtaining a sharp posterior density function,

$$f(\sigma^2), \text{ whose mode at } \frac{SSB}{I+3} - \frac{J Ca}{\lambda_a + 2}$$

terms of the gamma distribution. When $\gamma_2 > 0$, it is again the third term of the posterior density function of $\sigma^2$ which is dominant and $f(\sigma^2)$ is again sharp with mode at $\frac{SSB}{I+3} - \frac{J Ca}{\lambda_a + 2}$.

A final case studied is that of a non-informative prior distribution obtained by the method of Jeffreys and used extensively by Tiao and Tan. When this was considered it was unfortunately found that the results are disturbing and cannot be used.
7.3.2 The Variance, $\sigma^2$, of The Random Effect, $a_i$

Although mathematically it is always probable that the unbiased estimate of $\sigma^2$ arising from the sampling theory may take on a negative value, the theory is unable to explain or to interpret these negative values. In the Bayesian approach we try to explain this situation. We use the same type of prior as when investigating the distribution of $\sigma^2$ and a marginal likelihood function truncated to those three terms containing $SSB$, $SSW$, and $\gamma_2$ (the parameter of kurtosis).

Using this approach, the first case considered is that of $\gamma_2 = 0$, the normal case. Under the approximation of integration the results are found to be essentially the same as those of Hill, with only a slight change in mode and mean. Since the marginal likelihood function may be negative over parts of its domain, it is truncated to omit those parts and we try to understand the sharpness of the posterior density function in terms of the point of this truncation.

The next case is that when $SSB$ is at or near zero. It is found that the marginal likelihood function is flat so that the posterior density function is solely dependent upon the prior density function. This result is the same as that for the normal case.

Turning to the non-normal case, and when $I$ (the number of groups) is large, the coefficient of the third term of $p'(\sigma^2)$ is quite dominant and the result is the same as for the normal
case except that the first parameter is increased to \((I + 1)/2\) from \((I - 3)/2\). When SSB is not near zero and \(I\) is not large, the marginal likelihood function \(p'(\alpha_x^2)\) is the sum of three terms each containing SSB, SSW, \(I\), \(J\), and \(\gamma_2\), with none of the terms being dominant.

In the final case, the method of Jeffreys is used for a non-informative prior. When \(\alpha_x^2\) is negative the posterior density function of \(\alpha_x^2\) is flat, that is, does not give any information. But when SSB and \(I\) do not take extreme values the posterior density function of \(\alpha_x^2\) is the same as the marginal likelihood function \(p'(\alpha_x^2)\).

Briefly, then, the effect of negative \(\alpha_x^2\) is as follows. A negative \(\alpha_x^2\) leads to a likelihood factor with a maximum at \(\alpha_x^2 = 0\) and monotonically decreasing in \(\alpha_x^2\). The overall impact of such data is thus to give relatively more weight to small \(\alpha_x^2\) in the posterior than in the prior. But as \(\alpha_x^2\) becomes very large and negative the likelihood factor or the marginal likelihood function becomes more and more a straight line and the posterior density function is proportional to the prior itself. In such a case the data give no information about \(\alpha_x^2\). In short, the experiment gives no information at all.

7.3.3 The Ratio of Variances \(\tilde{c} = \sigma_x^2 / \sigma^2\)

The final operation in the analysis of variance is the application of the F test. We therefore investigate the distribution of the ratio \(\tilde{c} = \sigma_x^2 / \sigma^2\) (which leads to F) under non-normal conditions and various magnitudes of SSB and SSW. We
use the same prior as before and the likelihood functions of this ratio by a change of variable and integrating out $\sigma^2$.

It is somewhat difficult to derive general inferences from the distribution of $T^2 = \frac{\sigma^2}{\sigma^2}$. However, by assuming that $\lambda_\alpha = 0$, $\mu_\alpha = 0$, and approximating $(T^2)_{\text{in}}$ by $(T^2 + 1/j)^{-1}$ we obtain the same result as when the Jeffreys' prior is chosen.

The posterior density function of $T^2$ so obtained is composed of three terms, each of which depends on $SSW$, $SSB$, $I$, $J$, and $\gamma_2$. We find that the first term is the same as that obtained when a normal distribution is assumed. For the normal case we have examined examined $MSB \leq MSW$, $MSW/MSB \approx 1$, and $MSW/MSB \to \infty$.

For the non-normal case, $\gamma_2 \neq 0$, each term is an F density function truncated from below at $\frac{MSW}{MSB} \left( \frac{(MSW)(I+1)}{SSB} \right)$ and $\frac{(MSW)(I+3)}{SSB}$ respectively. When $MSW/MSB \approx 1$, each of the three F density functions has its maximum at the point of truncation and is monotonically decreasing to the right. When $MSW/MSB \to \infty$, we find that $g^{*}(T^2)$, the posterior density function of $T^2$, is flat. Therefore $g^{*}(T^2)$ is non-informative. In other words, the information on non-normality is not available and the data give no information about $T^2$ (i.e. the posterior density function is flat).

7.3.4. Summary

Overall, we have studied three posterior density functions, each of which is composed of three terms. We have found that
when I (the number of groups) is quite large, the coefficient of the third term, in each case, is quite dominant. When I is large, the effect of \( \gamma_2 \) (the parameter of kurtosis) is insignificant and the forms of the three density functions obtained are essentially the same as for the normal case. When \( \sigma_\alpha^2 \ll 0 \) all three posterior functions show us that the experiment is non-informative.

When I is small, i.e. less than 10, \( \gamma_2 \), the parameter of kurtosis, plays an important role.

7.4 Further Questions

We have treated only the case when the \( a_i \) are non-normally distributed and the \( e_{ij} \) are normally distributed. The reverse case, when the \( a_i \) are normally distributed and the \( e_{ij} \) are non-normally distributed, remains to be studied. Furthermore, the distribution of the \( a_i \) has been restricted to the class of exponential power functions, i.e. their distribution is symmetric. It would seem that there should be no great difficulty in applying the same method to the case of non-symmetric density functions. For different prior functions, similar analyses could be developed.
REFERENCES

(in order of citation)

[1] Bayes, T. R.,
An Essay towards Solving a Problem in the Doctrine of Chances,

[2] Bross, I.,
Fiducial Intervals for Variance Components,
Biometrics, 6 (1950) pp. 136-144.

[3] Bulmer, M. G.,
Approximate Confidence Limits for Components of Variance,
Biometrika, 44 (1957) pp. 159-167.

A Comparison of Three Different Procedures for Estimating Variance Components,

[5] Crump, S. L.,
Estimation of Variance Components in the Analysis of Variance,
[6] Crump, E. L.,
The Present Status of Variance Component Analysis,

[7] Daniels, H. E.,
The Estimation of Components of Variance,

[8] Fisher, R. A.,
The Fiducial Argument in Statistical Inference,

[9] Green, J. R.,
A Confidence Interval for Variance Components,

[10] Healy, M. J. R.,
Fiducial Limits for a Variance Component,

[11] Herbach, L. H.,
Properties of Model II-Type Analysis of Variance Tests,
[12] Thompson, W. A., Jr.,
The Problem of Negative Estimates of Variance Components,

[13] Scheffe, H.,
The Analysis of Variance,

[14] Tiw, G. C. and Tan W. Y.,
Bayesian Analysis of Random-Effect Models in the Analysis of Variance,

[15] Hill, B. M.,
Inference About Variance Components in the One-Way Model,

[16] Pearson, E. S.,
The Distribution of Frequency Constants in Small Samples from Symmetric Populations,

[17] Pearson, E. S.,
Some Notes on Sampling Tests with Two Variables,
[18] Geary, R. C.,
The Distribution of "Student's" Ratio for Non-Normal Samples,

[19] Gayen, A. K.,
The Distribution of "Student's" - t in Random Sample
of Any Size Drawn from Non-Normal Universe,

[20] Gayen, A. K.,
The Distribution of the Variance Ratio in Random
Samples of Any Size Drawn from a Non-Normal Universe,

[21] Box, G. E. P. and Tiao, G. C.,
A Bayesian Approach to the Importance of Assumptions
Applied to the Comparison of Variances,

[22] Diananda, P. H.,
Note on Some Properties of Maximum Likelihood Estimates,

[23] Box, G. E. P.,
A Note on Regions for Tests of Kurtosis,
[24] Turner, M. C.,
On Heuristic Estimation Methods,
Biometrics, 16 (1960) pp. 299-301.

The Use of OLUMV Estimates in Inference Robustness Studies of the Location Parameter of a Class of Symmetric Distributions,

[26] Jeffreys, H.,
Theory of Probability, third edition,

[27] Jeffreys, H. and Swirlee, B.,
Methods of Mathematical Physics,
REFERENCES AND AUTHOR INDEX

(Authors Alphabetically)

[1] Bayes,
[23] Box,
[21] Box and Tiao,
[2] Bros,  
[3] Bulmer,
[5] Crump,
[6] Crump,
[7] Daniels,
[22] Dianada,
[8] Fisher,
[19] Gayen,
[20] Gayen,
[18] Geary,
[9] Green,
[10] Healy,
[11] Herbach,
[15] Hill,
[26] Jeffreys,
[27] Jeffreys and Swirls,

[16] Pearson,
[17] Pearson,

[13] Scheffe,

[12] Thompson,
[25] Tiao and Lund,
[14] Tiao and Tan,
Sources Consulted but Not Cited

Abramowitz, M. and Stegun, I. A.,
Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,

Ali, M. M.,
Some Aspects of the One-Way Random-Effects Model and the Linear Model with Two Random Components,

Fartlett, M. S.,
The Effect of Non-Normality of the t-Distribution,

Birnbaum, A.,
On the Foundation of Statistical Inference,

Box, G. E. P.,
Non-Normality and Test on Variances,
Biometrika, 40 (1953a) pp. 318-335.
Box, G. E. P. and Anderson, S. L.,
Permutation Theory in the Derivation of Robust Criteria
and the Study of Departures from Assumptions,

Box, G. E. P. and Tiao, G. C.,
A Further Look at Robustness via Bayes' Theorem,
Biometrika, 49 (1962) pp. 419-432.

Box, G. E. P. and Tiao, G. C.,
A Note on Criterion Robustness and Inference Robustness,

Box, G. E. P. and Tiao, G. C.,
Multivariate Problems from a Bayesian Viewpoint,

Box, G. E. P. and Tiao, G. C.,
Bayesian Estimation of Means for the Random-Effect
Model,

Cochran, W. G. and Cox, G. M.,
Experimental Designs, second edition,
Cornish, E. A.,

The Multivariate t-Distribution Associated with a Set of
Normal Sample Deviates,

Cramér, H.,

Mathematical Methods of Statistics,

Daniels, H. E.,

Saddlepoint Approximations in Statistics,

David, F. N. and Johnson, N. L.,

The Effect of Non-Normality on the Power Function of the
F-Test in the Analysis of Variance,
Biometrika, 38 (1951) pp. 43-57.

De Pruijn, N. G.,

Asymptotic Methods in Analysis,
Amsterdam, North-Holland, 1961.

Dunnett, C. W. and Sobel, M.,

A Bivariate Generalization of Student's t-Distribution
with Tables for Certain Special Cases,
Cayen, A. K.,
The Significance of the Difference Between the Means of
Two Non-Normal Samples,

Cayen, A. K.,
The Frequency Distribution of the Product-Normal Correlation
Coefficient in Random Samples of Any Size Drawn from
Non-Normal Universes,

Geary, R. C.,
Testing for Normality,

Hill, B. M.,
Correlated Errors in the Random Model,

Johnson, R. A.,
An Asymptotic Expansion for Posterior Distributions,

Johnson, R. A.,
Asymptotic Expansions Associated with Posterior
Distributions,
Kempthorne, O.,
The Design and Analysis of Experiments,

Kendall, M. G. and Stuart, A.,
The Advanced Theory of Statistics,
vol. 2, second edition,

Lindley, D. V.,
Introduction to Probability and Statistics from a
Bayesian Viewpoint,

Lindley, D. V.,
The Use of Prior Probability Distributions in Statistical
Inference and Decisions,

Milne-Thomson, L. M.,
The Calculus of Finite Differences,

Pearson, E. S.,
The Analysis of Variance in Cases of Non-Normal Variation,
Biometrika, 23 (1931) pp. 114-133.
Pearson, E. S. and Hartley, H. O.,
Biometrika Tables for Statisticians, vol. 1,

Pearson, K.,
Tables of the Incomplete Beta-Function,

Raiffa, H. and Schlaifer, R.,
Applied Statistical Decision Theory,

Savage, L. J.,
The Foundations of Statistics, second edition,
New York, Dover, 1954.

Savage, L. J.,
The Foundations of Statistics Reconsidered,

Tiao, G. C. and Ali., M. M.,
Analysis of Correlated Random Effects: Linear Model with Two Random Components,

Tiao, G. C. and Box, G. E. P.,
Bayesian Analysis of a Three-Component Hierarchical Design Model,
Tiao, G. C. and Tan, W. Y.,
Bayesian Analysis of Random-Effect Models in the Analysis of Variance II: Effects of Autocorrelated Errors,

Tiao, G. C. and Zellner, A.,
Bayes' Theorem and the Use of Prior Knowledge in Regression Analysis,
Biometrika, 51 (1964b) pp. 219-230.

Wilks, S. S.,
Mathematical Statistics,

Yuen, Karen K.,
The Two-Sample Trimmed t for Unequal Population Variances,

Zellner, A.,
An Introduction to Bayesian Inference in Econometrics,
VITA AUCTORIS

Born: — Mokpo, Jeonnaem, Korea;
1944 November 26.

Schooling:
1957 Mokpo Bookkyo Elementary School, Korea.
1960 Mokpo Middle School, Korea.
1963 Jeonju Youngsang High School, Korea.

University Degrees:
1969 B.Sc. Chonpuk (Jeonbug) National University, Korea.

Military Service:
October 1965 to April 1966, military training in the
Republic of Korea Armed Forces.

Employed as teacher of Mathematics at the Chonju Keijon
Girl's High School in Chonju, Korea.

Employed as Visiting Assistant Professor of Statistics,
Memorial University of Newfoundland, St. John's,
Newfoundland, since September 1980.

Married in 1969 Miss Ok-Ja Kim; which union has since been
blessed with two daughters, Jean-Aha and Shun-Aha, and
one son, Charm-Aha.

- 0 0 0 -