Abstract

HUDSON-CURTIS, BUFFY L. GENERALIZATIONS OF THE MULTIVARIATE LOGISTIC DISTRIBUTION WITH APPLICATIONS TO MONTE CARLO IMPORTANCE SAMPLING. (Under the direction of Drs. John Monahan and Leonard Stefanski.)

Monte Carlo importance sampling is a useful numerical integration technique, particularly in Bayesian analysis. A successful importance sampler will mimic the behavior of the posterior distribution, not only in the center, where most of the mass lies, but also in the tails (Geweke, 1989).

Typically, the Hessian of the importance sampler is set equal to the Hessian of the posterior distribution evaluated at the mode. Since the importance sampling estimates are weighted averages, their accuracy is assessed by assuming a normal limiting distribution. However, if this scaling of the Hessian leads to a poor match in the tails of the posterior, this assumption may be false (Geweke, 1989). Additionally, in practice, two commonly used importance samplers, the Multivariate Normal Distribution and the Multivariate Student-\(t\) Distribution, do not perform well for a number of posterior distributions (Monahan, 2000).

A generalization of the Multivariate Logistic Distribution (the Elliptical Multivariate Logistic Distribution) is described and its properties explored. This distribution outperforms the Multivariate Normal distribution and the Multivariate Student-\(t\) dis-
tribution as an importance sampler for several posterior distributions chosen from the literature. A modification of the scaling by Hessians of the importance sampler and the posterior distribution is explained. Employing this alternate relationship increases the number of posterior distributions for which the Multivariate Normal, the Multivariate Student-$t$, and the Elliptical Multivariate Logistic Distribution can serve as importance samplers.
GENERALIZATIONS OF THE MULTIVARIATE LOGISTIC DISTRIBUTION WITH APPLICATIONS TO MONTE CARLO IMPORTANCE SAMPLING

by

BUFFY L. HUDSON-CURTIS

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in

STATISTICS

in the

GRADUATE SCHOOL

at

NC STATE UNIVERSITY

2001

____________________  ______________________
Dr. J. Monahan, Co-Chair    Dr. L. Stefanski, Co-Chair

____________________  ______________________
Dr. D. Boos               Dr. R. Berger
GENERALIZATIONS OF THE MULTIVARIATE LOGISTIC DISTRIBUTION WITH APPLICATIONS TO MONTE CARLO IMPORTANCE SAMPLING

Copyright 2001

by

BUFFY L. HUDSON-CURTIS
To Mark and Heidi.
Biography

Buffy L. Hudson-Curtis was born in Fayetteville, North Carolina on January 3, 1976. She graduated from Lincoln East High School in Lincoln, Nebraska in 1993 and received her Bachelor of Science degree in Actuarial Science and Mathematics from the University of Nebraska-Lincoln in 1996. During the last semester of her undergraduate career she decided to pursue a graduate degree in statistics, and received a Master of Statistics degree in 1998 from North Carolina State University. In 2001 Buffy completed the requirements for a Ph.D. in the Department of Statistics at North Carolina State University.
Acknowledgements

Thank you to my co-chairs Drs. John Monahan and Leonard Stefanski for all of their guidance and assistance. After stating that I had no interest in studying the field of statistical computing, they have taught me to never say never! Thank you to my committee members Drs. Dennis Boos and Roger Berger for their suggestions and instruction during my graduate career.

To my fellow students Jennifer Shannon Gauvin, Steve Novick, and Wendy Czika, thank you for your friendship and support (as well as the late nights spent studying for the Basic and Preliminary Written Exams). And to my undergraduate advisor and friend, Steve Linney, thank you for helping me discover what I wanted to do with my life.

Thank you to my family. To my parents, Berman Hudson and Sharon Pullen for giving me a love of learning. To my sister, Lori, for listening. And to my husband, Mark, and daughter, Heidi, for their love, support, patience, and understanding.
Contents

List of Figures viii

List of Tables xvi

1 Introduction 1
  1.1 Motivation .......................................................... 1
  1.2 Monte Carlo Importance Sampling ............................... 2
  1.3 Choosing an Importance Sampler .................................... 3
    1.3.1 Illustrative Example .......................................... 5
    1.3.2 Common Importance Samplers ................................. 8
  1.4 Multivariate Normal ................................................. 12
  1.5 Multivariate Student-$t$ ............................................. 13

2 Characterizations of the Multivariate Logistic Distribution 15
  2.1 Elliptical Multivariate Logistic Distribution (EMVL) ........... 16
    2.1.1 Generating from the EMVL .................................... 17
    2.1.2 Evaluating the Density Function ............................. 20
    2.1.3 Matching the Posterior and the EMVL ......................... 22
    2.1.4 Tail Behavior of the Marginal Distributions of the EMVL ... 24
  2.2 Simulation: Using the EMVL as an Importance Sampler for Dirichlet Posterior Distributions ............................... 30
    2.2.1 Evaluating the Weights ....................................... 33
    2.2.2 Unbiasedness .................................................. 43
    2.2.3 Error Estimates ............................................... 45
  2.3 Other Generalizations Considered .................................. 48
    2.3.1 Classical Multivariate Logistic Distribution .................. 48
    2.3.2 Multivariate Logistic Distribution 1 $MVL_{1}$ ................ 50
    2.3.3 Alternate Generalization .................................... 52
3 Simulation: Using the EMVL as an Importance Sampler for Posterior Distributions From the Literature

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Examples</td>
<td>57</td>
</tr>
<tr>
<td>3.1.1 Example 1 (d=3)</td>
<td>57</td>
</tr>
<tr>
<td>3.1.2 Example 2 (d=7)</td>
<td>57</td>
</tr>
<tr>
<td>3.1.3 Examples 3, 4, and 5 (d=3, 5, and 8)</td>
<td>58</td>
</tr>
<tr>
<td>3.1.4 Examples 6 and 7 (d=7 and 7)</td>
<td>58</td>
</tr>
<tr>
<td>3.1.5 Example 8, (d=9)</td>
<td>59</td>
</tr>
<tr>
<td>3.1.6 Example 9, d=10</td>
<td>59</td>
</tr>
<tr>
<td>3.1.7 Example 10, d=11</td>
<td>60</td>
</tr>
<tr>
<td>3.2 Results of Simulation Experiment</td>
<td>60</td>
</tr>
<tr>
<td>3.2.1 Evaluation of the Weights as a Function of the Radius</td>
<td>61</td>
</tr>
<tr>
<td>3.2.2 Evaluation of the Existence of the First and Second Moments of the Weights</td>
<td>73</td>
</tr>
<tr>
<td>3.2.3 Relative Standard Errors</td>
<td>82</td>
</tr>
<tr>
<td>3.3 Summary</td>
<td>83</td>
</tr>
</tbody>
</table>

4 Effect of Scale Factor

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Selecting ( c )</td>
<td>86</td>
</tr>
<tr>
<td>4.2 Evaluating Consistency of Selection</td>
<td>87</td>
</tr>
<tr>
<td>4.3 Results Produced When ( c ) is Altered for the EMVL</td>
<td>100</td>
</tr>
<tr>
<td>4.4 Recommendations for Selecting ( c )</td>
<td>113</td>
</tr>
<tr>
<td>4.5 Changing Scale with MVT and the MVN</td>
<td>114</td>
</tr>
<tr>
<td>4.6 Detailed Analysis of Example 8</td>
<td>135</td>
</tr>
<tr>
<td>4.7 Summary</td>
<td>139</td>
</tr>
</tbody>
</table>

A Fortran Code

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Fortran code for Example 1</td>
<td>144</td>
</tr>
<tr>
<td>A.2 Fortran code for Example 2</td>
<td>148</td>
</tr>
<tr>
<td>A.3 Fortran code for Example 3</td>
<td>151</td>
</tr>
<tr>
<td>A.4 Fortran code for Example 4</td>
<td>153</td>
</tr>
<tr>
<td>A.5 Fortran code for Example 5</td>
<td>155</td>
</tr>
<tr>
<td>A.6 Fortran code for Examples 6 and 7</td>
<td>158</td>
</tr>
<tr>
<td>A.7 Fortran code for Example 8</td>
<td>164</td>
</tr>
<tr>
<td>A.8 Fortran code for Example 9</td>
<td>166</td>
</tr>
<tr>
<td>A.9 Fortran code for Example 10</td>
<td>168</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Posterior Density, $p^*(t)$, and Importance Sampling Densities, $g_1(t)$ and $g_2(t)$ .................................................. 7
1.2 Corresponding Importance Sampling Weight Functions. A horizontal reference line is given at 1. The weights, $w_1$, correspond to the weights produced using $g_1$ as the importance sampling density and $w_2$ corresponds to the weights produced using the importance sampling density $g_2$. .................................................. 7
1.3 Importance sampling weights produced using four different importance samplers (dimension=3). The top left plot results from an importance sampler with tails that decay more slowly than those of $p^*(t)$ in all directions. the top right plot results from an importance sampler that matches the $p^*(t)$ exactly. The bottom left plot results from an importance sampler whose tails decay more quickly than those of the $p^*(t)$ in one dimension. The bottom right plot results from an importance sampler whose tails decay more slowly than those of the $p^*(t)$ in all three dimensions. See Section 1.3.1 for more details. .................. 9
2.1 Comparison of the marginal density estimates of the EMVL (d=3) produced by fitting Laguerre polynomials (dashed line) and using Simpson’s rule of integration (solid line). ................................. 30
2.2 Comparison of the marginal density estimates of the EMVL (d=5) produced by fitting Laguerre polynomials (dashed line) and using Simpson’s rule of integration (solid line). ................................. 31
2.3 Comparison of the marginal density estimates of the EMVL (d=10) produced by fitting Laguerre polynomials (dashed line) and using Simpson’s rule of integration (solid line). ................................. 31
2.4 Importance sampling weights using the posterior distributions from Problems 1 (top figure) and 2 (bottom figure), untransformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $\|v\|$. 35
2.5 Importance sampling weights using the posterior distributions from Problems 3 (top figure) and 4 (bottom figure), untransformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $||v||$. 36

2.6 Importance sampling weights using the posterior distributions from Problems 5 (top figure) and 6 (bottom figure), Untransformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $||v||$. 37

2.7 Importance sampling weights using the posterior distributions from Problems 1 (top figure) and 2 (bottom figure), transformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $||v||$. 38

2.8 Importance sampling weights using the posterior distributions from Problems 3 (top figure) and 4 (bottom figure), transformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $||v||$. 39

2.9 Importance sampling weights using the posterior distributions from Problems 5 (top figure) and 6 (bottom figure), transformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $||v||$. 40

2.10 Importance sampling weights using the posterior distributions from Problems 2 (top figure) and 6 (bottom figure), untransformed Dirichlet and the MVN as the importance sampler. The horizontal axes are $||v||$. 41

2.11 Importance sampling weights using the posterior distributions from Problems 1 (top figure) and 5 (bottom figure), transformed Dirichlet and the MVN as the importance sampler. The horizontal axes are $||v||$. 42

2.12 90% confidence intervals and horizontal reference lines for the equivalence interval $(0.995, 1.005)$ is given. If the 90% confidence interval is completely contained within the equivalence interval, one concludes that the estimates are equivalent to 1 with 95% confidence. . . . . . . . . . . . . . 45

3.1 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 1. The horizontal axes are $||v|| = ||L^{-1}(t - \hat{t})||$. A trend of weights that increase as $||v||$ increases is an indication that the importance sampling estimates may not be asymptotically normal. 63

3.2 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 2. The horizontal axes are $||v|| = ||L^{-1}(t - \hat{t})||$. A trend of weights that increase as $||v||$ increases is an indication that the importance sampling estimates may not be asymptotically normal. 64
3.3 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 3. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.

3.4 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 4. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.

3.5 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 5. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.

3.6 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 6. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.

3.7 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 7. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.

3.8 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 8. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.

3.9 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 9. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
3.10 Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 10. The horizontal axes are $||v|| = ||L^{-1}(t - \bar{t})||$. A trend of weights that increase as $||v||$ increases is an indication that the importance sampling estimates may not be asymptotically normal.

3.11 Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 1, the bottom to Example 2. A horizontal reference line is drawn at the value of $\hat{\beta}$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).

3.12 Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 3, the bottom to Example 4. A horizontal reference line is drawn at the value of $\hat{\beta}$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).

3.13 Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 5, the bottom to Example 6. A horizontal reference line is drawn at the value of $\hat{\beta}$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).

3.14 Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 7, the bottom to Example 8. A horizontal reference line is drawn at the value of $\hat{\beta}$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).

3.15 Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 9, the bottom to Example 10. A horizontal reference line is drawn at the value of $\hat{\beta}$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).

4.1 Estimates of $\beta$ and coefficient of variation calculated for Example 1 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \big|_{t=\bar{t}}$. .

4.2 Estimates of $\beta$ and coefficient of variation calculated for Example 2 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \big|_{t=\bar{t}}$. .
4.3 Estimates of $\alpha$ and coefficient of variation calculated for Example 3 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t_\text{opt}}$.

4.4 Estimates of $\alpha$ and coefficient of variation calculated for Example 4 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t_\text{opt}}$.

4.5 Estimates of $\alpha$ and coefficient of variation calculated for Example 5 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t_\text{opt}}$.

4.6 Estimates of $\alpha$ and coefficient of variation calculated for Example 6 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t_\text{opt}}$.

4.7 Estimates of $\alpha$ and coefficient of variation calculated for Example 7 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t_\text{opt}}$.

4.8 Estimates of $\alpha$ and coefficient of variation calculated for Example 8 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t_\text{opt}}$.

4.9 Estimates of $\alpha$ and coefficient of variation calculated for Example 9 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t_\text{opt}}$.

4.10 Estimates of $\alpha$ and coefficient of variation calculated for Example 10 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t_\text{opt}}$. 

4.11 Example 1: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 103

4.12 Example 2: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 104

4.13 Example 3: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 105

4.14 Example 4: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 106

4.15 Example 5: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 107

4.16 Example 6: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 108

4.17 Example 7: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 109

4.18 Example 8: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 110

4.19 Example 9: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 111

4.20 Example 10: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\). ................................. 112
4.21 Example 1: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.22 Example 2: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.23 Example 3: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.24 Example 4: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.25 Example 5: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.26 Example 6: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.27 Example 7: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.28 Example 8: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.29 Example 9: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).

4.30 Example 10: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).
4.31 Example 1: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (5 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).  

4.32 Example 3: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (5 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).  

4.33 Example 4: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (5 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).  

4.34 Example 5: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (5 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).  

4.35 Example 1: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (3 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).  

4.36 Example 3: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (3 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).  

4.37 Example 4: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (3 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).  

4.38 Example 5: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (3 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).
## List of Tables

2.1 Evaluation of Marginal Density Estimates for $d = 1$ .................. 27  
2.2 Estimated Fourier Coefficients for dimensions 1, 3, 5 and 10 resulting from fitting Laguerre polynomials to the marginal densities of the EMVL. 29  
2.3 Dirichlet Distribution Parameters ................................. 32  
2.4 Entries are the $t$-statistics for testing the hypotheses $H_0 : E(Z_i) = 1$ and $H_1 : E(Z_i) \neq 1$ ............................... 44  
2.5 Ratios of $RM_{E/(\sum_{i=1}^{n}(Z_i - \bar{Z})^2/n}$ .............................. 46  
2.6 Lack of Coverage of 95% Confidence Intervals. Entries represent the proportion of times that a 95% confidence interval constructed using importance sampling estimates that failed to contain the true value of 1. 47  
2.7 Results From Importance Sampling Using the EMVL ................. 48  
3.1 Estimates of $\beta$ for Examples 1-10 ............................... 75  
3.2 Relative standard errors of the importance sampling estimates of the posterior distributions for Examples 1-10. Relative standard errors were not given when empirical evidence suggested that the importance sampling estimates were not asymptotically normal. ................................. 83  
4.1 Resulting counts of the values of $c$ chosen from 100 simulations for each example, using the EMVL as the importance sampler.............. 89  
4.2 Table of $\hat{\beta}$, results of test of hypothesis of infinite variance, and relative standard errors produced when an alternate scale was used to match the importance sampler, the EMVL, to the posterior distributions. For comparison, relative standard errors that resulted using the usual scale, $c = 2$ are also given. Relative standard errors were not reported if there was an indication that the importance sampling estimates were not asymptotically normal. ................................. 102  
4.3 Hypothesis test for infinite variance of the importance sampling weights for importance samplers MVN, MVT (5 degrees of freedom), MVT (3 degrees of freedom) using an alternate scale...................... 115
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4</td>
<td>Comparison of relative standard errors for Examples 1-10</td>
<td>116</td>
</tr>
<tr>
<td>4.5</td>
<td>Frequency of visits by length of stay for 132 long-term schizophrenic patients (Wing, 1962)</td>
<td>139</td>
</tr>
<tr>
<td>4.6</td>
<td>Importance Sampling Posterior Means and Corresponding Standard Errors for Example 8</td>
<td>140</td>
</tr>
<tr>
<td>4.7</td>
<td>Importance Sampling Cell Count Estimates for Example 8. The difference between these estimates and the actual cell counts given by Wing (1962) are shown in parenthesis.</td>
<td>140</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation

In Bayesian statistics one often encounters the situation in which a random vector of interest, \( \mathbf{t} \subset \Theta \), varies according to a unnormalized posterior distribution, \( p^*(\mathbf{t}) \), with a unique form whose properties are only known by its evaluation at abscissas \( \mathbf{t} \). In this situation one frequently wishes to estimate some functional of \( \mathbf{t} \) - most commonly \( E_{p^*}[\mathbf{t}] \) (the mean), and \( E_{p^*}[\mathbf{tt}'] \) (the covariance matrix). Other expectations that may be of interest include the normalization constant of the posterior distribution and indicator functions. There are a number of techniques available to calculate

\[
E_{p^*}\{h(\mathbf{t})\} = \frac{\int h(\mathbf{t})p^*(\mathbf{t})d\mathbf{t}}{\int p^*(\mathbf{t})d\mathbf{t}},
\]

(1.1)
where $h(t)$ is some function of $t$. One typically considers analytical solutions first. However, the number of posterior distributions for which one can integrate Equation 1.1 analytically is considerably few, especially in higher dimensions.

This leads to the consideration of numerical methods. Monte Carlo simulation is one of the most commonly used numerical methods for integration. In both small and large dimensions, Monte Carlo simulation is an option that permits integration over a broad range of posterior distributions. Random vectors, $T^{(i)}$, $i = 1, \ldots, n$ are generated from the distribution $p^*(t)$. The average $n^{-1} \sum_{i=1}^{n} h(T^{(i)})$ is used to estimate $E_{p^*}[h(t)]$. This technique is feasible when it is easy to generate random variables from $p^*(t)$. However, in practice one rarely is able to generate from $p^*(t)$, and when it is possible, $E_{p^*}[h(t)]$ usually is known (Monahan, 2000). An alternative numerical method to consider is Monte Carlo Importance Sampling.

### 1.2 Monte Carlo Importance Sampling

This technique is similar to Monte Carlo integration, and can be used to integrate over unimodal, smooth posterior distributions. A sequence of $n$ random variables, $T^{(1)}, T^{(2)}, \ldots, T^{(n)}$ is generated from some density, $g(t)$, called the importance sampling density, or importance sampler. These generated values are used to calculate a weighted average of the $h(T^{(i)})$. One can then estimate $E_{p^*}[h(t)]$ by

$$
\bar{h}_n = E_g \{h(t)w(t)\} = n^{-1} \sum_{i=1}^{n} h(T^{(i)})w(T^{(i)}),
$$

(1.2)
where $\mathbf{T}^{(i)}$ is the $i^{th}$ vector generated from $g(t)$, and the function $h(t)$ evaluated at $\mathbf{T}^{(i)}$ is given weight equal to $w(\mathbf{T}^{(i)})$. The weights are a ratio of the posterior and importance sampling densities, so that $w(\mathbf{T}^{(i)}) = p^*(\mathbf{T}^{(i)})/g(\mathbf{T}^{(i)})$.

The use of Equation 1.2 as an estimate of $E_{p^*}[h(t)]$ follows from the fact that

$$E_g[h(t)w(t)] = \int h(t)[\frac{p^*(t)}{g(t)}]g(t)dt$$

$$= \int h(t)p^*(t)dt$$

$$= E_{p^*}[h(t)],$$

where $w(t) = p^*(t)/g(t)$.

### 1.3 Choosing an Importance Sampler

To be effective, an importance sampler should have certain properties. It should be relatively easy to generate random variables from $g(t)$. In order to calculate the weights, one should be able to evaluate the density of $g(t)$ up to a constant. In addition, one would like the density of $g(t)$ to be similar to $p^*(t)$.

It would also be useful to choose an importance sampler that results in importance sampling estimates that have a normal limiting distribution. It has been shown by Kloek and van Dijk (1987) and Geweke (1989) that $\bar{h}_n$ converges almost surely to $E_{p^*}[h(t)]$ if:

1. $p^*(t)$ is proportional to a proper probability function, defined on $\Theta$;
2. \(\{T^{(i)}\}_{i=1}^{\infty}\) is a sequence of iid random vectors, the common distribution having a probability distribution \(g(t)\);

3. The support of \(g(t)\) includes \(\Theta\); and

4. \(E_{\nu^*}[h(t)]\) exists and is finite.

This is basically an application of the strong law of large numbers. Based on this result, one expects \(\bar{h}_n\) to be a good estimate of \(E_{\nu^*}[h(t)]\) for large values of \(n\).

If, in addition to these assumptions, \(E_g[w(t)]\) and \(E_g[h(t)^2w(t)]\) are finite then (Geweke, 1989)

\[
\frac{1}{n} \{\bar{h}_n - E_{\nu^*}[h(t)]\} \Rightarrow N(0, \sigma^2),
\]

where

\[
\sigma^2 = \int \{h(t) - E_{\nu^*}[h(t)]\}^2 \left[\frac{p^*(t)}{g(t)} \right] g(t) dt.
\]

One can estimate \(\sigma^2\) by

\[
\frac{1}{n} \sum_{i=1}^{n} (h(T^{(i)})w(T^{(i)}) - \bar{h}_n)^2.
\]

In practice it is often difficult, and sometimes impossible to verify that \(E_g[w(t)]\) and \(E_g[h(t)^2w(t)]\) are finite, especially when the dimension of the posterior distribution is large. In the absence of analytical verification, empirical evidence can be used to justify the use of a particular importance sampler. An importance sampler, \(g(t)\), is likely to satisfy these conditions if its tails decay at least as slowly as those of the
posterior distribution (Geweke, 1989). Since it is not necessary for the weights to be bounded (Monahan, 2001), one can attempt to verify empirically that an importance sampler decays more slowly than the posterior distribution for a sufficient distance in the tails. It should be noted that this type of verification cannot be taken as a proof that the importance sampling estimates are asymptotically normal, but rather as informal evidence that one should not reject the assumption of asymptotic normality of the importance sampling estimates.

If this condition is not satisfied, and the tails of the importance sampler decay more quickly than those of the posterior distribution, the tails of the posterior distributions will rarely be observed. The few values that are observed from the tails will dominate the sample because they will be given relatively large weights. In this situation $\overline{h}_n$ is still an unbiased estimator of Equation 1.1, but its variance may not be finite, so one cannot assume that the estimate of $E_{p^*}[h(t)]$ is asymptotically normally distributed (Geweke, 1989).

1.3.1 Illustrative Example

To see the effect of mismatching the tails of the posterior distribution, consider a one-dimensional example given by Monahan (2001). Let $p^*(t)$ be the standard normal distribution, and consider two importance samplers, $g_1(t)$, the normal distribution with mean 0 and variance 1.2, and $g_2(t)$, the normal distribution with mean 0 and variance 0.8. Figures 1.1 and 1.2 display the density functions of these three
distributions and the weights $p^*(t)/g_1(t)$ and $p^*(t)/g_2(t)$. A reference line is drawn at $w(t) = p^*(t)/p^*(t) = 1$. As can be seen in these figures, the first importance sampling density, $g_1(t)$, has tails that are thicker than $p^*(t)$, and the resulting weights decrease as one moves away from the center of the distribution. In contrast, $g_2(t)$ has thinner tails than the posterior distribution, and the weights increase without bound as one moves away from the center.

A similar example can be given in three dimensions. Let $p^*(t)$ be normally distributed with its mean at the origin, and covariance matrix

$$
\Sigma_{p^*} = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}.
$$

Consider four normally distributed importance samplers ($g_1(t)$, $g_2(t)$ and $g_3(t)$, and $g_4(t)$) with means at the origin, and covariance matrices

$$
\Sigma_g = \begin{pmatrix}
\sigma_1^2 & 0 & 0 \\
0 & \sigma_2^2 & 0 \\
0 & 0 & \sigma_3^2
\end{pmatrix},
$$

where $(\sigma_1^2, \sigma_2^2, \sigma_3^2) = (3, 2, 3)$ for $g_1(t)$, $(1, 1, 1)$ for $g_2(t)$, $(0.4, 1, 1)$ for $g_3(t)$, and $(0.3, 0.2, 0.3)$ for $g_4(t)$. Importance sampling using 2000 iterations for each of the importance samplers was used to produce plots of the weights as a function of $\|t\|$. These plots are shown in Figure 1.3. The tails of the first importance sampler, $g_1(t)$, decay more slowly than those of the posterior distribution, an ideal situation in practice, as
Figure 1.1: Posterior Density, $p^*(t)$, and Importance Sampling Densities, $g_1(t)$ and $g_2(t)$.

Figure 1.2: Corresponding Importance Sampling Weight Functions. A horizontal reference line is given at 1. The weights, $w_1$, correspond to the weights produced using $g_1$ as the importance sampling density and $w_2$ corresponds to the weights produced using the importance sampling density $g_2$. 
the weights decrease as $\|t\|$ increases. The importance sampler, $g_2(t)$, matches $p^*(t)$ perfectly. In this situation, as one might expect, all of the weights are one. This would rarely, if ever, happen in practice. The remaining two importance samplers show the problem that can arise when the tails of the importance sampler decay more quickly than those of the posterior distribution. The plot produced using $g_3(t)$ demonstrates what can happen when the importance sampler decays more quickly than the posterior distribution in one dimension. The weights fan out and increase as $\|t\|$ increases. The importance sampler $g_4(t)$ decays more quickly than the posterior distribution in all dimensions. Here the weights steadily increase as $\|t\|$ increases.

Based on these examples, one might conclude that $g(t)$ should be chosen so that its tails decay as slowly as possible. However, when the tails of the importance sampler are much thicker than those of the posterior distribution, over sampling occurs in the tails of the posterior distribution. This results in miniscule weights being given to the observations from the tails of the posterior distribution. This can be a waste of computer time and can result in only a few observations contributing to the estimate of $E_{p^*}[h(t)]$.

1.3.2 Common Importance Samplers

Since the importance sampling density should mimic the posterior density, the importance sampling distribution is often “matched” to the posterior density at the mode of the posterior distribution by setting: $\nabla \ln g_T(t) \big|_{t=t^*} = 0$ and $-\nabla^2 \ln p(t) \big|_{t=t^*} = 0$. 
Figure 1.3: Importance sampling weights produced using four different importance samplers (dimension=3). The top left plot results from an importance sampler with tails that decay more slowly than those of $p^*(t)$ in all directions. The top right plot results from an importance sampler that matches the $p^*(t)$ exactly. The bottom left plot results from an importance sampler whose tails decay more quickly than those of the $p^*(t)$ in one dimension. The bottom right plot results from an importance sampler whose tails decay more slowly than those of the $p^*(t)$ in all three dimensions. See Section 1.3.1 for more details.
\(-\nabla^2 \ln g_{\mathbf{T}}(\mathbf{t}) \big|_{\mathbf{t} = \hat{\mathbf{t}}}\), where \(\hat{\mathbf{t}}\) is the posterior mode. This is based on large sample theory that the posterior distribution converges to the multivariate normal distribution (Geweke, 1989 and Monahan, 2001), and suggests that the multivariate normal distribution would be a good choice for an importance sampling density. However, if the prior is improper, or flat for one of the parameters, the multivariate normal is not a good choice for an importance sampling density (Geweke, 1989). It should be noted that for most posterior distributions it takes relatively little computer time to calculate the posterior mode and Hessian, as opposed to calculating posterior means and variances which involve integrating over the posterior distribution.

Since it is preferable to use an importance sampler that has just slightly thicker tails than the posterior distribution, the best form for \(g(\mathbf{t})\) depends heavily on the form of the posterior distribution. Several distributions have been considered as importance samplers in the past. These include the Multivariate Normal distribution, the Multivariate Student-\(t\) distribution with small degrees of freedom (Evans and Swartz, 1995, 2000), the Multivariate Split Normal and the Multivariate Split-\(t\) distribution (Geweke, 1989).

While these distributions work well for many problems, in practice there are times when the posterior distribution of interest has tail behavior that is too heavy for the multivariate normal distribution to be successful as an importance sampler, but is not heavy enough for the Multivariate Student-\(t\) distribution. These posterior distributions have tail behavior that is best characterized as exponential in nature (Monahan,
The following sections in this chapter will briefly address how the Multivariate Normal distribution, the Multivariate Student-\(t\) distribution perform as importance samplers when the posterior distribution has tails that decay exponentially.

It should be noted that the Split Normal and Split Student \(t\) distributions are useful when the posterior density is substantially asymmetric (Geweke, 1989), but do not address the issue of exponentially decaying tails. The basic idea when using these distributions is to explore the posterior density along the axes in each direction and to find the slowest rate of decline. The importance sampling density (either the Multivariate Normal or Multivariate Student-\(t\)) is modified along each of these axes by this rate of decline. This does little to address the problem that arises when the posterior distribution has exponential tail behavior, since the tail rates of decay for the Split Normal and Split Student \(t\) distributions are the same as those of the Multivariate Normal and Multivariate Student-\(t\) respectively.

We propose using a generalization of the Multivariate Logistic Distribution as an importance sampler when it is suspected that the posterior distribution decays exponentially. Chapter 2 discusses some different generalizations of the Multivariate Logistic Distribution and their potential use as importance samplers.

Chapter 3 studies the performance of the Ellipical Multivariate Logistic Distribution (EMVL), a generalization of the Multivariate Logistic distribution, as an importance sampler using several simulation experiments. Chapter 4 addresses the situation in which the behavior of the posterior distribution at the mode does not mimic the
behavior in the tails. When this happens any importance sampler may fail to perform well (Geweke, 1989). We suggest modifying the manner in which the importance sampler is matched to the posterior distribution.

1.4 Multivariate Normal

The Multivariate Normal distribution (MVN) is one of the distributions most commonly used as an importance sampler. To describe its form, let \( X \) be a \( d \)-dimensional vector. Then \( X \) has a multivariate normal distribution (MVN) with mean \( \mu \) and variance matrix \( \Sigma = LL' \) if its probability density function is of the form:

\[
f_X(x) = \frac{1}{(2\pi)^{d/2}|L|} \exp\left(\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\right),
\]

where \( LL' \) is the Cholesky factorization of \( \Sigma \). To use \( f_X(x) \) as an importance sampler approximating the posterior density, \( \Sigma^{-1} \) is set equal to the negative Hessian of the log posterior density evaluated at the mode, and \( \mu \) is set to the mode of the posterior.

The tails of the MVN decay at a rate of \( \exp(-cx^2) \), so the MVN does not perform satisfactorily as an importance sampler when the tails of the posterior distribution decay exponentially. If the MVN is used in this situation a few observations can dominate the sample, since tails of importance sampling density are not as thick as those of the posterior distribution. When this occurs, one should not assume that the importance sampling estimate is asymptotically normal.


1.5 Multivariate Student-t

The Multivariate Student-t (MVT) distribution is another distribution that is frequently used as an importance sampler. The $d$-dimensional vector, $\mathbf{X}$ has the MVT distribution with location parameter $\mu$, scale parameter $\Sigma$, and $k$ degrees of freedom if its probability density function is proportional to:

$$f \propto \left\{1 + \frac{1}{d+k}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)\right\}^{-\frac{d+k}{2}}.$$ 

This distribution has been proposed as an alternative to the MVN for use as an importance sampler when the posterior density has heavier tails than the multivariate normal distribution. To use the MVT distribution as an importance sampler, it is matched to the posterior distribution at the mode by setting $\mu$ equal to the mode of the posterior, and $\Sigma^{-1}$ equal to negative Hessian of the log of the posterior distribution.

When the MVT is chosen as an importance sampler, one is faced with the dilemma of how to choose the degrees of freedom. Again considering the case where the tails of the posterior distribution decay exponentially, if the degrees of freedom is too large, the tails of the importance sampler may not be thick enough. In this situation, one encounters with the same problems that are associated with using the MVN as the importance sampler. However, when the degrees of freedom is too small the tails of the posterior distribution may be over sampled, but given very little weight. This over sampling of the tails can be costly when one evaluation of the posterior distribution
takes a significant amount of computer time.
Chapter 2

Characerizations of the

Multivariate Logistic Distribution

This chapter discusses four different generalizations of the Multivariate Logistic Distribution. Their potential use as importance samplers when the posterior distribution of interest, $p^*(t)$, has tails that decay at a rate similar to $\exp(-|t|)$ is investigated. Three of the four generalizations have the form of the univariate logistic distribution when the one-dimensional case is considered. The univariate logistic distribution is described in Section 2.1. Each generalization $g(t)$ was evaluated based on four criteria:

1. Generating random variables from $g(t)$ should be relatively easy;

2. It should be possible to evaluate $g(t)$ for a given $t$;
3. It should be possible to match $g(t)$ to the posterior distribution at the mode;

4. The tails of $g(t)$ should decay at an approximately exponential rate.

The first three criteria can be considered minimum qualifications for any importance sampler. The final criterion is to determine if the proposed importance sampler would be useful when the posterior distribution has tails that decay at an exponential rate. Only one of the four generalizations (the Elliptical Multivariate Logistic Distribution) proposed in this chapter meets all four requirements. This distribution is investigated first. Chapter 4 discusses the results of simulation experiments using the Elliptical Multivariate Logistic Distribution as an importance sampler. The three remaining generalizations fail to meet the minimum requirements for an importance sampler. These distributions are not investigated further.

### 2.1 Elliptical Multivariate Logistic Distribution (EMVL)

This section discusses a generalization of the multivariate logistic distribution that meets all four of the criteria established at the beginning of this chapter. To define this distribution, let $T$ be a $d$-dimensional random vector and $k_d$ a normalization constant. The Elliptical Multivariate Logistic Distribution (EMVL) is defined as:

$$ g(t) = (2\pi)^{-\left(\frac{d-1}{2}\right)} \cdot k_d \cdot |L|^{-1} \frac{e^{-\sqrt{(t-t')\Sigma^{-1}(t-t')}}}{(1 + e^{-\sqrt{(t-t')\Sigma^{-1}(t-t')}})^2}, \quad (2.1) $$
where \(-\infty < t_i < \infty, i = 1, \ldots, d\) and \(\Sigma = LL'\).

When \(d = 1\) the EMVL reduces to:

\[
g(t) = \frac{\sigma^{-1} \exp(-t/\sigma)}{(1 + \exp(-t/\sigma))^2}, \quad -\infty < t < \infty,
\]

which is the form of the univariate logistic distribution with mean 0 and variance \(\sigma \pi^2/3\). Clearly, the tails of the joint density (2.1) have an exponential rate of decline, so all that remains to be investigated are the other three requisites established at the beginning of this chapter. To further understand the properties of this distribution, the tail behavior of the marginal distributions will be investigated as well.

### 2.1.1 Generating from the EMVL

The first criterion established for a viable importance sampler was that it must be possible to generate random vectors from its distribution. A technique for generating random vectors from the EMVL makes use of the fact that the distribution of \(V = L^{-1}(T - \hat{t})\), where \(\Sigma = LL'\), is spherically symmetric (i.e., the EMVL is elliptically symmetric).

First note that the density of \(V\) can be written as:

\[
g_V(v) = c_d \cdot \frac{e^{-\sqrt{v^T v}}}{(1 + e^{-\sqrt{v^T v}})^2}, \quad -\infty < v_i < \infty,
\]

where \(c_d\) is a normalization constant. Now, consider a transformation of \(V\), where \(V = R \cdot Z\), \(R\) and \(Z\) are independent, and

\[
Z = \frac{1}{\|V\|} \cdot V
\]
\[ R = \| \mathbf{V} \|. \]

This transformation is discussed in some detail in Anderson (1984). It will be shown that it is possible to generate the random variable \( R \) and the random vector \( \mathbf{Z} \) with relative ease, and hence possible to generate the random vector \( \mathbf{T} \).

The density function of the random variable \( R \) must be known in order to sample from its distribution. An expression for the density function can be found by noting that the density of \( \mathbf{V} \) can be expressed as a function of \( \mathbf{V} \mathbf{V} \). This implies that the density of \( R^2 = \mathbf{V} \mathbf{V} \) can be written as (Anderson, 1984):

\[
f_{R^2}(r^2) = \frac{1}{2} C_d \cdot m(r^2) \cdot (r^2)^{\frac{1}{2}d-1}
\]

where \( C_d = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \), the volume of a unit sphere in \( d \) dimensions, and

\[
m(u) = c \cdot \frac{e^{-\sqrt{u^2}}}{(1 + e^{-\sqrt{u^2}})^2}, \quad -\infty < u < \infty.
\]

A simple transformation of variables yields the distribution of \( R \):

\[
f_R(r) = C_d \cdot m(r) \cdot r^{d-1}
\]

Random variables can be generated from this distribution using a ratio of uniform random variables (Kinderman and Monahan, 1977). This is done by generating two uniform random variables \( A \) and \( B \), where the point \( (A,B) \) is uniformly distributed over the region

\[
C_f = \{(a, b) : 0 \leq a \leq f_R^{1/2}(b/a)\}.
\]
The ratio \( R = B/A \) has a density that is proportional to \( f_R \).

In practice this is done by noting that the region \( C_f \) often fits into a box with vertices \((0, b^*_+), (0, b^-), (a^*, b^*_+), (a^*, b^-)\). One method of computing the vertices of the box is to set:

\[
\begin{align*}
a^* &= \max_x \{ f^{1/2}(x) \} \\
b^*_+ &= \max_x \{ x \cdot f^{1/2}(x) \} \\
b^- &= \min_x \{ x \cdot f^{1/2}(x) \}.
\end{align*}
\]

Then one simply uses the following acceptance-rejection algorithm to generate \( R \) (Monahan, 2001, pp 283-284):

1. Generate \( A \sim \text{Uniform} \ (0, a^*) \)
2. Generate \( B \sim \text{Uniform} \ (b^-, b^*_+) \)
3. \( R = B/A \)
4. If \( A^2 \leq f_R(R) \) then deliver \( R \), else go to 1.

The second issue to address is the generation of the random vector \( Z \). Since \( g_V \) is a spherical distribution (i.e., \( g_V \) is left invariant under the set of \( d \times d \) orthogonal matrices), \( Z \) is uniformly distributed over a \( d \)-dimensional unit sphere (Anderson, 1984). To verify that \( g_V \) is a spherical distribution, consider \( Q \), an orthogonal \( d \times d \) matrix. Then,

\[
g_V(Qv) = c \cdot |Q| \frac{e^{-\sqrt{Q^T Q}v}}{(1 + e^{-\sqrt{Q^T Q}v})^2}
\]
\[
\begin{align*}
&= c \cdot \frac{e^{-\sqrt{\mathbf{v}^\prime \mathbf{v}}}}{(1 + e^{-\sqrt{\mathbf{v}^\prime \mathbf{v}}})^2} \\
&= m(\mathbf{v}^\prime \mathbf{v}) \\
&= g_V(\mathbf{v}).
\end{align*}
\]

Since \( \mathbf{Z} \) is uniformly distributed over a unit sphere, one can generate \( \mathbf{Y} \) where \( y_i \sim N(0, 1) \); then \( Z_i = Y_i/\|\mathbf{Y}\| \).

In conclusion, it is comparatively simple to generate random variables from this generalization of the Multivariate Logistic distribution since, to generate \( \mathbf{V} \), one simply needs to generate \( \mathbf{R} \) and \( \mathbf{Z} \). Then, since \( \mathbf{R} \) and \( \mathbf{V} \) are independent, \( \mathbf{V} = \mathbf{R} \cdot \mathbf{Z} \) and \( \mathbf{T} = \hat{t} + L \mathbf{V} \). Therefore, the EMVL satisfies the first requirement established at the beginning of this chapter.

### 2.1.2 Evaluating the Density Function

The second issue to address is the expression of the density function of the EMVL. The form for the density function is known and can be evaluated for a given \( \mathbf{t} \), except, perhaps, for the normalization constant, \( k_d \). The EMVL can be used as an importance sampler for \( h(\mathbf{t}) \neq 1 \), but in order to estimate the normalization constant of the posterior distribution, it is necessary to know \( k_d \). An expression for \( k_d \), however, can be found by noting that it is possible to write \( g(\mathbf{t}) \) as a mixture of normals where \( S \) is a random variable with density \( q_d(\mathbf{s}) \) and

\[
f_{\mathbf{T}|S}(s|\mathbf{t}) = (2\pi)^{-\frac{d}{2}} \left( \frac{1}{s} \right)^d e^{-\frac{1}{2s^2} \mathbf{t}' \mathbf{t}}
\]
\[ q_d(s) = s^{d-1}k_d \cdot dQ(s). \]

Here \( dQ(s) = \frac{d}{ds}L\left(\frac{1}{2}s\right) \), and \( L \) is the Kolmogorov-Smirnov (K-S) distribution

\[ L(s) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j-1}\exp(-2j^2s^2). \]

To show that

\[ g(t) = \int f_T|S(t|s)q_d(s)ds, \]

utilize the following result from Andrews and Mallows (1974) and Stefanski (1990). They demonstrate that the logistic distribution can be expressed as a Gaussian scale mixture:

\[ f(t) = \int_0^\infty \sigma^{-1}\phi(t/\sigma)dQ(\sigma) \]

\[ = \frac{e^{-t}}{(1+e^{-t})^2} \]

where \( \phi \) is the standard normal density. Using this fact,

\[ g(t) = \int \frac{(2\pi)^{-\frac{d}{2}}}{|L|} \left( \frac{1}{s} \right)^d e^{-\frac{1}{2\pi^2}(t-\bar{t})^\prime \Sigma^{-1}(t-\bar{t})} k_d \cdot s^{d-1}dQ(s) \]

\[ = \frac{(2\pi)^{-(d-1)}}{|L|} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\pi^2}(t-\bar{t})^\prime \Sigma^{-1}(t-\bar{t})} k_d \cdot s^{-1}dQ(s) \]

\[ = (2\pi)^{-(d-1)} \cdot k_d \cdot \frac{e^{-\sqrt{(t-\bar{t})^\prime \Sigma^{-1}(t-\bar{t})}}}{|L|} \cdot \frac{1}{(1 + e^{-\sqrt{(t-\bar{t})^\prime \Sigma^{-1}(t-\bar{t})}})^2}. \]

Since \( \int s^{d-1} \cdot k_d dQ(s) = 1 \), we see that \( k_d = 1 \) for \( d=1 \), and \( k_d \) is related to the inverse of the \((d-1)^{st}\) moment of the K-S distribution for \( k > 1 \). Using the form of the \((d-1)^{st}\) moment of the K-S distribution given in Johnson and Kotz (1970), one
finds:

\[
\frac{1}{k_d} = 2^{d-1}E((\frac{1}{2}s)^{d-1})
\]

\[
= 2^{(d-1)} \frac{\Gamma(\frac{1}{2} (d-1) + 1)}{2^{\frac{1}{2}(d-1) - 1}} \sum_{j=1}^{\infty} (-1)^{j-1} j^{-(d-1)}
\]

\[
= 2^{\frac{1}{2}(d+1)} \frac{\Gamma(\frac{1}{2} (d+1))}{2^{\frac{1}{2}(d-1)} - 1} \sum_{j=1}^{\infty} (-1)^{j-1} j^{-(d-1)}.
\]

Since the value of \(k_d\) can be computed using zeta functions (Abramowitz and Stegun, 1974, Sec. 23), the criterion requiring that it be possible to evaluate the importance sampling density has been satisfied.

\[\text{2.1.3 Matching the Posterior and the EMVL}\]

The third requisite for an importance sampler is that it must be possible to match the importance sampler to the posterior distribution at the posterior mode. In other words, it must be possible to set \(\nabla \ln g(t) \mid_{t=\hat{t}} = 0\) and \(-\nabla^2 \ln p^s(t) \mid_{t=\hat{t}} = -\nabla^2 \ln g(t) \mid_{t=\hat{t}}\), where \(\hat{t}\) is the posterior mode. Again, defining \(v = L^{-1}(t - \hat{t})\), and noting that \(\sqrt{v^T v} = \sqrt{\|v\|^2}\) one can write:

\[
\ln g(t) = \ln((2\pi)^{-\left(\frac{d-1}{2}\right)} \cdot k_d \cdot |L|^{-1}) - \sqrt{\|v\|^2} - 2 \ln(1 + e^{-\sqrt{\|v\|^2}})
\]

\[
\nabla \ln g(t) = -\frac{\Sigma^{-1}(t - \hat{t})}{\sqrt{\|v\|^2}} + 2 \frac{e^{-\sqrt{\|v\|^2}}}{1 + e^{-\sqrt{\|v\|^2}}} \frac{\Sigma^{-1}(t - \hat{t})}{\sqrt{\|v\|^2}}
\]

\[
= \frac{-\Sigma^{-1}(t - \hat{t})(1 - e^{-\sqrt{\|v\|^2}})}{\sqrt{\|v\|^2}} \frac{1}{(1 + e^{-\sqrt{\|v\|^2}})}.
\]

Expanding \(e^{-\sqrt{\|v\|^2}}\) for small \(\|v\|\),

\[
1 - e^{-\sqrt{\|v\|^2}} = (1 - \sqrt{\|v\|^2} + ...) - 1
\]
\[
\sqrt{\|v\|^2}
\]

and, since \(\lim_{t \to \hat{t}} (1 + e^{-\sqrt{\|v\|^2}}) = 2\),

\[
\lim_{t \to \hat{t}} \nabla \ln g(t) = \frac{0}{\sqrt{\|v\|^2}} \cdot \frac{\sqrt{\|v\|^2}}{2} = 0.
\]

An expression for the Hessian of the EMVL evaluated at \(\hat{t}\) can be found by writing:

\[
\nabla^2 \ln g(t) = \Sigma^{-1}(t - \hat{t})(t - \hat{t})' \Sigma^{-1} \frac{1 - e^{-\sqrt{\|v\|^2}}}{\|v\|^2 (1 + e^{-\sqrt{\|v\|^2}})} - \frac{\Sigma^{-1}}{\sqrt{\|v\|}^2} \left(1 - e^{-\sqrt{\|v\|^2}}\right) \frac{2e^{-\sqrt{\|v\|^2}}}{\|v\|^2 (1 + e^{-\sqrt{\|v\|^2}})^2} \\
- \Sigma^{-1}(t - \hat{t})(t - \hat{t})' \Sigma^{-1} \frac{1}{\|v\|^2} \left(1 + e^{-\sqrt{\|v\|^2}}\right)
\]

Expanding \(e^{-\sqrt{\|v\|^2}}\) for small \(\|v\|\),

\[
1 - e^{-2\sqrt{\|v\|^2}} - 2\sqrt{\|v\|^2} \cdot e^{-\sqrt{\|v\|^2}} = 1 - (1 - 2(\sqrt{\|v\|^2}) + 4(\frac{\|v\|^2}{2!})
\]

\[-8(\frac{(\|v\|^2)^{3/2}}{3!}) + ... \]

\[-2\sqrt{\|v\|^2}(1 - \sqrt{\|v\|^2} + \frac{\|v\|^2}{2!} - ...) \]

\[
\approx \frac{(\|v\|^2)^{3/2}}{3}
\]

and

\[
1 - e^{-\sqrt{\|v\|^2}} = (1 - \sqrt{\|v\|^2} + ...) - 1
\]

\[
\approx \sqrt{\|v\|^2}
\]
Since \( \lim_{t \to \hat{t}} (1 + e^{-\sqrt{||v||^2}}) = 2 \)

\[
\lim_{t \to \hat{t}} \Sigma^{-1}(t - \hat{t})(t - \hat{t})' \Sigma^{-1} \frac{1 - e^{-2\sqrt{||v||^2} - 2\sqrt{||v||^2} e^{-\sqrt{||v||^2}}}}{(\sqrt{||v||^2})^{3/2} (1 + e^{-\sqrt{||v||^2}})^2} = 0 \cdot \frac{1}{3} \cdot \frac{1}{4} = 0
\]

and

\[
\lim_{t \to \hat{t}} -\Sigma^{-1} \frac{1 - e^{\sqrt{||v||^2}}}{\sqrt{||v||^2}(1 + e^{\sqrt{||v||^2}})} = -\cdot \Sigma^{-1} \cdot \sqrt{||v||^2} \cdot \frac{1}{2\sqrt{||v||^2}} = -\frac{1}{2} \cdot \Sigma^{-1}
\]

Therefore,

\[
\nabla^2 \ln g(t) \big|_{t=\hat{t}} = -\frac{1}{2} \Sigma^{-1},
\]

and, to achieve proper scaling at the mode, set \( \Sigma^{-1} = -2\nabla^2 \ln p^*(t) \big|_{t=\hat{t}} \). Since it is possible to match the EMVL to the posterior distribution at the mode, the third requirement for an importance sampler has been met.

2.1.4 Tail Behavior of the Marginal Distributions of the EMVL

To further understand the properties of the EMVL, the marginal distributions of the EMVL, \( f_{T_i}(t_i), i=1,...,d \), were examined. Obviously, when \( d = 1 \), the distribution of \( T \) is simply the univariate logistic distribution. However, the task of finding the marginal distributions does not yield an easy analytical solution in higher dimensions.
Even when one considers the case where \(d=2\), the integral

\[
f_{T_2}(t_2) = \int_{-\infty}^{\infty} e^{-\sqrt{t_1^2+t_2^2}} \frac{1}{(1+e^{-\sqrt{t_1^2+t_2^2}})^2} dt_1
\]

cannot be solved using basic calculus techniques. To address this problem two numerical methods were employed. For simplicity, consider the situation where \(\hat{t} = 0\) and \(\Sigma = I\).

The first method takes advantage of the characterization of the EMVL as a scale mixture of the MVN where the scale is related to the Kolmogorov-Smirnov distribution (Section 2.1.2). Since \(T|S\) is distributed as \(N(0, I_{\frac{1}{s^2}})\) the marginal distribution of \(T_i|S \sim N(0, \frac{1}{s^2})\). Using this fact and the distribution of \(S\) as \(Q\), write

\[
f_{T_i}(t_i) = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} s^{d-2} \exp(-\frac{t_i^2}{s^2}) dQ(s)
\]

Again, save for \(d=1\), this integral cannot be evaluated easily using analytical methods, but since the integration problem has now been reduced to one dimension, a numerical method can be used. Unfortunately, any numerical method will only provide an approximation, no matter how good, of the true integral. However, these approximations can be examined for evidence of the behavior of the tails of the marginal distributions. One numerical method to consider is Simpson’s rule for integration. If \(f\) is to be integrated over the interval \([a,b]\), Simpson’s rule divides the interval into \(n\) subintervals of the size \(h = (b-a)/n\) and estimates the integral as:

\[
h[f(a) + 4 \sum_{i=1}^{n} f(a + h \frac{2i-1}{2}) + 2 \sum_{i=1}^{n-1} f(a + ih) + f(b)] / 6.
\]
It must be possible to evaluate the integrand at the endpoints of the subintervals in order to use Simpson’s rule. In this case, the value of \(dQ(s)\) is not easily evaluated for a given \(s\), but since \(Q(s)\) can be evaluated with reasonable ease (Monahan, 1989), it can be used to estimate \(dQ(s)\). This was done by fitting a natural cubic spline to \(Q(s)\) and using its derivative to estimate \(dQ(s)\). The domain of \(s\) was split into 30 subintervals over which a cubic function was fitted, forming the spline interpolant \(\widehat{Q}(s)\) and its corresponding derivative \(\widehat{q}(s)\). The estimate, \(\hat{f}_{T_1}(t_i)\) of the integral

\[
\hat{f}_{t_i}(t_i) = \int_0^\infty \frac{1}{\sqrt{2\pi}} s^{d-2} \exp\left(-\frac{t_i^2}{s^2}\right)q(s)ds
\]

was calculated using Simpson’s rule.

Just how good is the estimate of the marginal densities when the derivative of the spline is used to estimate \(dQ(s)\)? Since the value of the integral \(f_{T_1}(t_i)\) is known when \(d = 1\), it is possible to calculate the difference between the estimate and the actual density. Simpson’s rule was used to calculate:

\[
\hat{I}_1 = \int_0^\infty |\hat{f}_{T_1}(t_1) - f_{T_1}(t_1)|dt_1
\]

\[
\hat{I}_2 = \sqrt{\int_0^\infty (\hat{f}_{T_1}(t_1) - f_{T_1}(t_1))^2dt_1}
\]

for \(k=5, 10, 15, 20, 25\) and 30 subintervals (Table 2.1). The derivative of the spline estimate of \(Q(s)\) provides a good approximation of \(dQ(s)\) when \(d = 1\) and a relatively large number of intervals is used.

A second method of approximating the marginal densities was used to examine the tail rate of decay of the marginal densities. The previous method of using splines did
<table>
<thead>
<tr>
<th></th>
<th>$I_1$</th>
<th>$I_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$1.9577 \times 10^{-1}$</td>
<td>$1.8584 \times 10^{-1}$</td>
</tr>
<tr>
<td>10</td>
<td>$3.0959 \times 10^{-2}$</td>
<td>$4.2805 \times 10^{-2}$</td>
</tr>
<tr>
<td>15</td>
<td>$7.7655 \times 10^{-3}$</td>
<td>$8.6364 \times 10^{-3}$</td>
</tr>
<tr>
<td>20</td>
<td>$2.8487 \times 10^{-3}$</td>
<td>$3.8848 \times 10^{-3}$</td>
</tr>
<tr>
<td>25</td>
<td>$6.7694 \times 10^{-4}$</td>
<td>$9.5239 \times 10^{-4}$</td>
</tr>
<tr>
<td>30</td>
<td>$1.6685 \times 10^{-4}$</td>
<td>$2.5576 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 2.1: Evaluation of Marginal Density Estimates for $d = 1$

not assume anything about the tail behavior of the marginal distributions. However, since the joint density decays at an exponential rate, if the assumption is made that the tails of the marginal density also decay exponentially, Laguerre polynomials can be used to approximate the marginal densities by expanding the marginal density, $f_{T_i}$, in a Fourier series (Cencov, 1962 and Monahan, 2001): $f_{T_i}(t_i) = \sum_{j=0}^{\infty} c_j p_j(t_i)$, where $p_j(t) = \phi_j(t) \cdot e^{-t}$. The $\phi_j(t)$, $j = 0, 1, 2, \ldots$ are the series of complete orthonormal functions as given in Monahan (2001):

$$
\begin{align*}
\phi_0(t) &= 1 \\
\phi_1(t) &= 1 - t \\
\phi_2(t) &= (2 - 4t - t^2)/2 \\
&\vdots
\end{align*}
$$

with recurrence formula: $(n + 1)\phi_{n+1}(t) = (2n + 1 - t)\phi_n(t) - n\phi_{n-1}(t)$.

The Fourier coefficients were estimated using $\hat{c}_j = \frac{1}{n} \sum_{i=1}^{n} p_j(T_{i(t)}) w(T_{i(t)})$, where
\(w(t) = e^t\) so that the \(p_j\) were orthogonal with respect to \(w(t)\). The marginal random variables, \(T_{i(l)}, l = 1, \ldots, n\) were generated from a EMVL distribution. If the assumption about the tail rate of decay was reasonable, the Fourier coefficients should become very small as \(j\) increases.

Forty-thousand random variables were generated from the EMVL distribution for each of \(d=1, 3, 5,\) and 10. A technical problem arose as the domain of \(T_i\) is \((-\infty, \infty)\) and the domain of the Laguerre polynomials is \((0, \infty)\). This problem was solved by noting that EMVL is symmetric about 0. The marginal densities then were estimated using the absolute value of the generated \(T_i\) values to fit the Fourier coefficients. Then \(2 \ast f_{T_i}(t_i)\) was estimated by the Fourier series using the first \(k = 10\) estimated coefficients: 
\[
\widehat{f}_{T_i}(t_i) = \sum_{j=0}^{k} c_j p_j(t_i).
\]
Table 2.2 shows the estimated coefficients up to \(k = 15\) for \(d = 1, 3, 5\) and 10.

The fact that the Fourier coefficients become very small as \(j\) increases provides further evidence in favor of the assumption that the marginal distributions have approximately exponential tail behavior. As a further check, one can compare these marginal density estimates to those that were derived by fitting a spline. The marginal density estimates produced using Simpson’s rule for \(d=1, 3, 5,\) and 10 were calculated using 30 intervals to estimate the density at each \(t_i\). If our choice of Laguerre polynomials “forced” the marginal densities to exhibit exponential tail behavior, one would not expect their density estimates to closely match the estimates that result from fitting a spline to \(Q(s)\). It is apparent from Figures 2.1-2.1.4 that the two estimates match
<table>
<thead>
<tr>
<th>$j$</th>
<th>$d=1$</th>
<th>$d=3$</th>
<th>$d=5$</th>
<th>$d=10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.38554025</td>
<td>0.33821484</td>
<td>0.30710647</td>
<td>0.23282228</td>
</tr>
<tr>
<td>1</td>
<td>0.12935252</td>
<td>0.09083394</td>
<td>0.0742228</td>
<td>0.03436207</td>
</tr>
<tr>
<td>2</td>
<td>0.03011608</td>
<td>0.00848164</td>
<td>0.00648283</td>
<td>-0.00468211</td>
</tr>
<tr>
<td>3</td>
<td>-0.00361001</td>
<td>-0.01222448</td>
<td>-0.00755043</td>
<td>-0.00663721</td>
</tr>
<tr>
<td>4</td>
<td>-0.01189556</td>
<td>-0.01253462</td>
<td>-0.00713078</td>
<td>-0.00341662</td>
</tr>
<tr>
<td>5</td>
<td>-0.01142526</td>
<td>-0.0078907</td>
<td>-0.00450211</td>
<td>-0.00122777</td>
</tr>
<tr>
<td>6</td>
<td>-0.00863972</td>
<td>-0.00341648</td>
<td>-0.00259758</td>
<td>-0.00025016</td>
</tr>
<tr>
<td>7</td>
<td>-0.00574272</td>
<td>-0.00036605</td>
<td>-0.00159952</td>
<td>0.00013308</td>
</tr>
<tr>
<td>8</td>
<td>-0.00333541</td>
<td>0.00129475</td>
<td>-0.00114828</td>
<td>0.00030734</td>
</tr>
<tr>
<td>9</td>
<td>-0.00149763</td>
<td>0.00192163</td>
<td>-0.0009505</td>
<td>0.00041537</td>
</tr>
<tr>
<td>10</td>
<td>-0.00018355</td>
<td>0.00187115</td>
<td>-0.00086045</td>
<td>0.00048509</td>
</tr>
<tr>
<td>11</td>
<td>0.00066309</td>
<td>0.00142055</td>
<td>-0.00082753</td>
<td>0.00051319</td>
</tr>
<tr>
<td>12</td>
<td>0.00109363</td>
<td>0.00076615</td>
<td>-0.00083828</td>
<td>0.00049852</td>
</tr>
<tr>
<td>13</td>
<td>0.00116378</td>
<td>0.00004048</td>
<td>-0.00088221</td>
<td>0.00044822</td>
</tr>
<tr>
<td>14</td>
<td>0.00094103</td>
<td>-0.00067022</td>
<td>-0.00094021</td>
<td>0.00037392</td>
</tr>
<tr>
<td>15</td>
<td>0.00050452</td>
<td>-0.00131243</td>
<td>-0.00098585</td>
<td>0.00028690</td>
</tr>
</tbody>
</table>

Table 2.2: Estimated Fourier Coefficients for dimensions 1, 3, 5 and 10 resulting from fitting Laguerre polynomials to the marginal densities of the EMVL.
very closely. Based on these empirical results, it is not unreasonable to conclude that
the tails of the marginal distributions decay at a rate similar to $\exp(-|t|)$.

2.2 Simulation: Using the EMVL as an Importance Sampler for Dirichlet Posterior Distributions

The EMVL passed all four requirements that were established at the beginning
of the Chapter. As a further check, a simulation experiment was used to verify that
using the EMVL as an importance sampler:

1. produces unbiased estimates, and;
Figure 2.2: Comparison of the marginal density estimates of the EMVL (d=5) produced by fitting Laguerre polynomials (dashed line) and using Simpson’s rule of integration (solid line).

Figure 2.3: Comparison of the marginal density estimates of the EMVL (d=10) produced by fitting Laguerre polynomials (dashed line) and using Simpson’s rule of integration (solid line).
Table 2.3: Dirichlet Distribution Parameters

2. produces standard errors of the estimates that are reasonable approximations of the true variability.

Twelve different normalized posterior distributions were considered. These posteriors were used by Monahan and Genz (1997) and consist of:

- six Dirichlet distributions with dimensions 3,3,3,7,7,7
- six transformed Dirichlet distributions with the same dimensions, transformed by \( e^{t} \frac{e^{t}}{1+e^{t}} \).

Table 2.3 gives the \( \alpha \) parameters of the six densities from the Dirichlet family. The support of these densities is limited to the simplex \( \sum t_{j} \leq 1 \), and the densities take the form

\[
\prod_{j=1}^{d} t_{j}^{\alpha_{j}-1} (1 - \sum t_{j})^{\alpha_{d+1}-1} \Gamma(\sum_{j=1}^{d+1} \alpha_{j}) / \prod_{j=1}^{d+1} \Gamma(\alpha_{j})
\]

<table>
<thead>
<tr>
<th>Problem</th>
<th>dimension</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>21 22 23 24</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3 4 5 6</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5 10 15 20</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>21 22 23 24 25 26 27 28</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>3 4 5 10 16 21 22 23</td>
</tr>
</tbody>
</table>
The transformed Dirichlet densities use the change in variables
\[ t_j = \frac{e^{x_j}}{1 + \sum_{k=1}^{d} e^{x_k}}, \]
for \( j = 1, \ldots, d \), with the resulting density over all \( R^d \)
\[ \frac{e^{\sum_{j=1}^{d} \alpha_j x_j}}{(1 + \sum_{j=1}^{d} e^{x_j})^{\sum_{j=1}^{d+1} \alpha_j}} \frac{\Gamma(\sum_{j=1}^{d+1} \alpha_j)}{\prod_{j=1}^{d+1} \Gamma(\alpha_j)}. \]

A simulation experiment using 40,000 evaluations was done 100 times for each posterior distribution to estimate the integral \( \int h(t)p^*(t)dt \), where \( h(t) = 1 \). Since the posterior distributions are normalized, the estimate of this integral should be one.

### 2.2.1 Evaluating the Weights

Before evaluating the estimates produced by the simulation study, it is important to determine that the requirements established in Chapter 1 appealing to the CLT have not been violated. One method of doing this is to examine the behavior of the weights, \( w(t) \), with respect to \( \|v\| \), where \( V = L^{-1}(T - \hat{t}) \) and
\[
 w(t) = \frac{p^*(t)}{g(t)} = \frac{p^*(t)}{(2\pi)^{-\left(\frac{d+1}{2}\right)} k_d \frac{1}{|L^{-1}|} \frac{e^{-\sqrt{\sigma \nu}}}{(1+e^{-\sqrt{\sigma \nu}})^2}}.
\]
If the weights decrease as \( \|v\| \) increases, it is reasonable to assume that the tails of the EMVL are thicker than those of the posterior distribution, and to proceed as though the importance sampling estimates are asymptotically normal.

Four thousand weights were calculated for each problem. Figures 2.2.1 - 2.9 plot the weights normalized by \( w(\hat{t}) \) versus the norm of its associated vector, \( v = (t - \hat{t})L^{-1} \). Since the weights decrease as \( \|v\| \) increases, these plots do not indicate
any reason to question the assumption of asymptotic normality of the estimates of $E_p[h(t)]$.

For reference, Figures 2.2.1 and 2.11 plot the normalized weights that result when the multivariate normal distribution is used as the importance sampling distribution for four of the problems. These plots again demonstrate the problem that arises when the posterior density has heavier tails than the importance sampling distribution. When the MVN is used as an importance sampler, the normalized weights tend to fan out and become quite large as $\|v\|$ increases. In these situations, it would be unwise to assume the estimates are asymptotically normal.
Figure 2.4: Importance sampling weights using the posterior distributions from Problems 1 (top figure) and 2 (bottom figure), untransformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $\|v\|$. 
Figure 2.5: Importance sampling weights using the posterior distributions from Problems 3 (top figure) and 4 (bottom figure), untransformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $\|v\|$. 
Figure 2.6: Importance sampling weights using the posterior distributions from Problems 5 (top figure) and 6 (bottom figure), Untransformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $\|v\|$. 
Figure 2.7: Importance sampling weights using the posterior distributions from Problems 1 (top figure) and 2 (bottom figure), transformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $\|v\|$. 
Figure 2.8: Importance sampling weights using the posterior distributions from Problems 3 (top figure) and 4 (bottom figure), transformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $\|v\|$. 
Figure 2.9: Importance sampling weights using the posterior distributions from Problems 5 (top figure) and 6 (bottom figure), transformed Dirichlet and the EMVL as the importance sampler. The horizontal axes are $\|v\|$. 
Figure 2.10: Importance sampling weights using the posterior distributions from Problems 2 (top figure) and 6 (bottom figure), untransformed Dirichlet and the MVN as the importance sampler. The horizontal axes are $\|\mathbf{v}\|$. 
Figure 2.11: Importance sampling weights using the posterior distributions from Problems 1 (top figure) and 5 (bottom figure), transformed Dirichlet and the MVN as the importance sampler. The horizontal axes are $\|\mathbf{v}\|$. 
2.2.2 Unbiasedness

Two different methods were used for evaluating the unbiasedness of the importance sampling estimates of $\int h(t)p^*(t)dt$. For each posterior distribution, a simulation experiment was used to calculate 100 estimates, $Z_i$, of $E_{p^*}[h(t)]$. Each $Z_i$ was computed using 40,000 evaluations. Since $h(t) = 1$ and the posterior distributions are normalized, $E_{p^*}[h(t)] = 1$. If the importance sampling estimates are unbiased one should find $E(Z_i) = 1$.

The first method for evaluating these estimates performs the traditional hypothesis test:

$$H_0 : E(Z_i) = 1$$
$$H_1 : E(Z_i) \neq 1.$$ 

The test statistic $t$ was computed for each simulation, where:

$$t = \frac{\sqrt{n}(\overline{Z} - 1)}{\sqrt{\sum_i (Z_i - \overline{Z})^2/(n - 1)}}.$$ 

The resulting $t$-statistics are given in Table 2.4. The critical 95% $t$-value for these simulations is 1.98. Since the absolute values of all of the $t$-statistics fall below this value, one concludes that there is no evidence to suggest that the simulation experiments produced biased estimates.

An alternative approach uses an equivalence test (Berger and Hsu, 1996) to examine unbiasedness:

$$H_0 : |E(Z_i) - 1| \geq \delta$$
Table 2.4: Entries are the $t$-statistics for testing the hypotheses $H_0 : E(Z_i) = 1$ and $H_1 : E(Z_i) \neq 1$

$$H_1 : |E(Z_i) - 1| < \delta,$$

where the value of $\delta$ is determined beforehand by the experimenter. This hypothesis test is done by first constructing a $(1 - 2 \times \alpha)$ confidence interval. If this interval is completely contained in the predetermined equivalence interval, one concludes that $E(Z_i)$ is equivalent to one with $(1 - \alpha)$% confidence. An equivalence interval of $(0.995, 1.005)$ was used to test for unbiasedness. The choice of $\delta = 0.005$ was arbitrary, and a wider or narrower interval could be used instead. Figure 2.12 shows the 90%, $(\alpha=0.05)$ confidence intervals resulting from the 100 simulations done for each problem. Horizontal reference lines are drawn at 0.995 and 1.005. Since all of the confidence intervals fall between the reference lines, one concludes that the estimates are equivalent to one for all the simulations, and hence, that the importance sampling estimates are unbiased.
Figure 2.12: 90% confidence intervals and horizontal reference lines for the equivalence interval (0.995, 1.005) is given. If the 90% confidence interval is completely contained within the equivalence interval, one concludes that the estimates are equivalent to 1 with 95% confidence.

2.2.3 Error Estimates

Two approaches were used to determine if the 100 standard errors that resulted from each simulation were reasonable approximations of the true variance. The first approach compared the assessment of variability measured in the root mean estimated variance, $RMEV = \sqrt{\frac{\sum_{i=1}^{n} s_i^2}{n}}$ with a measurement of the true variability, $\sqrt{\frac{\sum_{i=1}^{n} (Z_i - \bar{Z})^2}{n}}$.

The estimation of the standard errors of these ratios is based on a Delta-method approximation to the asymptotic ($n \to \infty$) distribution of the ratio. Under the assumption that the ratio is consistently estimating 1, and the denominator is approximately Chi-squared with $n - 1$ degrees of freedom, the resulting large-sample
<table>
<thead>
<tr>
<th>Problem</th>
<th>Untransformed Dirichlet</th>
<th>Transformed Dirichlet</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.023</td>
<td>1.192</td>
</tr>
<tr>
<td>2</td>
<td>0.997</td>
<td>0.856</td>
</tr>
<tr>
<td>3</td>
<td>0.856</td>
<td>0.919</td>
</tr>
<tr>
<td>4</td>
<td>0.995</td>
<td>1.035</td>
</tr>
<tr>
<td>5</td>
<td>0.873</td>
<td>0.928</td>
</tr>
<tr>
<td>6</td>
<td>0.958</td>
<td>0.935</td>
</tr>
</tbody>
</table>

Table 2.5: Ratios of \( RMEV/\sqrt{\sum_{i=1}^{n}(Z_i - \bar{Z})^2/n} \).

The standard error is \( \sqrt{2/n} \) (Stefanski, 2001).

If the importance sampling standard errors are reasonably consistent across simulations, the ratios, \( RMEV/\sqrt{\sum_{i=1}^{n}(Z_i - \bar{Z})^2/n} \), should cluster about one. Ratios that deviate significantly from one are evidence that the standard errors of the importance sampling estimates tend to underestimate (ratios less than one) or overestimate (ratios greater than one) the true variability. The ratios produced for these test problems are shown in Table 2.5. Since the standard error for these ratios is \( \sqrt{2/100} = 0.14 \), ratios that fall within the interval (.72, 1.28) indicate that the standard errors produced by each simulation are reasonable estimates of the true variability. All of the ratios fall within this interval, indicating that the standard errors of the importance sampling estimates are reasonable.

The second approach used to assess the error estimates consisted of inspecting the coverage of a 95% confidence interval of \( h(t) \). One-hundred intervals were constructed for each posterior distribution:

\[ Z_i \pm 1.96 \cdot s_i, \]
Table 2.6: Lack of Coverage of 95% Confidence Intervals. Entries represent the proportion of times that a 95% confidence interval constructed using importance sampling estimates that failed to contain the true value of 1.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Untransformed Dirichlet</th>
<th>Transformed Dirichlet</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6/100</td>
<td>8/100</td>
</tr>
<tr>
<td>2</td>
<td>2/100</td>
<td>4/100</td>
</tr>
<tr>
<td>3</td>
<td>2/100</td>
<td>4/100</td>
</tr>
<tr>
<td>4</td>
<td>6/100</td>
<td>9/100</td>
</tr>
<tr>
<td>5</td>
<td>2/100</td>
<td>2/100</td>
</tr>
<tr>
<td>6</td>
<td>7/100</td>
<td>4/100</td>
</tr>
<tr>
<td>Total</td>
<td>25/600</td>
<td>31/600</td>
</tr>
</tbody>
</table>

where $Z_i$ is the importance sampling estimate of $h(t)$ and $s_i$ is the standard error of $Z_i$. Since an examination of the weights in Section 2.2.1 indicated that it is reasonable to assume that the importance sampling estimates have a normal limiting distribution, if $s_i$ is a good estimate of the true variability, these confidence intervals should demonstrate 95% coverage of $E(Z_i) = 1$. Table 2.6 gives the number of times the confidence intervals do not capture 1, and provides further evidence that $s_i$ is a good estimate of the true variability.

The resulting importance sampling estimates can be evaluated by investigating the bias ($E_p[h(t)] - 1$) and the standard error of the bias. These results are shown in Table 2.7. The small biases and standard errors that were observed further demonstrate that the EMVL was a successful importance sampler for these problems.
2.3 Other Generalizations Considered

The distributions considered in this section failed to meet at least one of the first three requisites for an importance sampler. The reasons for their failure and some of their other properties are explored here.

2.3.1 Classical Multivariate Logistic Distribution

The most obvious multivariate logistic distribution to consider as an importance sampler is the full classical d-variate logistic distribution with real location parameters $\mu_i$ and positive scale parameters $\sigma_i$ as described by Arnold (1992). If $X$ has the classical multivariate logistic distribution, its cumulative distribution function is of
the form:

\[ F_X(x) = \{1 + \sum_{i=1}^{d} \exp\left[-\left(\frac{x_i - \mu_i}{\sigma_i}\right)\right]\}^{-1} \]

Random variables can be generated from this distribution (Arnold, 1992) and its density can be calculated, so this distribution meets the first two criteria that were established. One way to generate random variables from the Classical Multivariate Logistic distribution is to generate \( d + 1 \) independent and identically distributed random variables, \( Y_0, Y_1, Y_2, \ldots, Y_d \) from the extreme value distribution with density

\[ f_Y(y) = e^{-y} \exp(e^{-y}), \quad -\infty < y < \infty. \]

Then, if \( T_i = Y_i - Y_0, i = 1, \ldots, d \) is the \( i \)th element of the random vector \( T, T \) has the Classical Multivariate Logistic Distribution (Arnold, 1992).

The marginal densities, \( f_{T_i}(t_i) \), have the form:

\[ f_{X_i}(x_i; \mu_i, \sigma_i) = \frac{e^{(x_i-\mu_i)/\sigma_i}}{(1 + e^{(x_i-\mu_i)/\sigma_i})^2}. \]

Since these marginal densities are simply univariate logistic densities (whose tails decay at an exponential rate), the fourth criterion is also met.

Unfortunately, this distribution fails to meet the third criterion. If the Classical Multivariate Logistic Distribution is to be of use as an importance sampling density, it must be possible to standardize \( X \) so that its distribution matches the posterior distribution at the posterior mode, \( \hat{t} \), by setting: \( -\nabla^2 \ln p(t) \big|_{t=\hat{t}} = -\nabla^2 \ln f_X(x) \big|_{x=\hat{t}} \) and \( \nabla \ln f_X(x) \big|_{x=\hat{t}} = 0. \) The correlation between \( X_i \) and \( X_j \) is fixed once \( \sigma_i \) and \( \sigma_j \)
have been specified. This restricts the number of posterior distributions for which the Classical Multivariate Logistic distribution can be used, and so the Classical Multivariate Logistic Distribution fails to meet the third criterion.

### 2.3.2 Multivariate Logistic Distribution 1 \( MVL_1 \)

Another generalization of the Multivariate Logistic Distribution is given below for the two dimensional case. Let \( Z_i, i = 1, 2 \) be independent standard normal random variables. Let \( S \) have density \( dQ(s) = \frac{d}{ds} L(\frac{1}{2}s) \), where \( L \) is the Kolmogorov-Smirnov distribution

\[
L(s) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j-1} \exp(-2j^2s^2).
\]

Assume \( S \) and \( Z_i, i = 1, 2 \) are independent, and construct

\[
T_1 = (aZ_1 + bZ_2)S \\
T_2 = (cZ_1 + dZ_2)S
\]

where \( a^2 + b^2 = 1 \) and \( c^2 + d^2 = 1 \). The correlation of \( T_1 \) and \( T_2 \) is

\[
\rho = \text{corr}(T_1, T_2) \\
= \text{corr}(aZ_1 + bZ_2, cZ_1 + dZ_2) \\
= ac + bd
\]

Since the only restrictions on \( a, b, c, \) and \( d \) are \( a^2 + b^2 = 1 \) and \( c^2 + d^2 = 1 \) it is clear that \(-1 \leq \text{corr}(T_1, T_2) \leq 1\).
Random variables can be generated from this distribution with relative ease since one need only generate independent random variables from the normal distribution and the Kolmogorov-Smirnov distribution. Since generating normal random variables is relatively easy and it has been shown (Monahan, 1989) that one can generate random variables from the Kolomogorov-Smirnov distribution with minimal difficulty, the first of the four criteria has been met.

To address the fourth criterion, one can examine the behavior of the marginal distributions. Since the marginal distribution of \( T^{(i)} \) is a scale mixture of a normal random variable and a Kolmogorov-Smirnov random variable, it has a univariate logistic distribution (Andrews and Mallows, 1974; Stefanski, 1990). This result is discussed in further detail in Section 2.1.2. Since the logistic distribution has exponential tail behavior, this distribution meets the fourth criterion that we established at the beginning of this chapter.

The difficulty arises in meeting the second criterion: it must be easy to evaluate the density of the importance sampler. Note that the cumulative distribution function of \( T_1, T_2 \) can be written as:

\[
Pr(T_1 \leq t_1, T_2 \leq t_2) = \int \Phi(t_1, t_2; 0, 0, s^2, s^2, \rho s^2) dQ(s)
\]

where \( \rho = ac + bd \), and \( \Phi \) is the bivariate normal cumulative distribution function

\[
\Phi(t_1, t_2; 0, 0, s^2, s^2, \rho s^2) = \\
\int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \frac{1}{2\pi s^4 \sqrt{1 - \rho s^2}} \exp\left\{ -\frac{1}{2(1 - \rho^2 s^4)} \left( \frac{t_1^2}{s^2} - 2\rho t_1 t_2 + \frac{t_2^2}{s^2} \right) \right\} dt_2 \, dt_1.
\]
One can see from this expression that the probability density function is not easily computed – even for only two dimensions.

This Multivariate Logistic distribution did not meet all of the established criteria. Therefore it was concluded that while this generalization of the Multivariate Logistic Distribution may be interesting to explore in and of itself, it is not useful as an importance sampling distribution.

### 2.3.3 Alternate Generalization

An alternative generalization of the Multivariate Logistic Distribution, is defined as:

$$f_T(t) = c_d \cdot \frac{e^{-(t-\bar{t})'\Sigma^{-1}(t-\bar{t})}}{(1 + e^{-(t-\bar{t})'\Sigma^{-1}(t-\bar{t})})^2}, -\infty < t_i < \infty,$$

where $c_d$ is the normalization constant and $d$ is the dimension of the random vector $t$.

One should note that this distribution does not meet the fourth criterion established at the beginning of the chapter since its tails decay at a rate of $\exp(-ct^2)$. Additionally, one should note, that unlike the other generalizations discussed in this chapter, when $d = 1$ this generalization does not have the univariate logistic distribution. Because of this, this distribution, is not, in a strict sense, a generalization of the logistic distribution. However, it is included in this section since its distribution has a similar functional form to the EMVL. Unfortunately, this distribution failed to meet the third criterion; it is not possible to match its Hessian evaluated at the mode to that of the posterior distribution’s.
Using the procedure described in Section 2.1.1, one can generate random variables from this distribution since it is spherically symmetric. In order to evaluate the probability density function (and fulfill the second requirement we established for an importance sampler) we need to know the normalization constant, $c_d$. However, even if the normalization constant is not known, one can still estimate $E_{p_r}[h(t)]$ for most $h(t)$; the notable exception is $h(t) = 1$.

First, consider the case where $d = 1$. To solve for the normalization constant, $c_1$, define:

$$A_k = \int_{-\infty}^{\infty} -\exp(-kx^2) \frac{1}{1 + \exp(-x^2)} \, dx.$$  

Then

$$\frac{1}{c_1} = \int_{-\infty}^{\infty} \exp(-x^2) \frac{1}{(1 + \exp(-x^2))^2} \, dx = \sum_{k=1}^{\infty} (-1)^{k+1} A_k,$$

where

$$A_1 = \sqrt{\pi} \sum_{j=1}^{\infty} (-1)^{j+1} j^{-1/2}$$

and

$$A_{k+1} = \frac{\sqrt{\pi}}{k} - A_k.$$

Proof:

$$\frac{1}{c_1} = \int_{-\infty}^{\infty} \frac{\exp(-x^2)}{(1 + \exp(-x^2))^2} \, dx = \sum_{k=1}^{\infty} (-1)^{k} \int_{-\infty}^{\infty} \frac{\exp(-kx^2)}{1 + \exp(-x^2)} \, dx = -\sum_{k=1}^{\infty} (-1)^{k} \int_{-\infty}^{\infty} \frac{\exp(-kx^2)}{1 + \exp(-x^2)} \, dx$$
\[ A_{k+1} = \int_{-\infty}^{\infty} \exp(-x^2) \frac{\exp(-kx^2)}{1 + \exp(-x^2)} \, dx \]
\[ = \int_{-\infty}^{\infty} (1 + \exp(-x^2) - 1) \frac{\exp(-kx^2)}{1 + \exp(-x^2)} \, dx \]
\[ = \int_{-\infty}^{\infty} \exp(-kx^2) - \int_{-\infty}^{\infty} \frac{\exp(-kx^2)}{1 + \exp(-x^2)} \, dx \]
\[ = \sqrt{\frac{\pi}{k}} - A_k \]

\[ A_1 = \int_{-\infty}^{\infty} \frac{\exp(-x^2)}{1 + \exp(-x^2)} \, dx \]
\[ = \int_{-\infty}^{\infty} \sum_{j=1}^{\infty} (-1)^{j+1} \exp(-jx^2) \, dx \]
\[ = \sum_{j=1}^{\infty} (-1)^{j+1} \int_{-\infty}^{\infty} \exp(-jx^2) \, dx \]
\[ = \sum_{j=1}^{\infty} (-1)^{j+1} \sqrt{\frac{\pi}{j}} \]

Clearly, even for one dimension, analytically calculating the normalization constant is not an easy task. However, since one can otherwise evaluate the probability density function of this distribution, this difficulty alone would not rule out its use as an importance sampler.

Unfortunately, this generalization fails to meet the third criterion for an importance sampler: that it must be possible to set \( \nabla \ln g_T(t) \big|_{t=\hat{t}} = 0 \) and \( -\nabla^2 \ln p^*(t) \big|_{t=\hat{t}} = -\nabla^2 \ln g_T(t) \big|_{t=\hat{t}} \). Since \( \nabla^2 \ln f_T(t) \big|_{t=\hat{t}} = 0 \), this distribution cannot be used as an importance sampler for any posterior distribution. To demonstrate this, write:

\[ \ln f_T(t) = \ln(c_d) - (t - \hat{t})'\Sigma^{-1}(t - \hat{t}) - 2\ln(1 + \exp\{-(t - \hat{t})'\Sigma^{-1}(t - \hat{t})\}) \]
and

\[ \nabla \ln f_T(t) = -2 \Sigma^{-1}(t - \hat{t}) + 4 \Sigma^{-1}(t - \hat{t}) \cdot \frac{e^{-(t - \hat{t})'\Sigma^{-1}(t - \hat{t})}}{1 + e^{-(t - \hat{t})'\Sigma^{-1}(t - \hat{t})}}. \]

Evaluating \( \nabla \ln f_T(t) \) at the mode of the posterior distribution, \( \hat{t} \), one finds \( \nabla \ln f_T(t) \big|_{t=\hat{t}} = 0 \), as expected. However, investigating \( \nabla^2 \ln f_T(t) \), one finds:

\[
\nabla^2 \ln f_T(t) = -2 \Sigma^{-1} + 4 \Sigma^{-1} \cdot \frac{\exp\{- (t - \hat{t})'\Sigma^{-1}(t - \hat{t})\}}{1 + \exp\{- (t - \hat{t})'\Sigma^{-1}(t - \hat{t})\}} \\
-8 \cdot \Sigma^{-1}(t - \hat{t})(t - \hat{t})'\Sigma^{-1} \cdot \frac{\exp\{- (t - \hat{t})'\Sigma^{-1}(t - \hat{t})\}}{1 + \exp\{- (t - \hat{t})'\Sigma^{-1}(t - \hat{t})\}} \\
+8 \cdot \Sigma^{-1}(t - \hat{t})(t - \hat{t})'\Sigma^{-1} \cdot \frac{\exp\{-2 (t - \hat{t})'\Sigma^{-1}(t - \hat{t})\}}{(1 + \exp\{- (t - \hat{t})'\Sigma^{-1}(t - \hat{t})\})^2}.
\]

Evaluating at the mode,

\[
\nabla^2 \ln f_T(t) \big|_{t=\hat{t}} = -2 \cdot \Sigma^{-1} + 4 \cdot \Sigma^{-1} \cdot \frac{1}{2} - 0 + 0 \\
= 0
\]

Because this distribution failed to meet the third criterion, it cannot be used as an importance sampler.
Chapter 3

Simulation: Using the EMVL as an Importance Sampler for Posterior Distributions From the Literature

Ten examples were chosen from the literature to evaluate the performance of the EMVL as an importance sampler. These examples, used by Monahan and Genz (1997), range in dimension from 3 to 11. The unnormalized posterior distribution is known for each example. Simulation experiments were used to estimate the normalization constant of the posterior distribution ($h(t) = 1$) using the EMVL, the MVN, the MVT with 5 degrees of freedom and the MVT with 3 degrees of freedom as importance samplers. We chose to compare the performance of the MVN and the MVT to the EMVL since they are the most common importance samplers used
in practice. The ten examples are described briefly in Section 3.1, and references where more detail can be found are given. Additionally, the Fortran 95 code that was used to evaluate the likelihood functions for each of the examples is provided in the appendix. Section 3.2 discusses the results of the simulation experiments.

3.1 Examples

3.1.1 Example 1 (d=3)

This example comes from a model that was proposed by Turnbull et al. (1974) and has been analyzed as a Bayesian problem in the past by Naylor and Smith (1982), Tierney and Kadane (1986), Tierney, Kass and Kadane (1989), and others. While it is of relatively small dimension (d=3), the posterior distribution in this problem suffers from heavy tail behavior. This example uses data from the Stanford heart transplant study. Individual patients are assumed to have exponential lifetimes after entering the study. A uniform prior was used for this example.

3.1.2 Example 2 (d=7)

The proportional hazards model in this example was proposed by Lawless (1982) and has a relatively small sample size (n=65), with seven dimensions and five explanatory variables. The problem relates survival times in months for multiple myeloma patients to several prognostic variables: log of a blood urea nitrogen measurement
at diagnosis, hemoglobin measurement at diagnosis, age at diagnosis, sex and serum calcium measurement at diagnosis. The data used for this problem were presented by Lawless. They are a subset of the original data presented by Krall, Uthoff, and Harley (1975).

3.1.3 Examples 3, 4, and 5 (d=3, 5, and 8)

This ten observation macroeconometric model was used in an example presented by Johnston (1963, p.268ff) and examined by Kloek and van Dijk (1978) and van Dijk and Kloek (1980, 1984). Example 5 examines the problem using all eight parameters. Example 4 uses only five parameters after analytically integrating out the three covariance parameters. Example 3 considers the problem with two additional parameters integrated out analytically. This is the three dimensional form used by Kloek and van Dijk (1978).

3.1.4 Examples 6 and 7 (d=7 and 7)

These two examples of high dimension logistic regression utilize a study of the incidence of low birth weight in infants that was described by Hosmer and Lemeshow (1989, p.91). A subset of six explanatory variables (and an intercept) were used: mother’s age, weight, smoking, premature labors, hypertension, and uterine irritability. Data were collected on 189 women, 59 of whom had low birth weight babies and 130 of whom had normal birthweight babies. Example 6 uses one fifth of these
observations. Example 7 uses all of the observations.

### 3.1.5 Example 8, (d=9)

This contingency table model was presented by Evans, Gilula and Guttman (1989). One-hundred thirty-two long-term schizophrenic patients were classified into three row categories based on the frequency of hospital visits and three column categories involving the length of stay. Evans, Gilula and Guttman modeled a $3 \times 3$ cross-classification by Wing (1962) with the $(i, j)^{th}$ cell probability

$$p_{ij} = \theta \alpha_i(1)\beta_j(1) + (1 - \theta)\alpha_i(2)\beta_j(2)$$

where $\sum \alpha_i = 1$ and $\sum \beta_j = 1$. The prior distributions are:

$$\theta \sim U(0, 1/2)$$

$$(\alpha_1(i), \alpha_2(i), \alpha_3(i)) \sim \text{Dirichlet}(1,1,1), i = 1, 2$$

$$(\beta_1(i), \beta_2(i), \beta_3(i)) \sim \text{Dirichlet}(1,1,1), i = 1, 2$$

A logit transformation was used to transform $\theta$ and a similar transformation was used on the other eight parameters in pairs to extend the parameter space to $R^9$.

### 3.1.6 Example 9, d=10

This simulated example presented by Evans and Swartz (1995) has a model that is essentially a nine-level, one-way layout with independent and identically distributed
Student-t (with 3 degrees of freedom) errors and the log transformed scale parameter as the tenth parameter. The posterior distribution in this example takes the form:

\[ f(\theta) = \exp\{-9n\theta_{10}\} \prod_{i=1}^{9} \prod_{j=1}^{n} g_\lambda \left( \frac{y_{ij} - \theta_i}{\exp(\theta_{10})} \right) \]

where

\[ g_\lambda(z) = \frac{\Gamma((\lambda + 1)/2)}{\Gamma(1/2)\Gamma(\lambda/2)} \left( 1 + \frac{z^2}{\lambda - 2} \right)^{-\lambda/2} \frac{1}{\sqrt{\lambda - 2}} \]

with \( \lambda = 3 \) and \( n = 5 \). Curiously, posterior means and variances can be reduced to one-dimensional integrals.

### 3.1.7 Example 10, d=11

This eleven-dimensional ordinal regression problem for forecasting tornado intensity was used by Monahan, Schrab, and Anderson (1993). The sample size for this problem (157) is relatively large. There were five explanatory variables: two measurements of thunderstorm cell circulation, vorticity, storm relative ambient wind, and intercept. The seven ordinal levels accounted for the other six parameters.

### 3.2 Results of Simulation Experiment

Monte Carlo Importance Sampling was performed for each example using 40,000 evaluations and the EMVL as the importance sampler. The results from these simulations were compared to those obtained when the Multivariate Normal, the Multivariate Student-t with 3 degrees of freedom and the Multivariate Student-t with
5 degrees of freedom were used as importance samplers. For each simulation, the behavior of the importance sampler with respect to the posterior density (especially in the tails) was investigated. This was done in two ways:

1. The behavior of the weights in relation to \( \|v\| = \|L^{-1}(t - \hat{t})\| \) was examined, where \( -\nabla^2 \ln p(t) \bigg|_{t=\hat{t}} = LL' \). Weights that become increasingly large as \( \|v\| \) increases are evidence of tails that are “mismatched,” and indicate that the importance sampling estimates may not be asymptotically normally distributed. These analyses are presented in Section 3.2.1.

2. We used hypothesis testing suggested by Monahan (1993, 2001) to investigate the existence of the second moment of the weights. If the variance of the weights is not finite one should not assume asymptotic normality of the importance sampling estimates. In addition, it could indicate the non-existence of other expectations, such as the posterior variance (Monahan, 2001). These hypothesis tests are discussed in Section 3.2.2.

### 3.2.1 Evaluation of the Weights as a Function of the Radius

To ascertain how well the importance samplers matched the posterior distributions in these ten examples, we looked at how the weights behaved in relation to \( \|v\| = \|L^{-1}(t - \hat{t})\| \), where \( -\nabla^2 \ln p(t) \bigg|_{t=\hat{t}} = LL' \). Weights that decrease as the radius increases indicate that the importance sampler has thicker tails than the posterior distribution. In contrast, if the weights increase as the radius increases, one should
suspect that the importance sampler has thinner tails than the posterior distribution, and one should not assume the importance sampling estimates are asymptotically normal. Plots of $w(T^{(i)})/p^*(\hat{t}) = p^*(T^{(i)})/[p^*(\hat{t}) g(T^{(i)})]$ by $\|v\|$ for Examples 1-10 are given in Figures 3.1-3.10 for the four importance samplers (EMVL, MVN, MVT with 3 degrees of freedom, and the MVT with 5 degrees of freedom).

Judging from the plots, the EMVL matches the tails of the posterior distributions for Examples 2, 6, 7, 8, 9 and 10, and provide evidence that the requirements appealing to the CLT have been met. Since the tails were not matched as well in Examples 1, 3, 4, and 5 asymptotic normality of these estimates should not be assumed.

Similar conclusions can be drawn from the plots associated with the MVT with 3 degrees of freedom, and the MVT with 5 degrees of freedom. The tails appear well matched for Examples 2, 6, 7, 8, 9 and 10, but not for Examples 1, 3, 4 and 5. The MVN had thinner tails than the posterior distribution for all the Examples. One should not assume that the estimates from these simulations are asymptotically normal.
Figure 3.1: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 1. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.2: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 2. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.3: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 3. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.4: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 4. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.5: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 5. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.6: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 6. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.7: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 7. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.8: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 8. The horizontal axes are $\|v\| = \|L^{-1}(t - \hat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.9: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 9. The horizontal axes are $\|\mathbf{v}\| = \|\mathbf{L}^{-1}(\mathbf{t} - \hat{\mathbf{t}})\|$. A trend of weights that increase as $\|\mathbf{v}\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
Figure 3.10: Importance sampling weights resulting from four importance samplers (EMVL, MVN, MVT with 5 degrees of freedom and 3 degrees of freedom) used to estimate the normalization constant of the posterior distribution in Example 10. The horizontal axes are $\|v\| = \|L^{-1}(t - \widehat{t})\|$. A trend of weights that increase as $\|v\|$ increases is an indication that the importance sampling estimates may not be asymptotically normal.
3.2.2 Evaluation of the Existence of the First and Second Moments of the Weights

This section evaluates the existence of the first and second moments of the weights produced for Examples 1-10. The distribution function of the weights can be modeled as (Monahan, 2001):

$$1 - F(w) = Cw^{-1/\beta}(1 + Dw^{-1} + o(w^{-1})).$$

The term $1/\beta$ has been referred to as the coefficient of regular variation. A common example given for $\beta = 1$ is the distribution of a Cauchy random variable, which has an undefined mean, and hence its variance does not exist. Similarly, a common example for $\beta = 1/2$ is the Student t distribution with 2 degrees of freedom. This distribution has finite mean but infinite variance. These approximations can be used to construct hypotheses tests for finite means and variances.

The hypotheses to test for finite mean are:

$$H_0 : \beta \geq 1\text{(infinite mean)}$$

$$H_1 : \beta < 1\text{(finite mean)}.$$ 

The hypotheses for finite variance are:

$$H_0 : \beta \geq 1/2\text{(infinite variance)}$$

$$H_1 : \beta < 1/2\text{(finite variance)}.$$
To carry out these tests, an estimate of \( \beta \) is needed. One of the most widely used estimates was proposed by Hill (1975). If \( W(j) \) is the \( j^{th} \) smallest order statistic from a sample size of \( n \) weights then the Hill estimate is given by:

\[
\hat{\beta} = \frac{1}{k} \sum_{j=1}^{k} \log(W_{(n-j+1)}) - \log(W_{(n-k)}).
\]

Determining the best technique for choosing \( k \) has been the topic of some discussion. One technique is to produce a Hill plot by graphing \( f(k, \hat{\beta}_k), 1 \leq k \leq n \). The investigator then attempts to determine the value of \( \beta \) from a stable region of the plot. An alternative to this plot was suggested by Resnick and Starcia (1997), who propose graphing \( \{(\theta, \hat{\beta}_{n\theta}), 0 \leq \theta \leq 1\} \), where \( k = n^\theta \). This method of displaying the data can make it easier to identify a stable region of the graph. As an alternative, Monahan (2001) suggests that setting \( k = 4n^{1/3} \) works well for many problems.

Hill plots as proposed by Resnick and Starcia (1997) were generated for each example using 40,000 evaluations and the EMVL as the importance sampler. These plots are shown in Figures 3.11 - 3.15. For comparison, the value of \( \hat{\beta} \) for \( k = 4n^{1/3} \) is plotted on the graphs as a horizontal reference line. Estimating \( \hat{\beta} \) using \( k = 4n^{1/3} \) produces similar estimates to those chosen from determining a stable region of the graph for these examples. Based on this result, \( k = 4n^{1/3} \) was used to calculate \( \hat{\beta} \) throughout this document.

The result from Haeusler and Teugels (1985) and Hall (1982) can be used to carry out the hypothesis tests for finite mean and variance (Monahan, 2001). If
Table 3.1: Estimates of $\beta$ for Examples 1-10

<table>
<thead>
<tr>
<th>Example</th>
<th>EMVL</th>
<th>MVN</th>
<th>MVT (5 df)</th>
<th>MVT (3 df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5948 *</td>
<td>0.7194 *</td>
<td>0.5059 *</td>
<td>0.3067</td>
</tr>
<tr>
<td>2</td>
<td>0.1500</td>
<td>0.4777 *</td>
<td>0.0654</td>
<td>0.0419</td>
</tr>
<tr>
<td>3</td>
<td>0.5755 *</td>
<td>0.8016 *</td>
<td>0.5254 *</td>
<td>0.4922 *</td>
</tr>
<tr>
<td>4</td>
<td>0.6182 *</td>
<td>0.7056 *</td>
<td>0.6059 *</td>
<td>0.5024 *</td>
</tr>
<tr>
<td>5</td>
<td>0.7131 *</td>
<td>0.9271 *</td>
<td>0.6328 *</td>
<td>0.6308 *</td>
</tr>
<tr>
<td>6</td>
<td>0.1283</td>
<td>0.6546 *</td>
<td>0.2966</td>
<td>0.2096</td>
</tr>
<tr>
<td>7</td>
<td>0.0400</td>
<td>0.3532</td>
<td>0.0147</td>
<td>0.0260</td>
</tr>
<tr>
<td>8</td>
<td>0.4064</td>
<td>0.8153 *</td>
<td>0.7878 *</td>
<td>0.5345 *</td>
</tr>
<tr>
<td>9</td>
<td>0.2033</td>
<td>0.6331 *</td>
<td>0.1978</td>
<td>0.1818</td>
</tr>
<tr>
<td>10</td>
<td>0.3554</td>
<td>0.5472 *</td>
<td>0.1548</td>
<td>0.1559</td>
</tr>
</tbody>
</table>

* indicates null hypothesis of infinite variance was not rejected

$k = o(n^{\beta/(\beta+0.5)})$ then

$$\sqrt{k}(\hat{\beta} - \beta) \sim AN(0, \beta^2)$$

The null hypothesis of infinite mean is rejected if $\hat{\beta} < 1 - z_{\alpha}/k$ where $k = 4n^{1/3}$ and $z_{\alpha}$ is the upper $\alpha$ critical value of the standard normal distribution. Similarly, the null hypothesis of infinite variance is rejected if $\hat{\beta} < (1 - z_{\alpha}/k)/2$. The values of $\hat{\beta}$ from the simulation experiments are given in Table 3.1, where $k = 133$ and the $\alpha = 0.05$ critical value for testing for finite mean is 0.8573, for finite variance is 0.4286.

As can be seen from the table, there are some examples where the conclusions about the reliability of the importance sampling estimates differ from those of the previous section. The null hypothesis of infinite variance was rejected for Example 1, when the MVT with 3 degrees of freedom was used as an importance sampler and for Example 7, where the MVN was used as the importance sampler, giving no
indication that the estimates of $E_p[h(t)]$ are not asymptotically normal. However, one should exercise caution in relying on these estimates since the plots of the weights as a function of the radius indicated that the importance samplers did not match the tails of the posterior distributions well. Additionally, the null hypothesis of infinite variance of the weights was not rejected for Example 8 for the MVT with 3 and 5 degrees of freedom. These importance sampling estimates should not be relied upon.
Figure 3.11: Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 1, the bottom to Example 2. A horizontal reference line is drawn at the value of $\hat{\beta}$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).
Figure 3.12: Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 3, the bottom to Example 4. A horizontal reference line is drawn at the value of $\hat{\beta}$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).
Figure 3.13: Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 5, the bottom to Example 6. A horizontal reference line is drawn at the value of $\tilde{\beta}$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).
Figure 3.14: Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 7, the bottom to Example 8. A horizontal reference line is drawn at the value of $\beta$ produced using $k = 4n^{1/3}$ as suggested by Monahan (2001).
Figure 3.15: Hill Plots, as described in Section 3.2.2. The plots were produced using the EMVL as the importance sampler. The top figure corresponds to Example 9, the bottom to Example 10. A horizontal reference line is drawn at the value of \( \hat{\beta} \) produced using \( k = 4n^{1/3} \) as suggested by Monahan (2001).
3.2.3 Relative Standard Errors

The relative standard errors (or coefficients of variation) of the estimates of $E_p[h(t)]$ can be used to compare the performance of the four importance samplers. Since $h(t) = 1$ (normalization constant), the relative standard errors were calculated as:

$$\frac{(n - 1)^{-1}(\sum_{i=1}^{n} w_i^2 - \frac{1}{n}(\sum_{i=1}^{n} w_i)^2)}{\sum_{i=1}^{n} w_i},$$

where $w_i$ is the $i^{th}$ weight calculated using $n$ iterations. The two previous sections evaluated the reliability of the importance sampling estimates, and as a conservative measure, relative standard errors are not reported if either the plot of the weights versus the radius or the test for finite variance of the weights indicated that the importance sampler did not match the posterior distribution well.

The relative standard errors are given in Table 3.2. The MVN did such a poor job of matching the posterior distributions that relative standard errors were not reported for any of the examples. The EMVL outperformed the MVT and for Example 2; the relative standard error is 0.0054 compared to 0.0645 and 0.0419 for the MVT with 5 and 3 degrees of freedom respectively. The EMVL and the MVT seemed to perform equally well for Examples 6 and 9. The MVT with 3 degrees of freedom performed the best for Example 7, although the EMVL and the MVT with 5 degrees of freedom produced relatively small standard errors as well. The EMVL was the only one of the four importance samplers that matched the posterior distribution in Example 8, and produced a relative standard error of 0.025. Based on these results, one can
Table 3.2: Relative standard errors of the importance sampling estimates of the posterior distributions for Examples 1-10. Relative standard errors were not given when empirical evidence suggested that the importance sampling estimates were not asymptotically normal.

conclude that the EMVL is a viable alternative to the MVT and the MVN for use as an importance sampler.

3.3 Summary

The simulation studies in this chapter demonstrated the advantages of using the EMVL as an importance sampler. Three common importance samplers, the MVN, the MVT with 3 and 5 degrees of freedom, were compared to the EMVL using simulation experiments to calculate the normalization constant of ten different posterior distributions chosen from the literature. This empirical study demonstrated that the EMVL can perform better than two of the most commonly used importance samplers, the MVT and the MVN, resulting in more reliable estimates of the expectation of
interest for a range of posterior distributions.

When the EMVL was used as the importance sampler the normalization constant and its relative standard error were successfully estimated for six of the ten posterior distributions. In contrast, the MVT with 3 and 5 degrees of freedom performed well for only five of the ten examples, and the MVN did not work for any of the examples. A comparison of the relative standard errors produced by these simulations demonstrated that the relative standard errors produced by the MVT and the EMVL were comparable for four of the examples. For one of the examples the relative standard error produced by the EMVL was dramatically smaller than those produced by the MVT (0.005 compared to 0.06 and 0.04), resulting in more precise estimates of $E_{p^*}[h(t)]$. 
Chapter 4

Effect of Scale Factor

In importance sampling one traditionally matches the posterior and the importance sampler at the mode of the posterior distribution, \( \hat{t} \), by setting \( \nabla \ln g_T(t) |_{t=\hat{t}} = 0 \) and \( -\nabla^2 \ln p(t) |_{t=\hat{t}} = -\nabla^2 \ln g_T(t) |_{t=\hat{t}} \). However, this can result in poor performance from an importance sampler when the Hessian of the posterior distribution evaluated at the mode does not reflect the behavior of its tails (Geweke, 1989). This problem can be addressed by modifying how the spread of the importance sampler relates to the spread of the posterior distribution at the mode.

The MVN, the MVT (5 degrees of freedom), the MVT (3 degrees of freedom) and the EMVL are defined by three parameters: the dimension, \( d \); a centering vector, \( \mu \); and a scaling matrix, \( \Sigma \). When these importance samplers are matched to the posterior at the mode, \( \Sigma^{-1} = -c\nabla^2 \ln p^*(t) |_{t=\hat{t}} \), where \( c = 1 \) for the MVT and MVN distributions, and \( c = 2 \) for the EMVL. If \( c \) is increased from these “standard” values,
the tails of the importance sampling distribution become thinner. They become thicker when \( c \) is decreased. Changing the value of \( c \) can improve the performance of the importance sampler when there is reason to suspect that the behavior of the tails of the posterior distribution does not mimic the behavior of the posterior at the mode. A method for selecting \( c \) is proposed in the next section.

4.1 Selecting \( c \)

Two different techniques for selecting \( c \) are considered. Since there is a high coefficient of variation (CV) when a few weights dominate the sample (Evans and Swartz, 1995) one option is to choose the \( c \) so that the CV is minimized. Another option is to choose the value of \( c \) that results in the smallest value of \( \hat{\beta} \) (as defined in Section 3.2.2), since large values of \( \hat{\beta} \) can be indicative of infinite or nonexistent variance.

To investigate these options, a simulation experiment using the EMVL as the importance sampler was run. The CV’s and \( \hat{\beta} \) were calculated using \( n = 4,000 \) evaluations for each of the ten examples from Chapter 3 for values of \( c \) that change by increments of 0.25. The increment of 0.25 was chosen so that a reasonable number of scale factors were investigated without becoming impractical to implement. One could just as easily have chosen increments of 0.125 or 0.5. If one chooses too small an increment, the method becomes of questionable practical value, as it will require a large number of evaluations of the posterior distribution. In contrast, if the incre-
ments chosen are too large, the procedure might miss a value of $c$ that would provide improvement.

Figures 4.1-4.10 display plots of the CV’s and $\hat{\beta}$ for the ten examples. There does not seem to be a clear choice of $c$ from the plots of the coefficient of variation for any of these examples. While the coefficient of variation is certainly lower for some values of $c$ than others, there is not one value that seems to result in significantly lower values of the CV than the others. On the other hand one can easily determine a value of $c$ that produces the smallest $\hat{\beta}$ for Examples 1, 3, and 4. The choice is not as dramatically clear for Example 5, but $c = 1$ produces a smaller $\hat{\beta}$ than the other values. These are the same examples for which it was determined that the EMVL did not produce reliable results when matched at the mode. There is not a clear choice of $c$ for the remaining examples, but the EMVL performed well in these examples when matched to the mode of the posterior.

### 4.2 Evaluating Consistency of Selection

To determine how consistently a particular a value of $c$ minimizes the estimate of $\hat{\beta}$, 100 simulations were run for each example. The counts of the values of $c$ that minimized $\hat{\beta}$ from these simulations are displayed in Table 4.1. Altering the scale factor, $c$, from its standard value of $c = 2$ has the effect of either “thickening” or “thinning” the tails of the importance sampler. All 100 simulations resulted in a $c$ that produces thicker importance sampling tails for Examples 1, 3, 4 and 6, and
thinner tails for Examples 2 and 10. Most of the simulations (92) suggested thicker tails for Example 5. The rest of the simulations did not produce such consistent results. The simulation results for Examples 7, 8 and 9 suggested that the standard scaling worked fairly well, but also suggested both thicker and thinner tails. Most importantly, the simulation experiments resulted in consistent suggestions for altering the scale for the four examples (1, 3, 4 and 5) where the EMVL was not a successful importance sampler when $c = 2$. The next section examines the effect of using the value of $c$ that minimizes $\hat{\beta}$ on importance sampling results.
<table>
<thead>
<tr>
<th>Example</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
<th>1.75</th>
<th>2.00</th>
<th>2.25</th>
<th>2.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33</td>
<td>67</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th>2.25</th>
<th>2.5</th>
<th>2.75</th>
<th>3.00</th>
<th>3.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>45</td>
<td>38</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
<th>1.75</th>
<th>2.00</th>
<th>2.25</th>
<th>2.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>97</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>91</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>26</td>
<td>49</td>
<td>13</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th>1.25</th>
<th>1.5</th>
<th>1.75</th>
<th>2.0</th>
<th>2.25</th>
<th>2.5</th>
<th>2.75</th>
<th>3.0</th>
<th>3.25</th>
<th>3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>10</td>
<td>77</td>
<td>13</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>70</td>
<td>9</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>16</td>
<td>36</td>
<td>29</td>
<td>11</td>
<td>6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>47</td>
<td>37</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Example</th>
<th>3.50</th>
<th>3.75</th>
<th>4.00</th>
<th>4.25</th>
<th>4.50</th>
<th>4.75</th>
<th>5.00</th>
<th>5.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7</td>
<td>25</td>
<td>27</td>
<td>23</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4.1: Resulting counts of the values of c chosen from 100 simulations for each example, using the EMVL as the importance sampler.
Figure 4.1: Estimates of \( \beta \) and coefficient of variation calculated for Example 1 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, \( c \) that was used in matching the EMVL to the posterior distribution, \( p^*(t) \) by setting \( \Sigma^{-1} \), the inverse of the scaling matrix of the EMVL, equal to \( -c\nabla^2 \ln p^*(t) \mid_{t=t} \).
Figure 4.2: Estimates of $\beta$ and coefficient of variation calculated for Example 2 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t}$. 
Figure 4.3: Estimates of $\beta$ and coefficient of variation calculated for Example 3 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c \nabla^2 \ln p^*(t) \big|_{t=t}$. 
Figure 4.4: Estimates of $\beta$ and coefficient of variation calculated for Example 4 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) |_{t=t}$. 
Figure 4.5: Estimates of $\beta$ and coefficient of variation calculated for Example 5 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \big|_{t=t}$.
Figure 4.6: Estimates of $\beta$ and coefficient of variation calculated for Example 6 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \big|_{t=t}$. 
Figure 4.7: Estimates of $\beta$ and coefficient of variation calculated for Example 7 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \big|_{t=t}$. 
Figure 4.8: Estimates of $\beta$ and coefficient of variation calculated for Example 8 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c^2 \ln p^*(t) \big|_{t=t}$. 
Figure 4.9: Estimates of $\beta$ and coefficient of variation calculated for Example 9 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c \nabla^2 \ln p^*(t) \big|_{t=t}$. 

\[ \text{beta} \]

\[ \text{CV} \]

\[ \text{Scale} \]
Figure 4.10: Estimates of $\beta$ and coefficient of variation calculated for Example 10 using the EMVL as the importance sampler. The horizontal axes represent the scale factor, $c$ that was used in matching the EMVL to the posterior distribution, $p^*(t)$ by setting $\Sigma^{-1}$, the inverse of the scaling matrix of the EMVL, equal to $-c\nabla^2 \ln p^*(t) \mid_{t=t}$. 
4.3 Results Produced When $c$ is Altered for the EMVL

A simulation experiment was used to evaluate the effect of changing $c$. Table 4.2 lists the value of $c$ that produced the smallest $\hat{\beta}$ for each example. Importance sampling using these values of $c$ and 40,000 evaluations was performed for each example. The weights that resulted were examined to determine how well the EMVL matched the posterior distribution when the new scale was used. Plots of the weights as a function of the radius, $\|v\| = \|L^{-1}(t - \hat{t})\|$, where $(LL')^{-1} = -\nabla^2 \ln p(t) \mid_{t=\hat{t}}$ were produced for each example. These plots are shown in Figures 4.11-4.20. Since the weights tended to decrease as the radius increased, changing the scale had a clear beneficial effect on the weights in Examples 1, 3, 4 and 5. Little or no improvement was seen for the remaining examples - but these plots did not indicate that the tails were mismatched when the Hessians were matched at the mode. Hypothesis testing as described in Chapter 3 was used to investigate the finiteness of the variance. Again, $k = 4n^{1/3}$ was used to calculate

$$\hat{\beta} = \frac{1}{k} \sum_{j=1}^{k} \log(W_{(n-j+1)}) - \log(W_{(n-k)}).$$

The results of these hypotheses tests are shown in Table 4.2. Changing the scale had a negative impact on the results for Examples 6 and 8, since the null hypothesis of infinite variance of the weights was not rejected in these examples, but had been rejected when $c = 2$ was used. There was little effect for Examples 2, 7, 9 and
10. There was a definite beneficial effect on Examples 1, 3, 4 and 5 as the null hypothesis of infinite variance was rejected, in contrast to the results obtained when the EMVL was matched to the posterior distribution at the mode. If one compares the relative standard errors from Table 4.2 it is apparent that altering $c$ resulted in higher relative standard errors for Examples 7 and 9, and lower standard errors for Examples 2 and 10. This shows that, for simulations where the importance sampler was successfully matched at the mode, changing $c$ did not guarantee a reduction in the relative standard errors.

Based on these results, we concluded that when the traditional approach works well, altering $c$ had either no effect or a negative effect. However, for a number of examples, if an examination of the weights indicated that the importance sampling estimate was not asymptotically normal when the posterior distribution and importance sampler were matched at the mode, changing the scale improved the situation.
Table 4.2: Table of $\tilde{\beta}$, results of test of hypothesis of infinite variance, and relative standard errors produced when an alternate scale was used to match the importance sampler, the EMVL, to the posterior distributions. For comparison, relative standard errors that resulted using the usual scale, $c = 2$ are also given. Relative standard errors were not reported if there was an indication that the importance sampling estimates were not asymptotically normal.

<table>
<thead>
<tr>
<th>Example</th>
<th>Scale</th>
<th>$\beta_{new}$</th>
<th>infinite variance?</th>
<th>s.e.</th>
<th>s.e. (old)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.50</td>
<td>0.0653</td>
<td>rejected</td>
<td>0.0099</td>
<td>**</td>
</tr>
<tr>
<td>2</td>
<td>2.50</td>
<td>0.2080</td>
<td>rejected</td>
<td>0.0042</td>
<td>0.0050</td>
</tr>
<tr>
<td>3</td>
<td>0.50</td>
<td>0.0552</td>
<td>rejected</td>
<td>0.0115</td>
<td>**</td>
</tr>
<tr>
<td>4</td>
<td>0.75</td>
<td>0.3025</td>
<td>rejected</td>
<td>0.0140</td>
<td>**</td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
<td>0.4168</td>
<td>rejected</td>
<td>0.0272</td>
<td>**</td>
</tr>
<tr>
<td>6</td>
<td>0.50</td>
<td>0.6459</td>
<td>not rejected</td>
<td>**</td>
<td>0.0089</td>
</tr>
<tr>
<td>7</td>
<td>1.50</td>
<td>0.1292</td>
<td>rejected</td>
<td>0.0146</td>
<td>0.0101</td>
</tr>
<tr>
<td>8</td>
<td>2.25</td>
<td>0.4461</td>
<td>not rejected</td>
<td>**</td>
<td>0.0250</td>
</tr>
<tr>
<td>9</td>
<td>1.75</td>
<td>0.2862</td>
<td>rejected</td>
<td>0.0222</td>
<td>0.0180</td>
</tr>
<tr>
<td>10</td>
<td>3.75</td>
<td>0.1680</td>
<td>rejected</td>
<td>0.0121</td>
<td>0.0327</td>
</tr>
</tbody>
</table>
Figure 4.11: Example 1: Comparison of importance sampling weights for conventional and alternate scale ($c$) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$.
Figure 4.12: Example 2: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).
Figure 4.13: Example 3: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).
Figure 4.14: Example 4: Comparison of importance sampling weights for conventional and alternate scale (c) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$.
Figure 4.15: Example 5: Comparison of importance sampling weights for conventional and alternate scale ($c$) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or $\|\mathbf{v}\| = \|L^{-1}(\mathbf{t} - \hat{\mathbf{t}})\|$. 
Figure 4.16: Example 6: Comparison of importance sampling weights for conventional and alternate scale (c) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$. 
Figure 4.17: Example 7: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).
Figure 4.18: Example 8: Comparison of importance sampling weights for conventional and alternate scale (c) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$. 
Figure 4.19: Example 9: Comparison of importance sampling weights for conventional and alternate scale (c) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or $||v|| = ||L^{-1}(t - \hat{t})||$. 
Figure 4.20: Example 10: Comparison of importance sampling weights for conventional and alternate scale (c) when the EMVL was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$.
4.4 Recommendations for Selecting $c$

Altering the scale parameter can dramatically improve the behavior of the weights when one fails to get desirable results from matching the importance sampler to the posterior distribution. However, if the importance sampler performed well when it was matched to the posterior distribution at the mode, altering the scale parameter did little to improve the situation (and in a few cases, worsened it.)

Although it would be desirable to automate the selection of $c$, these simulations illustrate the difficulties involved in doing so. The estimates of $\hat{\beta}$ do not produce a “smooth” curve – even for situations where there is a $c$ that clearly produces a minimum value, $\hat{\beta}$ does not consistently increase as one moves away from that point. This makes it difficult to automate a method for choosing $c$. Additionally, it is important to remember that if the importance sampler and the posterior distribution are well matched at the mode altering the value of $c$ can produce undesirable results. Therefore, the results obtained when $c = 2$ should be investigated before one considers altering $c$.

Taking these considerations into account, the following procedure is recommended:

1. Perform Monte-Carlo importance sampling, matching the importance sampler to the posterior distribution at the mode: $\nabla \ln f_X(x) \big|_{x=\hat{t}} = 0$ and $-\nabla^2 \ln p^*(t) \big|_{t=\hat{t}} = -\nabla^2 \ln g(t) \big|_{t=\hat{t}}$.

2. Evaluate the weights by testing for infinite variance, and examining the plot of
the weights vs \( \|v\| \).

3. If a problem with the weights is indicated, choose \( c \) so that \( \hat{\beta} \) is minimized.

4. Evaluate the weights that result from the new scale factor.

### 4.5 Changing Scale with MVT and the MVN

If changing the scale for the EMVL improves the behavior of the weights, changing the scale factor for the MVT and the MVN should have a beneficial effect as well. When the MVN was matched to the posterior distribution at the mode it did not perform well as an importance sampler for any of the ten examples. There was evidence that the tails of the MVN were thinner than those of the posterior distribution, and that the variance of the weights was not finite. Alternate scales for each example were chosen using the method described in the previous section. These scales and the corresponding \( \hat{\beta} \)'s are given in Table 4.3. Changing the scale factor improved the situation for seven of the examples, but infinite variance still cannot be ruled out for Examples 4, 5, and 6. Plots of the weights versus radius are given in Figures 4.21-4.30. These plots show marked improvement, leading one to conclude that the MVN can be used for any of the examples where the null hypothesis of infinite variance was rejected (Examples 1, 2, 3, 7, 8, 9, 10).

There were five examples (1, 3, 4, 5, and 8) where the MVT with 5 degrees of freedom and 3 degrees of freedom were unsuccessful as importance samplers when
Table 4.3: Hypothesis test for infinite variance of the importance sampling weights for importance samplers MVN, MVT (5 degrees of freedom), MVT (3 degrees of freedom) using an alternate scale.

c = 1. The values of c that were chosen for each example are displayed in Table 4.3. Using these values of c, infinite variance could not be rejected for Example 5 when either of the MVT distributions were used, or for Example 8 when the MVT with 5 degrees of freedom was used. Figures 4.31-4.38 show the plots of the weights versus \|v\| for Examples 1, 3, 4, and 5 that resulted from using the alternate scales for the importance samplers MVT with 3 degrees of freedom and MVT with 5 degrees of freedom. When these two importance samplers were matched to the posterior distributions at the mode an examination of the plots indicated that the tails of the importance sampler were thinner than those of the posterior distribution. When the scale factor c was altered, the plots improved dramatically, and give no indication that the importance sampling estimates are not asymptotically normal.
Table 4.4: Comparison of relative standard errors for Examples 1-10

The relative standard errors produced by each of the importance samplers are given in Table 4.4. Relative standard errors are not given when there was an indication that the importance sampler estimates were not asymptotically normal. The EMVL, MVN and the MVT distributions produced similar relative standard errors, however, the EMVL was successful for all examples, including Example 5 where both the MVN and MVT distributions failed, and for Example 8 where the MVT with 5 degrees of freedom failed. In addition, choosing to use the EMVL as an importance sampler has the added benefit that the investigator does not have to choose among various degrees of freedom, as is necessary for the MVT distributions.
Figure 4.21: Example 1: Comparison of importance sampling weights for conventional and alternate scale (c) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$.
Figure 4.22: Example 2: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \(\|\mathbf{v}\| = \|L^{-1}(\mathbf{t} - \hat{\mathbf{t}})\|\).
Figure 4.23: Example 3: Comparison of importance sampling weights for conventional and alternate scale (c) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).
Figure 4.24: Example 4: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).
Figure 4.25: Example 5: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( ||v|| = ||L^{-1}(t - \hat{t})|| \).
Figure 4.26: Example 6: Comparison of importance sampling weights for conventional and alternate scale (c) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$.
Figure 4.27: Example 7: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \hat{t})\|\).
Figure 4.28: Example 8: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).
Figure 4.29: Example 9: Comparison of importance sampling weights for conventional and alternate scale ($c$) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or $\|\mathbf{v}\| = \|L^{-1}(\mathbf{t} - \hat{\mathbf{t}})\|$. 
Figure 4.30: Example 10: Comparison of importance sampling weights for conventional and alternate scale (c) when the MVN was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$.
Figure 4.31: Example 1: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (5 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \tilde{t})\|\).
Figure 4.32: Example 3: Comparison of importance sampling weights for conventional and alternate scale ($c$) when the MVT (5 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or $\|\mathbf{v}\| = \|L^{-1}(\mathbf{t} - \mathbf{\hat{t}})\|$. 
Figure 4.33: Example 4: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVT (5 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t - \hat{t})\| \).
Figure 4.34: Example 5: Comparison of importance sampling weights for conventional and alternate scale \( (c) \) when the MVT (5 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \( \|v\| = \|L^{-1}(t-t)\| \).
Figure 4.35: Example 1: Comparison of importance sampling weights for conventional and alternate scale ($c$) when the MVT (3 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or $||v|| = ||L^{-1}(t - \hat{t})||$. 
Figure 4.36: Example 3: Comparison of importance sampling weights for conventional and alternate scale \(c\) when the MVT (3 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \tilde{t})\|\).
Figure 4.37: Example 4: Comparison of importance sampling weights for conventional and alternate scale (c) when the MVT (3 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or $\|v\| = \|L^{-1}(t - \hat{t})\|$. 
Figure 4.38: Example 5: Comparison of importance sampling weights for conventional and alternate scale \((c)\) when the MVT (3 degrees of freedom) was used as the importance sampler. The horizontal axes represent the radius, or \(\|v\| = \|L^{-1}(t - \tilde{t})\|\).
4.6 Detailed Analysis of Example 8

This section examines Example 8 in greater detail to demonstrate how the EMVL performs as an importance sampler and how the introduction of a scaling factor, \(c\), can improve importance sampling results. This example was described by Evans, Gilula and Guttman (1989) and Evans and Swartz (1995). A Bayesian analysis of a contingency table involving a cross-classification of 132 long-term schizophrenic patients into three row categories (frequency of hospital visits) and three column categories (length of stay) was used to fit a latent variable model. Table 4.5 displays the data as given by Wing (1962). As described in Chapter 3, the \((i, j)^{th}\) cell probability is:

\[
p_{ij} = \theta \alpha_i(1) \beta_j(1) + (1 - \theta) \alpha_i(2) \beta_j(2)
\]

where

\[
1 = \sum_{i=1}^{3} \alpha_i(1) = \sum_{i=1}^{3} \alpha_i(2) = \sum_{j=1}^{3} \beta_i(1) = \sum_{j=1}^{3} \beta_i(2)
\]

Independence is assumed and the prior distributions are: \(\theta \sim U(0, 1/2)\), \((\alpha_1(i), \alpha_2(i), \alpha_3(i)) \sim \text{Dirichlet}(1,1,1)\), and \((\beta_1(i), \beta_2(i), \beta_3(i)) \sim \text{Dirichlet}(1,1,1)\) for \(i = 1, 2\). A transformation to \(t\) is performed so that:

\[
\begin{align*}
\theta &= \frac{\exp(t_1)}{2(1 + \exp(t_1))} \\
\alpha_1(1) &= \frac{\exp(t_2)}{(1 + \exp(t_2) + \exp(t_3))} \\
\alpha_2(1) &= \frac{\exp(t_3)}{(1 + \exp(t_2) + \exp(t_3))} \\
\alpha_3(1) &= \frac{1}{(1 + \exp(t_2) + \exp(t_3))}
\end{align*}
\]
Monte Carlo importance sampling using 40,000 evaluations was performed using three different importance samplers: the MVN, the MVT with 3 df, and the EMVL. Using the results of the previous section, the scale factor $c = 0.25, 0.75, \text{ and } 2$, respectively.

In Section 4.5 importance sampling was used to estimate the normalization constant of the posterior distribution, and the relative standard errors of these estimates were reported. In this section, as part of a more detailed analysis, a simulation experiment was used to calculate 100 estimates, $Z_i$, of the normalization constant of the posterior distribution. Each estimation and corresponding standard error ($s_i$) was calculated using 40,000 evaluations. The accuracy of these standard errors was investigated by examining the ratio of the RMEV $= \sqrt{\frac{\sum_{i=1}^{n} s_i^2}{n}}$ and a measurement of the true variability, $\sqrt{\frac{\sum_{i=1}^{n}(Z_i - \overline{Z})^2}{n}}$ as described in Chapter 2. If the standard errors
are reasonably consistent across simulations, the ratios, \( RMEV = \frac{\sqrt{\sum_{i=1}^{n}(Z_i - \bar{Z})^2}}{n} \), should cluster about one. Ratios that deviate significantly from one are evidence that the standard errors of the importance sampling estimates tend to underestimate (ratios less than one) or overestimate (ratios greater than one) the true variability. The ratios produced were: 0.96, 1.03, and 1.08 with standard errors equal to 0.14, for the MVN, MVT with 3 df and the EMVL, respectively. These ratios indicate that the standard errors produced by each of these important samplers are reasonable estimates of the true variability of the importance sampling estimates.

These importance samplers should provide reasonable estimates of the \( E[t] \). For comparison, the MVN with \( c = 1 \) was also used as an importance sampler. This distribution should not perform as well as the other three since an evaluation of the weights it produced indicated that its tails were not as thick as those of the posterior distribution, and hence asymptotic normality of these estimates cannot be assumed. Table 4.6 displays the estimates of \( E[t] \) (the posterior means) and their corresponding standard errors. Since the posterior distribution was unnormalized, the estimation of the normalization constant has to be taken into account when calculating the standard errors. If \( T^{(i)} \) is the \( i^{th} \) vector generated from the importance sampler, \( g \), then label

\[
W_i = \frac{p^*(T^{(i)})}{g(T^{(i)})} \\
Z_i = h(T^{(i)})W_i
\]
\[ S_{ZW} = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z})(W_i - \bar{W}) \]
\[ S_{WW} = \frac{1}{n-1} \sum_{i=1}^{n} (W_i - \bar{W})^2 \]
\[ S_{ZZ} = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i - \bar{Z})^2. \]

Then, if \( h(T) = t_j \), the standard errors for the \( j^{th} \) element of the posterior mean can be estimated by (Monahan, 2001)

\[
\frac{1}{\sqrt{nW}} [S_{ZZ} - 2\frac{\bar{Z}}{\bar{W}}S_{ZW} + \frac{\bar{Z}^2}{\bar{W}^2}]^{1/2}.
\]

The standard errors for the MVN when \( c = 1 \) are ten times greater than those produced using the other importance samplers, but since we were unable to reject the null hypothesis of infinite variance of the weights (Chapter 3), they should not be relied upon.

Table 4.7 displays the cell counts estimated using the three importance samplers \( (E(n \cdot p_{ij}) = 132 * \hat{p}_{ij}) \) for the 132 patients. The difference between the true cell counts and the estimated cell counts are shown in parentheses. Differences greater than one are italicized. It is apparent from this table that the MVN \((c=0.25)\), the MVT with 3 df, and the EMVL performed well for this example. The estimated cell counts are very close to the actual cell counts; cell (1,1) was the only cell where the difference was greater than 1. The MVN \((c = 1)\) did not perform as well since the counts for cells (1,2) and (3,1) differ from the actual accounts by more than 2, and differ by more than 1 for cells (1,3), (2,3) and (3,3). This is not unexpected, as the analysis of the weights in Chapter 3 indicated that the MVN did not match the
### Table 4.5: Frequency of visits by length of stay for 132 long-term schizophrenic patients (Wing, 1962)

<table>
<thead>
<tr>
<th>Frequency of visiting</th>
<th>At least 2 years but less than 10 years</th>
<th>At least 10 years but less than 20 years</th>
<th>At least 20 years</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goes home or visited regularly</td>
<td>43</td>
<td>16</td>
<td>3</td>
<td>62</td>
</tr>
<tr>
<td>Visited less than once a month does not go home</td>
<td>6</td>
<td>11</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>Never visited and never goes home</td>
<td>9</td>
<td>18</td>
<td>16</td>
<td>43</td>
</tr>
<tr>
<td>Totals</td>
<td>58</td>
<td>45</td>
<td>29</td>
<td>132</td>
</tr>
</tbody>
</table>

The results of the simulation studies in this chapter demonstrate the advantage of introducing a scaling factor that dictates how the importance sampler relates to the posterior distribution at the mode. The same ten examples and four importance samplers (the MVN, the MVT with 3 and 5 degrees of freedom, and the EMVL) that were investigated in Chapter 3 were used in this chapter. The scale factor was chosen that minimized $\bar{\beta}$, where $1/\beta$ is the coefficient of regular variation of the weights produced by importance sampling.

The use of a scale factor greatly expanded the number of examples for which the posterior distribution well when $c = 1$.

### 4.7 Summary

The results of the simulation studies in this chapter demonstrate the advantage of introducing a scaling factor that dictates how the importance sampler relates to the posterior distribution at the mode. The same ten examples and four importance samplers (the MVN, the MVT with 3 and 5 degrees of freedom, and the EMVL) that were investigated in Chapter 3 were used in this chapter. The scale factor was chosen that minimized $\bar{\beta}$, where $1/\beta$ is the coefficient of regular variation of the weights produced by importance sampling.
<table>
<thead>
<tr>
<th></th>
<th>EMVL</th>
<th>MVT 3 df</th>
<th>MVN c = 1</th>
<th>MVN c = 0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_1$</td>
<td>1.95</td>
<td>2.03</td>
<td>2.39</td>
<td>1.94</td>
</tr>
<tr>
<td>$t_2$</td>
<td>-2.53</td>
<td>-2.41</td>
<td>-1.75</td>
<td>-2.40</td>
</tr>
<tr>
<td>$t_3$</td>
<td>-0.49</td>
<td>-0.48</td>
<td>-0.60</td>
<td>-0.49</td>
</tr>
<tr>
<td>$t_4$</td>
<td>1.87</td>
<td>1.85</td>
<td>2.84</td>
<td>1.86</td>
</tr>
<tr>
<td>$t_5$</td>
<td>-0.44</td>
<td>-0.38</td>
<td>0.61</td>
<td>-0.33</td>
</tr>
<tr>
<td>$t_6$</td>
<td>-1.85</td>
<td>-1.85</td>
<td>0.00001</td>
<td>-1.89</td>
</tr>
<tr>
<td>$t_7$</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.16</td>
<td>0.024</td>
</tr>
<tr>
<td>$t_8$</td>
<td>2.95</td>
<td>2.99</td>
<td>2.89</td>
<td>3.05</td>
</tr>
<tr>
<td>$t_9$</td>
<td>1.97</td>
<td>1.97</td>
<td>1.90</td>
<td>2.03</td>
</tr>
</tbody>
</table>

Table 4.6: Importance Sampling Posterior Means and Corresponding Standard Errors for Example 8

<table>
<thead>
<tr>
<th>cell</th>
<th>actual</th>
<th>EMVL</th>
<th>MVT 3 df</th>
<th>MVN c=0.25</th>
<th>MVN c=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,1)</td>
<td>43</td>
<td>41.7 (-1.3)</td>
<td>41.6 (-1.4)</td>
<td>41.8 (-1.2)</td>
<td>43.7 (0.7)</td>
</tr>
<tr>
<td>(1,2)</td>
<td>16</td>
<td>16.8 (0.8)</td>
<td>16.2 (0.2)</td>
<td>16.5 (0.5)</td>
<td>18.9 (2.9)</td>
</tr>
<tr>
<td>(1,3)</td>
<td>3</td>
<td>3.5 (0.5)</td>
<td>3.5 (0.5)</td>
<td>3.4 (0.4)</td>
<td>4.9 (1.9)</td>
</tr>
<tr>
<td>(2,1)</td>
<td>6</td>
<td>5.7 (-0.3)</td>
<td>6.0 (0)</td>
<td>6.1 (0.1)</td>
<td>6.5 (0.5)</td>
</tr>
<tr>
<td>(2,2)</td>
<td>11</td>
<td>11.0 (0)</td>
<td>11.3 (0.3)</td>
<td>11.2 (0.2)</td>
<td>11.1 (0.1)</td>
</tr>
<tr>
<td>(2,3)</td>
<td>10</td>
<td>10.0 (0)</td>
<td>10.1 (0.1)</td>
<td>9.9 (0)</td>
<td>8.2 (-1.8)</td>
</tr>
<tr>
<td>(3,1)</td>
<td>9</td>
<td>8.9 (-0.1)</td>
<td>9.0 (0)</td>
<td>8.9 (-0.1)</td>
<td>5.9 (-3.1)</td>
</tr>
<tr>
<td>(3,2)</td>
<td>18</td>
<td>18.0 (0)</td>
<td>18.0 (0)</td>
<td>17.9 (-0.1)</td>
<td>18.1 (0.1)</td>
</tr>
<tr>
<td>(3,3)</td>
<td>16</td>
<td>16.4 (0.4)</td>
<td>16.3 (0.3)</td>
<td>16.2 (0.2)</td>
<td>14.7 (-1.3)</td>
</tr>
</tbody>
</table>

Table 4.7: Importance Sampling Cell Count Estimates for Example 8. The difference between these estimates and the actual cell counts given by Wing (1962) are shown in parenthesis.
importance samplers produced reliable estimates of the normalization constant of the posterior distribution. Again, even with the improvements due to the introduction of a scale factor, the EMVL was a successful importance sampler for more posterior distributions than were the other importance samplers in estimating the normalization constant of the posterior distribution. Based on the number of examples for which the EMVL, the MVN and the MVT with 3 and 5 degrees of freedom successfully estimated the normalization constant, the EMVL performed better than these importance samplers.

A scale factor was used for the examples in which the importance sampler failed to produce reliable estimates of the normalization constant when it was matched to the posterior distribution at the mode. With the introduction of the scale parameter, the EMVL successfully estimated the normalization constant for all ten of the examples. The MVN distribution worked for seven of the ten examples, the MVT with 5 degrees of freedom for eight, and the MVT with 3 degrees of freedom for nine. The EMVL worked better for these posterior distributions, and is an improvement over using the MVT distribution for which the investigator must choose the degrees of freedom that s/he believes will work best for a particular posterior distribution.

In order to fully demonstrate the advantages of using the EMVL and of introducing a scale factor, a detailed analysis of Example 8 was performed. It was shown that, with the introduction of a scale factor, $c$, the MVN, the MVT distribution with 3 degrees of freedom, and the EMVL were all successful importance samplers for
estimating the posterior means. Since an examination of the weights produced when
the MVT distribution with 5 degrees of freedom was used an importance sampler
indicated that the CLT should not be used, the advantage of using the EMVL in a
situation like this is clear. When the MVT distribution is used, the investigator must
choose the degrees of freedom. This choice can result in very different outcomes. One
does not have to determine the degrees of freedom when the EMVL is used as an
importance sampler, and it performed well on a range of posterior distributions from
the literature.
Appendix A

Fortran Code

A.1 Fortran code for Example 1

function rlglik(t)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
Use the_data
Implicit none
! log-likelihood function for the heart transplant data
!integer, parameter :: double = selected_real_kind(15,200)
real(kind=2) :: rlglik
real(kind=2) :: rlt1,rlt2,rlt3,rls
real(kind=2), dimension(3) :: t
integer :: i
!real(kind=2), dimension(82) :: sb, sa
!
! first check to see if out of range
if( t(1) .le. 0. ) go to 9
if( t(2) .le. 0. ) go to 9
if( t(3) .le. 0. ) go to 9
!
rlt1 = dlog(t(1))
rlt2 = dlog(t(2))
rlt3 = dlog(t(3))
!
rlglik = 0.
do i = 1,82
rls = t(2) + sb(i) + t(1)*sa(i)
if( rls .le. 0. ) go to 9
rls = dlog( rls )
rlglik = rlglik + id(i)*(rlt3+it(i)*rlt1) + t(3)*rlt2 &
- (t(3)+id(i))*rls
end do
! use unlikelihood so that hessian is pos def
rlglik = - rlglik
return
9 continue
! if out of range set to big number
rlglik = 9999.
return
end function rlglik

Data: it, sb, id, sa
  0  49  1  0
  0   5  1  0
  0  17  1  0
  0   2  1  0
  0  39  1  0
  0  84  1  0
  0   7  1  0
  0   0  1  0
  0  35  1  0
  0  36  1  0
  0 1400  0  0
  0   5  1  0
  0  34  1  0
  0  15  1  0
  0  11  1  0
  0   2  1  0
  0   1  1  0
  0  39  1  0
146

0  8  1  0
0 101  1  0
0   2  1  0
0 148  1  0
0   1  1  0
0   68  1  0
0   31  1  0
0   1  1  0
0   20  1  0
0 118  0  0
0   91  0  0
0  427  0  0
1   0  1  15
1  35  1  3
1  50  1  624
1   11  1  46
1  25  1  127
1   16  1  61
1  36  1  1350
1  27  1  312
1   19  1  24
1   17  1  10
1   7  1  1024
1   11  1  39
1   2  1  730
1  82  1  136
1  24  0  1379
1   70  1  1
1  15  1  836
1   16  1  60
1  50  0  1140
1  22  0  1153
1   45  1  54
1  18  1  47
1   4  1  0
1   1  1  43
1  40  0  971
1  57  0  868
1   0  1  44
1   1  0  780
1   20  1  51
1  35  0  710
<table>
<thead>
<tr>
<th>1</th>
<th>82</th>
<th>0</th>
<th>663</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>1</td>
<td>253</td>
</tr>
<tr>
<td>1</td>
<td>40</td>
<td>1</td>
<td>147</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>1</td>
<td>51</td>
</tr>
<tr>
<td>1</td>
<td>66</td>
<td>0</td>
<td>479</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>1</td>
<td>322</td>
</tr>
<tr>
<td>1</td>
<td>77</td>
<td>0</td>
<td>442</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>65</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>0</td>
<td>419</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>0</td>
<td>362</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td>1</td>
<td>56</td>
<td>1</td>
<td>228</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>65</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>0</td>
<td>264</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>0</td>
<td>193</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>0</td>
<td>196</td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>1</td>
<td>63</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>45</td>
<td>0</td>
<td>103</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
<td>43</td>
</tr>
</tbody>
</table>
A.2 Fortran code for Example 2

function rlglik(t)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
Use the_data
implicit none
! log-likelihood function for the Lawless survival problem
Real(kind=2), dimension(1) :: t
Real(kind=2) :: rho,zb,y, rlglik
!
integer :: i, j
!
 rho = dexp(t(1))
!
 rlglik = 0.
do i = 1,65
 zb = 0.d0
 do j = 1,6
 zb = zb + t(j+1)*g(i,j)
 End do !loop on j
 y = log(ti(i))
 rlglik = rlglik - exp( rho*y + zb )
 if( ic(i) .eq. 1 ) &
 & rlglik = rlglik + t(1) + (rho-1.)*y + zb
End do !loop on i
!
 ! use unlikelihood so that hessian is pos def
 rlglik = - rlglik
return
End

Data:

do i = 1,65
read(8,23) ti(i),(g(i,j),j=2,6)
g(i,1) = 1.d0
ic(i) = 1
if( i .gt. 48 ) ic(i) = 0
write(6,22) i,ic(i),(g(i,j),j=1,6)
center some of the variables

\[ g(i,2) = g(i,2) - 1.391769 \]
\[ g(i,3) = g(i,3) - 10.20154 \]
\[ g(i,4) = g(i,4) - 60.15385 \]
\[ g(i,6) = g(i,6) - 10.12308 \]

end do !loop on i

data to be read in:

1  2.218  9.4  67  0  10
1  1.940  12.0  38  0  18
2  1.519  9.8  81  0  15
2  1.748  11.3  75  0  12
2  1.301  5.1  57  0  9
3  1.544  6.7  46  1  10
5  2.236  10.1  50  1  9
5  1.681  6.5  74  0  9
6  1.362  9.0  77  0  8
6  2.114  10.2  70  1  8
6  1.114  9.7  60  0  10
6  1.415  10.4  67  1  8
7  1.978  9.5  48  0  10
7  1.041  5.1  61  1  10
7  1.176  11.4  53  1  13
9  1.724  8.2  55  0  12
11  1.114  14.0  61  0  10
11  1.230  12.0  43  0  9
11  1.301  13.2  65  0  10
11  1.508  7.5  70  0  12
11  1.079  9.6  51  1  9
13  0.778  5.5  60  1  10
14  1.398  14.6  66  0  10
15  1.602  10.6  70  0  11
16  1.342  9.0  48  0  10
16  1.322  8.8  62  1  10
17  1.230  10.0  53  0  9
17  1.591  11.2  68  0  10
18  1.447  7.5  65  1  8
19  1.079  14.4  51  0  15
19  1.255  7.5  60  1  9
24  1.301  14.6  56  1  9
25  1.000  12.4  67  0  10
<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>1.230</td>
<td>11.2</td>
<td>49</td>
<td>1</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1.322</td>
<td>10.6</td>
<td>46</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>1.114</td>
<td>7.0</td>
<td>48</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>1.602</td>
<td>11.0</td>
<td>63</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1.000</td>
<td>10.2</td>
<td>69</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>1.146</td>
<td>5.0</td>
<td>70</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>1.568</td>
<td>7.7</td>
<td>74</td>
<td>0</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>1.000</td>
<td>10.1</td>
<td>60</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>1.255</td>
<td>9.0</td>
<td>49</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>1.204</td>
<td>12.1</td>
<td>42</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>1.447</td>
<td>6.6</td>
<td>59</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>1.322</td>
<td>12.8</td>
<td>52</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>1.176</td>
<td>10.6</td>
<td>47</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>1.322</td>
<td>14.0</td>
<td>63</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>1.431</td>
<td>11.0</td>
<td>58</td>
<td>1</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.945</td>
<td>10.2</td>
<td>59</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.924</td>
<td>10.0</td>
<td>49</td>
<td>1</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.114</td>
<td>12.4</td>
<td>48</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.532</td>
<td>10.2</td>
<td>81</td>
<td>0</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.079</td>
<td>9.9</td>
<td>57</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.146</td>
<td>11.6</td>
<td>46</td>
<td>1</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.613</td>
<td>14.0</td>
<td>60</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.398</td>
<td>8.8</td>
<td>66</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1.663</td>
<td>4.9</td>
<td>71</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1.146</td>
<td>13.0</td>
<td>55</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1.322</td>
<td>13.0</td>
<td>59</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1.322</td>
<td>10.8</td>
<td>69</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>1.230</td>
<td>7.3</td>
<td>82</td>
<td>1</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>1.756</td>
<td>12.8</td>
<td>72</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>1.114</td>
<td>12.0</td>
<td>66</td>
<td>0</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>1.255</td>
<td>12.5</td>
<td>66</td>
<td>0</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>1.079</td>
<td>14.0</td>
<td>60</td>
<td>0</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
function rlglik(t)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
Implicit none
! log-likelihood function for three parameter
! Kloek and van Dijk prob
Real(kind=2), dimension(d) :: t
Real(kind=2), dimension(2,2) :: bd
Real(kind=2) :: s11, s12, s22
! don't need individual data
! omegn is sse matrix of y’s (c i) on x’s (z lagi)
Real(kind=2), dimension(d) :: &
omegn=(/ .89754012354d0, .23746932393d0, .073946591778d0 /)
! zpb1z is cov mx of (z lagi)*df or centered x’x
Real(kind=2), dimension(d) :: &
zpb1z=(/ 1.4176836d0, .1507912d0, .5799304d0 /)
! pi2 is bottom part of reduced form estimates
Real(kind=2), dimension(2,2) :: pi2= &
& reshape((/ .424716d0, 2.757635d0, &
.049935d0, 1.034737d0 /), (/2,2/))
Real(kind=2) :: rlglik
! check parameter constraints
rlglik = 1.d9
if( (t(1) .le. 0.d0) .or. (t(1) .ge. 1.d0) ) return
if( (t(2) .le. 0.d0) .or. (t(2) .ge. 1.d0) ) return
if( (t(3) .le. 0.d0) .or. (t(3) .ge. 1.d0) ) return
if( t(1) + t(2) .ge. 1.d0 ) return
! gamma matrix is ( t(1)-1 t(2) )
! ( t(1) t(2)-1 )
!
! put n*gamma' omega*gamma into ss
!
s11 = (t(1)-1.d0)*(t(1)-1.d0)*omegn(1) + &
& 2.d0*(t(1)-1.d0)*t(1)*omegn(2) + &
& t(1)*t(1)*omegn(3)
s12 = (t(1)-1.d0)*t(2)*omegn(1) + ( t(1)*t(2) + &
& (t(1)-1.d0)*(t(2)-1.d0) )*omegn(2) + &
\[ t(1) \times (t(2) - 1.0) \times \text{omegn}(3) \]
\[
s22 = (t(2) - 1.0) \times (t(2) - 1.0) \times \text{omegn}(3) + \&
\& 2.0 \times (t(2) - 1.0) \times t(2) \times \text{omegn}(2) + \&
\& t(2) \times t(2) \times \text{omegn}(1) \]

! now difference in beta from estimate
! get beta estimate from reduced form \( \pi_2 \) and gamma
\[
\text{bd}(1,1) = t(1) - \pi_2(1,1) + t(1) \times (\pi_2(1,1) + \pi_2(1,2))
\]
\[
\text{bd}(2,1) = 0 - \pi_2(2,1) + t(1) \times (\pi_2(2,1) + \pi_2(2,2))
\]
\[
\text{bd}(1,2) = t(2) - \pi_2(1,2) + t(2) \times (\pi_2(1,1) + \pi_2(1,2))
\]
\[
\text{bd}(2,2) = t(3) - \pi_2(2,2) + t(2) \times (\pi_2(2,1) + \pi_2(2,2))
\]

! now get ss -- add \( n \times \text{gam} \times \text{omeg} \times \text{gam} \) to qf in bd and \( z'z \)
\[
s11 = s11 + \text{bd}(1,1) \times \text{bd}(1,1) \times \text{zpb1z}(1) + \&
\& 2 \times \text{bd}(1,1) \times \text{bd}(2,1) \times \text{zpb1z}(2) + \&
\& + \text{bd}(2,1) \times \text{bd}(2,1) \times \text{zpb1z}(3)
\]
\[
s12 = s12 + \text{bd}(1,1) \times \text{bd}(1,2) \times \text{zpb1z}(1) + \&
\& \text{bd}(1,1) \times \text{bd}(2,2) \times \text{zpb1z}(2) + \&
\& \text{bd}(2,1) \times \text{bd}(1,2) \times \text{zpb1z}(2) + \&
\& \text{bd}(2,2) \times \text{bd}(2,1) \times \text{zpb1z}(3)
\]
\[
s22 = s22 + \text{bd}(1,2) \times \text{bd}(1,2) \times \text{zpb1z}(1) + \&
\& 2 \times \text{bd}(1,2) \times \text{bd}(2,2) \times \text{zpb1z}(2) + \&
\& + \text{bd}(2,2) \times \text{bd}(2,2) \times \text{zpb1z}(3)
\]
\[
\text{rlglik} = 10.0 \times \text{dlog}(1.0 - t(1) - t(2)) + \&
\& - 4.5 \times \text{dlog}(s11 \times s22 - s12 \times s12)
\]

! use unlikelihood so that hessian is pos def
\[ \text{rlglik} = - \text{rlglik} \]
\[ \text{return} \]
\[ \text{end} \]
A.4 Fortran code for Example 4

function rlglik(t)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
use the_data
! log-likelihood function for five parameter Kloek and van Dijk prob
Use the_data
Implicit none
Real(kind=2), dimension(d), intent(in) :: t
Real(kind=2) :: rlglik, u1,u2,s11,s12,s22
integer :: i
! check parameter constraints
rlglik = 1.d9
if( (t(3) .le. 0.d0) .or. (t(3) .ge. 1.d0) ) return
if( (t(4) .le. 0.d0) .or. (t(4) .ge. 1.d0) ) return
if( (t(5) .le. 0.d0) .or. (t(5) .ge. 1.d0) ) return
if( t(3) + t(4) .ge. 1.d0 ) return
!
s11 = 0.d0
s12 = 0.d0
s22 = 0.d0
! note index starts at 2 because of lag investment
do i = 2,11
! form residuals
u1 = yciz(i,2) - t(1) - t(3)*yciz(i,1)
u2 = yciz(i,3) - t(2) - t(4)*yciz(i,1) - t(5)*yciz(i-1,3)
! now get ss
s11 = s11 + u1*u1
s12 = s12 + u1*u2
s22 = s22 + u2*u2
End do !loop on i
rlglik = 10.d0*dlog(1.-t(3)-t(4)) &
& - 5.d0*dlog(s11*s22-s12*s12)
! use unlikelihood so that hessian is pos def
rlglik = - rlglik
return
end
Data

do i = 1,11
    read(5,23) (yciz(i,j),j=1,4)
    yciz(i,1) = yciz(i,1)/1000.d0
    yciz(i,2) = yciz(i,2)/1000.d0
    yciz(i,3) = yciz(i,3)/1000.d0
    yciz(i,4) = yciz(i,4)/1000.d0
end do

yciz:
1948  13895  10706  1024   2165
1949  14377  10940  1078   2359
1950  14843  11250  1123   2470
1951  15307  11089  1052   3166
1952  15360  11023  980    3357
1953  15951  11474  1073   3404
1954  16680  12023  1281   3376
1955  17237  12443  1474   3320
1956  17547  12548  1591   3408
1957  17788  12802  1668   3318
1958  17699  13096  1709   2894
A.5 Fortran code for Example 5

```fortran
function rlglik(t)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
! log-likelihood function for
! eight parameter Kloek and van Dijk prob
use_the_data
Implicit none
Real(kind=2), intent(IN), dimension(d) :: t
Real(kind=2) :: u1,u2,s11,s12,s22,dlgsig,dlggam
Integer :: i
real(kind=2) :: rlglik
! restrictions on beta and sigma
rlglik = 1.d9
if( (t(3) .le. 0.d0) .or. (t(3) .ge. 1.d0) ) return
if( (t(4) .le. 0.d0) .or. (t(4) .ge. 1.d0) ) return
if( (t(5) .le. 0.d0) .or. (t(5) .ge. 1.d0) ) return
if( t(6) .le. 0.d0 ) return
if( t(8) .le. 0.d0 ) return
if( t(3) + t(4) .ge. 1.d0 ) return
!
s11 = 0.d0
s12 = 0.d0
s22 = 0.d0
! note index starts at 2 because of lag investment
do i = 2,11
! form residuals
u1 = yciz(i,2) - t(1) - t(3)*yciz(i,1)
u2 = yciz(i,3) - t(2) - t(4)*yciz(i,1) - t(5)*yciz(i-1,3)
! multiply by inv of Cholesky factor
u1 = u1 / t(6)
u2 = ( u2 - t(7)*u1 ) / t(8)
!
now get ss
s11 = s11 + u1*u1
s12 = s12 + u1*u2
s22 = s22 + u2*u2
End do !loop on i
!
End function rlglik
```

log of determinants
dlgsig = dlog(t(6)*t(8))
dlggam = dlog( 1.d0 - t(3) - t(4) )
rlglik = 10.d0*dlggam - 13.d0*dlgsig - (s11+s22)/2.d0

! use unlikelihood so that hessian is pos def
rlglik = - rlglik
return
End

Data

do i = 1,11
    read(5,23) (yciz(i,j),j=1,4)
    yciz(i,1) = yciz(i,1)/1000.d0
    yciz(i,2) = yciz(i,2)/1000.d0
    yciz(i,3) = yciz(i,3)/1000.d0
    yciz(i,4) = yciz(i,4)/1000.d0
end do
<table>
<thead>
<tr>
<th>Year</th>
<th>Value1</th>
<th>Value2</th>
<th>Value3</th>
<th>Value4</th>
<th>Value5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1948</td>
<td>13895</td>
<td>10706</td>
<td>1024</td>
<td>2165</td>
<td></td>
</tr>
<tr>
<td>1949</td>
<td>14377</td>
<td>10940</td>
<td>1078</td>
<td>2359</td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>14843</td>
<td>11250</td>
<td>1123</td>
<td>2470</td>
<td></td>
</tr>
<tr>
<td>1951</td>
<td>15307</td>
<td>11089</td>
<td>1052</td>
<td>3166</td>
<td></td>
</tr>
<tr>
<td>1952</td>
<td>15360</td>
<td>11023</td>
<td>980</td>
<td>3357</td>
<td></td>
</tr>
<tr>
<td>1953</td>
<td>15951</td>
<td>11474</td>
<td>1073</td>
<td>3404</td>
<td></td>
</tr>
<tr>
<td>1954</td>
<td>16680</td>
<td>12023</td>
<td>1281</td>
<td>3376</td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td>17237</td>
<td>12443</td>
<td>1474</td>
<td>3320</td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>17547</td>
<td>12548</td>
<td>1591</td>
<td>3408</td>
<td></td>
</tr>
<tr>
<td>1957</td>
<td>17788</td>
<td>12802</td>
<td>1668</td>
<td>3318</td>
<td></td>
</tr>
<tr>
<td>1958</td>
<td>17699</td>
<td>13096</td>
<td>1709</td>
<td>2894</td>
<td></td>
</tr>
</tbody>
</table>
A.6 Fortran code for Examples 6 and 7

```fortran
double precision function rlglik(t)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
! log-likelihood function for the Birthweight problem
  Use the_data
    Implicit None
    Real(kind=2) :: s,e
    Real(kind=2), dimension(d), intent(IN) :: t
  !
    Integer :: I, j
  !
    rlglik = 0.d0
      ! use only one obs in five for now (Example 6)
      ! to use all the data (Example 7), change to:
      do i = 1,189
        do i = 1,189,5
          e = t(1)
          do j = 1,6
            e = e + t(j+1)*g(i,j)
          End do !loop on j
          rlglik = rlglik + h(i)*e - dlog( 1.d0 + dexp(e) )
        End do !loop on i
      ! use unlikelihood so that hessian is pos def
      rlglik = - rlglik
      return
  end
  !
  Data:
    do i = 1,189
      read(8,23) h(i),(g(i,j),j=1,6)
      g(i,4) = min0( 1, g(i,4) )
    End do !loop on i
  h(i), g(i,j):
    0 19 182 2 0 0 0 1 0 2523
    0 33 155 3 0 0 0 0 3 2551
```
<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20</td>
<td>105</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2557</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
<td>108</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2594</td>
</tr>
<tr>
<td>0</td>
<td>18</td>
<td>107</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2600</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
<td>124</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2622</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>118</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2637</td>
</tr>
<tr>
<td>0</td>
<td>17</td>
<td>103</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2637</td>
</tr>
<tr>
<td>0</td>
<td>29</td>
<td>123</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2663</td>
</tr>
<tr>
<td>0</td>
<td>26</td>
<td>113</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2665</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>95</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2722</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>150</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2733</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>95</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2751</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>107</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2750</td>
</tr>
<tr>
<td>0</td>
<td>18</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2769</td>
</tr>
<tr>
<td>0</td>
<td>18</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2769</td>
</tr>
<tr>
<td>0</td>
<td>15</td>
<td>98</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2778</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>118</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2782</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>120</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2807</td>
</tr>
<tr>
<td>0</td>
<td>28</td>
<td>120</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2821</td>
</tr>
<tr>
<td>0</td>
<td>32</td>
<td>121</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2835</td>
</tr>
<tr>
<td>0</td>
<td>31</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2835</td>
</tr>
<tr>
<td>0</td>
<td>36</td>
<td>202</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2836</td>
</tr>
<tr>
<td>0</td>
<td>28</td>
<td>120</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2863</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>120</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2877</td>
</tr>
<tr>
<td>0</td>
<td>28</td>
<td>167</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2877</td>
</tr>
<tr>
<td>0</td>
<td>17</td>
<td>122</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2906</td>
</tr>
<tr>
<td>0</td>
<td>29</td>
<td>150</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2920</td>
</tr>
<tr>
<td>0</td>
<td>26</td>
<td>168</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2920</td>
</tr>
<tr>
<td>0</td>
<td>17</td>
<td>113</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2920</td>
</tr>
<tr>
<td>0</td>
<td>17</td>
<td>113</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2920</td>
</tr>
<tr>
<td>0</td>
<td>24</td>
<td>90</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2948</td>
</tr>
<tr>
<td>0</td>
<td>35</td>
<td>121</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2948</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>155</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2977</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>125</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2977</td>
</tr>
<tr>
<td>0</td>
<td>29</td>
<td>140</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2977</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>138</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2977</td>
</tr>
<tr>
<td>0</td>
<td>27</td>
<td>124</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2922</td>
</tr>
<tr>
<td>0</td>
<td>31</td>
<td>215</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3005</td>
</tr>
<tr>
<td>0</td>
<td>33</td>
<td>109</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3033</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
<td>185</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3042</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
<td>189</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>3062</td>
</tr>
<tr>
<td>0</td>
<td>23</td>
<td>130</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3062</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
<td>160</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3062</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>90</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>90</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3062</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>132</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>132</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>115</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>85</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>120</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>128</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>130</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>95</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>115</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>110</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>110</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>153</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>103</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>119</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>119</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>119</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>110</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3232</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>140</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3234</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>133</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3260</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>169</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3274</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>115</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3274</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>250</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>3303</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>141</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3317</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>158</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3317</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>112</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3317</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>150</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3321</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>115</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3331</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>112</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>135</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3374</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>229</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3402</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>140</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3416</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>134</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>3430</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>121</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3444</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>190</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3459</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>131</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3460</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>170</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3473</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>110</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3544</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>127</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3487</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>123</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3544</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>120</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3572</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>187</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2367</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>122</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2381</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>105</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2381</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>115</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2381</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>120</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2410</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>142</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2410</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>22</td>
<td>130</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2410</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>120</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2414</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>110</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2424</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>120</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2438</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>154</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2442</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>105</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2450</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>26</td>
<td>190</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2466</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>101</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2466</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>95</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2466</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>100</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2495</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>94</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2495</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>142</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2495</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21</td>
<td>130</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2495</td>
<td></td>
</tr>
</tbody>
</table>
A.7 Fortran code for Example 8

function rlglik(t)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
! log-likelihood function for
! the second Evans & Swartz problem
!
implicit none
real(kind=2) :: theta,omth, e1,e2,s, pij, rlglik
real(kind=2), intent(in), dimension(9) :: t
real(kind=2), dimension(3,3) :: p
real(kind=2), dimension(2,3) :: alph, beta
integer :: i, j
integer, dimension(3,3):: nij= &
reshape((/43, 6, 9, 16, 11, 18, 3, 10, 16/), (/3,3/))
!
  e1 = dexp(t(1))
s = 2.d0*( 1.d0 + e1 )
theta = e1 / s
omth = (2.d0+e1) / s
e1 = dexp(t(2))
e2 = dexp(t(3))
s = 1.d0 + e1 + e2
alph(1,1) = e1 / s
alph(1,2) = e2 / s
alph(1,3) = 1.d0 / s
e1 = dexp(t(4))
e2 = dexp(t(5))
s = 1.d0 + e1 + e2
alph(2,1) = e1 / s
alph(2,2) = e2 / s
alph(2,3) = 1.d0 / s
e1 = dexp(t(6))
e2 = dexp(t(7))
s = 1.d0 + e1 + e2
beta(1,1) = e1 / s
beta(1,2) = e2 / s
beta(1,3) = 1.d0 / s
e1 = dexp(t(8))
e2 = dexp(t(9))
s = 1.d0 + e1 + e2
beta(2,1) = e1 / s
beta(2,2) = e2 / s
beta(2,3) = 1.d0 / s

rlglik = 0.d0
do i = 1,3
do j = 1,3
   pij = theta*alph(1,i)*beta(1,j) + omth*alph(2,i)*beta(2,j)
   rlglik = rlglik + nij(i,j)*dlog(pij)
end do !loop on j
! determinant of Jacobian
   rlglik = rlglik + &
      & dlog( alph(1,i)*alph(2,i)*beta(1,i)*beta(2,i) )
end do !loop on i
rlglik = rlglik + dlog( theta*(omth-theta) )
! use unlikenlihood so that hessian is pos def
rlglik = - rlglik
return
end
A.8 Fortran code for Example 9

function rlglik(t)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
! log-likelihood function for the Evans & Swartz problem
use the_data
real(kind=2):: rlglik
real(kind=2), dimension(d) :: t
real(kind=2) :: s,e
!
integer :: i
!
!
s = dexp(t(10))
!
rlglik = 0.d0
do i = 1,9
  do j = 1,5
    e = ( g(i,j) - t(i) ) / s
    rlglik = rlglik - dlog( 1.d0 + e*e )
  end do !loop on j
end do !loop on i
rlglik = 2.d0*rlglik - 45.d0*t(10) - 20.32122173d0
! use unlikelihood so that hessian is pos def
rlglik = - rlglik
return
end

Data:

do i = 1,9
  read(8,23) (g(i,j),j=1,5)
end do !loop on i

0.151754965 1.085632419 0.591451482 -0.068423422 -0.378073973
-0.638321684 0.186636475 -0.009514517 0.495339037 0.261643198
2.833250710 1.098441287 -0.476050335 -0.383252862 1.128895153
0.041946195 0.174042986 -0.581710481 0.447193674 -0.071116204
A.9  Fortran code for Example 10

FUNCTION rlglik(TH)
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
! log-likelihood function for ordinal regression model
!
use the_data
implicit none
Real(kind=2), dimension(d) :: xi
Real(kind=2), dimension(d) :: th
REAL(kind=2), external :: DLNPRB
INTEGER :: KM1,KW,i,j, ki
Real(kind=2) :: rlglik
!
! check condition that cutoffs are increasing
rlglik = 1000d0
km1 = km - 1
kw = 0
do j = 2,km1
  if( th(j) .lt. th(j-1) ) kw = kw + 1
end do !loop on j
if( kw .ne. 0 ) return
!
rlglik = 0.D0
!
DO I = 1,N
KI = KL(I)
IF( P .EQ. 0 ) GO TO 3
DO J = 1,P
XI(J) = g(I,J)
End do !loop on j
3 CONTINUE
rlglik = rlglik - DLNPRB(KI,KM,TH,XI,P)
end do !loop on i
RETURN
END

FUNCTION DLNPRB(K,KM,ALPH,X,P)
! EVALUATE PROBABILITY FUNCTION WITH LOGISTIC ORDINAL MODEL
! original code written by
! J F MONAHAN DEPT OF STATISTICS, NCSU, RALEIGH, NC USA
! converted from Fortran 77 to Fortran 95 by
! BUFFY HUDSON-CURTIS (2000) DEPT OF STATISTICS,
! NCSU, RALEIGH, NC USA
!
!
! K INTEGER OBSERVED CATEGORY
! KM INTEGER NUMBER OF CATEGORIES
! ALPH REAL PARAMETERS
! ALPH(1),...,ALPH(KM-1) CUT POINTS
! ALPH(KM),...,ALPH(KM+P-1) COEFFICIENTS
! X REAL COVARIATES
! P INTEGER NUMBER OF COVARIATES
!
!
Implicit none

Real(kind=2) :: dlnprb
INTEGER :: I,K,KM,P

Real(kind=2), dimension(1) :: alph, x
REAL(kind=2) :: BX,XA,EB,EXA,EXB,AB,EMD,OMED,DEL

! GET REGRESSION PART B’X
BX = 0.D0
IF( P .EQ. 0 ) GO TO 3
DO I =1,P
BX = BX + ALPH(KM-1+I)*X(I)
End do !loop on i
!
3 CONTINUE
!
EVALUATE PROBABILITY DEPENDING ON CATEGORY
IF( K .NE. 1 ) GO TO 4
!
CASE K = 1
BX = ALPH(1) - BX
EB = 0.D0
AB = DABS(BX)
IF( AB .LT. 125.D0 ) EB = DEXP(-AB)
DLNPRB = - DLOG( 1.D0 + EB )
IF( BX .LT. 0.D0 ) DLNPRB = DLNPRB + BX
RETURN
!
NEXT LOOK AT THE CASE K = KM
CONTINUE
IF( K .NE. KM ) GO TO 7
BX = ALPH(KM-1) - BX
EB = 0.D0
AB = DABS(BX)
IF( AB .LT. 125.D0 ) EB = DEXP(-AB)
DLNPRB = - DLOG( 1.D0 + EB )
IF( BX .GT. 0.D0 ) DLNPRB = DLNPRB - BX
RETURN

! NOW THE MIDDLE VALUES OF K

CONTINUE
XA = ALPH(K-1) - BX
DEL = ALPH(K) - ALPH(K-1)
EXA = 0.D0
AB = DABS(XA)
IF( AB .LT. 125.D0 ) EXA = DEXP(-AB)

! NOW SPLIT CASES DEPENDING ON SIZE OF DEL
IF( DEL .GT. .01D0 ) GO TO 8
OMED = DEL*(1-(DEL/2)*(1-(DEL/3)*(1-(DEL/4)*(1-DEL/5))))
EMD = 1.D0 - OMED
OMED = OMED / (1.D0+EXA)
IF( XA .GT. 0.D0 ) OMED = OMED / (1.D0 + EXA*EMD)
IF( XA .LE. 0.D0 ) OMED = OMED / (EXA + EMD)
DLNPRB = -AB + DLOG(OMED)
RETURN

! NOW DO MIDDLE SIZE VALUES OF DEL

CONTINUE
IF( DEL .GT. 40.D0 ) GO TO 9
EMD = DEXP( -DEL )
IF( XA .GT. 0.D0 ) DLNPRB = -AB + DLOG( (1.D0-EMD) / &
    ( (1.D0 + EXA) * (1.D0 + EXA*EMD) ) )
& IF( XA .LE. 0.D0 ) DLNPRB = -AB + DLOG( (1.D0-EMD) / &
    ( (1.D0 + EXA) * (EXA + EMD) ) )
RETURN

! NOW BIG VALUES OF DEL

CONTINUE
BX = ALPH(K) - BX
EXB = 0.D0
AB = DABS(BX)
IF( AB .LT. 125.D0 ) EXB = DEXP(-AB)
DLNPRB = - DLOG( (1.D0 + EXA) * (1.D0 + EXB) )
IF( XA .GT. 0.D0 ) DLNPRB = DLNPRB - XA
IF( BX .LT. 0.D0 ) DLNPRB = DLNPRB + BX
RETURN
END

! read in the data

DO I = 1,N
READ(8,22) IU,IM,KI,IV,SRAWI,GRJ
! WRITE(6,22) IU,IM,KI,IV,SRAWI,GRJ
! IU umax
! IM mda
! KI max intensity on Fujita scale
! IV surface vorticity
! SRAW storm relative ambient wind
! GRJ outbreak group number
! add one to intensity so range is 1 to 7
KI = KI + 1
!
! store it (it’s our Y) in KL
KL(I) = KI
!
! count the number in each intensity category
KCAT(KI) = KCAT(KI) + 1
!
!
g(I,1) = real(IU)
g(I,2) = real(IM)
g(I,3) = SRAWI
!
g(I,4) = g(I,1)*g(I,1)/100.d0
!
!
End do !loop on i

data (Variables 1, 2, and 7 not read in)

OKA 0.00 10 -4 0 18 1399 19.2 1
OKB 0.00 15 15 0 18 3506 11.1 1
OKC 2.50 20 15 3 18 3506 19.2 1
OKD 0.00 10 11 0 18 1514 14.8 1
OKE 0.00 13 14 0 18 1514 18.4 1
OKF 0.00 18 11 0 18 1514 15.1 1
OKG 0.00 8 -5 0 18 1399 23.7 1
OKH 0.00 5 5 0 18 1247 16.2 1
<table>
<thead>
<tr>
<th>Code</th>
<th>Value</th>
<th>Column1</th>
<th>Column2</th>
<th>Column3</th>
<th>Column4</th>
<th>Column5</th>
<th>Column6</th>
<th>Column7</th>
<th>Column8</th>
</tr>
</thead>
<tbody>
<tr>
<td>OKI</td>
<td>0.00</td>
<td>4</td>
<td>-4</td>
<td>0</td>
<td>18</td>
<td>1247</td>
<td>22.6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>OKJ</td>
<td>1.25</td>
<td>18</td>
<td>13</td>
<td>2</td>
<td>18</td>
<td>1247</td>
<td>12.3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>OKK</td>
<td>0.00</td>
<td>12</td>
<td>17</td>
<td>0</td>
<td>18</td>
<td>1514</td>
<td>12.9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>OKL</td>
<td>0.00</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>18</td>
<td>1399</td>
<td>22.4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>OKM</td>
<td>0.00</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>18</td>
<td>1514</td>
<td>18.6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>OKN</td>
<td>0.00</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>1514</td>
<td>15.0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>OKO</td>
<td>0.00</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>18</td>
<td>1514</td>
<td>17.7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>OKP</td>
<td>2.00</td>
<td>15</td>
<td>21</td>
<td>2</td>
<td>18</td>
<td>397</td>
<td>3.7</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WIA</td>
<td>0.00</td>
<td>15</td>
<td>16</td>
<td>0</td>
<td>19</td>
<td>2070</td>
<td>9.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIB</td>
<td>0.00</td>
<td>12</td>
<td>16</td>
<td>0</td>
<td>19</td>
<td>2070</td>
<td>11.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIC</td>
<td>0.00</td>
<td>10</td>
<td>15</td>
<td>0</td>
<td>19</td>
<td>4121</td>
<td>15.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WID</td>
<td>0.00</td>
<td>13</td>
<td>10</td>
<td>0</td>
<td>19</td>
<td>2070</td>
<td>16.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WID'</td>
<td>1.00</td>
<td>15</td>
<td>30</td>
<td>1</td>
<td>19</td>
<td>2070</td>
<td>13.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIE</td>
<td>0.00</td>
<td>16</td>
<td>22</td>
<td>0</td>
<td>19</td>
<td>4121</td>
<td>11.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIE'</td>
<td>3.00</td>
<td>33</td>
<td>32</td>
<td>4</td>
<td>19</td>
<td>4121</td>
<td>15.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIF</td>
<td>0.00</td>
<td>30</td>
<td>7</td>
<td>0</td>
<td>19</td>
<td>4121</td>
<td>8.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIF'</td>
<td>3.00</td>
<td>27</td>
<td>35</td>
<td>5</td>
<td>19</td>
<td>4121</td>
<td>7.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIG</td>
<td>0.00</td>
<td>20</td>
<td>16</td>
<td>0</td>
<td>19</td>
<td>4121</td>
<td>12.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIH</td>
<td>0.00</td>
<td>15</td>
<td>22</td>
<td>0</td>
<td>19</td>
<td>4121</td>
<td>10.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIJ</td>
<td>2.00</td>
<td>16</td>
<td>34</td>
<td>2</td>
<td>19</td>
<td>4121</td>
<td>8.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIN</td>
<td>5.00</td>
<td>30</td>
<td>51</td>
<td>5</td>
<td>19</td>
<td>4121</td>
<td>9.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIQ</td>
<td>0.00</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>19</td>
<td>1788</td>
<td>14.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIS</td>
<td>0.00</td>
<td>8</td>
<td>10</td>
<td>0</td>
<td>19</td>
<td>1788</td>
<td>17.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WIT</td>
<td>0.00</td>
<td>12</td>
<td>21</td>
<td>0</td>
<td>19</td>
<td>1788</td>
<td>8.0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>IAA</td>
<td>3.00</td>
<td>22</td>
<td>11</td>
<td>3</td>
<td>32</td>
<td>1071</td>
<td>20.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAB</td>
<td>4.00</td>
<td>25</td>
<td>13</td>
<td>4</td>
<td>32</td>
<td>1071</td>
<td>21.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAC</td>
<td>3.25</td>
<td>27</td>
<td>15</td>
<td>5</td>
<td>32</td>
<td>1071</td>
<td>21.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAD</td>
<td>2.61</td>
<td>50</td>
<td>15</td>
<td>4</td>
<td>32</td>
<td>2517</td>
<td>19.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAE</td>
<td>0.00</td>
<td>12</td>
<td>8</td>
<td>0</td>
<td>32</td>
<td>1071</td>
<td>18.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAE'</td>
<td>0.00</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>1071</td>
<td>21.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAG</td>
<td>0.00</td>
<td>12</td>
<td>5</td>
<td>0</td>
<td>32</td>
<td>1823</td>
<td>14.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAH</td>
<td>0.00</td>
<td>12</td>
<td>8</td>
<td>0</td>
<td>32</td>
<td>2517</td>
<td>17.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAI</td>
<td>0.00</td>
<td>15</td>
<td>-1</td>
<td>0</td>
<td>32</td>
<td>1720</td>
<td>16.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAJ</td>
<td>2.67</td>
<td>24</td>
<td>13</td>
<td>3</td>
<td>32</td>
<td>1071</td>
<td>28.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAJ'</td>
<td>2.50</td>
<td>20</td>
<td>12</td>
<td>3</td>
<td>32</td>
<td>1071</td>
<td>31.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IAL</td>
<td>0.00</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>32</td>
<td>1350</td>
<td>3.0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>NEA</td>
<td>4.00</td>
<td>15</td>
<td>21</td>
<td>4</td>
<td>54</td>
<td>2130</td>
<td>15.7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEB</td>
<td>0.00</td>
<td>7</td>
<td>-9</td>
<td>0</td>
<td>54</td>
<td>2130</td>
<td>10.5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEC</td>
<td>0.00</td>
<td>5</td>
<td>-3</td>
<td>0</td>
<td>54</td>
<td>1400</td>
<td>5.7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NED</td>
<td>0.00</td>
<td>5</td>
<td>13</td>
<td>0</td>
<td>54</td>
<td>2018</td>
<td>9.7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEE</td>
<td>0.00</td>
<td>10</td>
<td>-13</td>
<td>0</td>
<td>54</td>
<td>2018</td>
<td>8.0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEF</td>
<td>3.00</td>
<td>20</td>
<td>12</td>
<td>4</td>
<td>54</td>
<td>2130</td>
<td>13.8</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Value1</td>
<td>Value2</td>
<td>Value3</td>
<td>Value4</td>
<td>Value5</td>
<td>Value6</td>
<td>Value7</td>
<td>Value8</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>NEG</td>
<td>4.33</td>
<td>30</td>
<td>18</td>
<td>5</td>
<td>54</td>
<td>2130</td>
<td>16.1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEH</td>
<td>0.00</td>
<td>10</td>
<td>-3</td>
<td>0</td>
<td>54</td>
<td>2130</td>
<td>11.6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEI</td>
<td>0.00</td>
<td>2</td>
<td>-2</td>
<td>0</td>
<td>54</td>
<td>2018</td>
<td>10.5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEJ</td>
<td>0.00</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>54</td>
<td>2130</td>
<td>11.5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEK</td>
<td>0.00</td>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>54</td>
<td>2130</td>
<td>11.3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEL</td>
<td>0.00</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>54</td>
<td>2018</td>
<td>10.7</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>NEM</td>
<td>2.00</td>
<td>20</td>
<td>9</td>
<td>2</td>
<td>54</td>
<td>2018</td>
<td>7.2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>OHA</td>
<td>0.00</td>
<td>15</td>
<td>17</td>
<td>0</td>
<td>9</td>
<td>479</td>
<td>26.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHBC</td>
<td>4.00</td>
<td>38</td>
<td>38</td>
<td>5</td>
<td>9</td>
<td>479</td>
<td>23.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHD</td>
<td>0.00</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>479</td>
<td>15.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHE</td>
<td>0.00</td>
<td>10</td>
<td>-5</td>
<td>0</td>
<td>9</td>
<td>479</td>
<td>11.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHF</td>
<td>0.00</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>479</td>
<td>9.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHG</td>
<td>4.00</td>
<td>25</td>
<td>36</td>
<td>5</td>
<td>9</td>
<td>479</td>
<td>17.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHH</td>
<td>3.33</td>
<td>25</td>
<td>54</td>
<td>6</td>
<td>9</td>
<td>479</td>
<td>13.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHI</td>
<td>2.67</td>
<td>26</td>
<td>29</td>
<td>4</td>
<td>9</td>
<td>479</td>
<td>14.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHJ2</td>
<td>3.00</td>
<td>31</td>
<td>16</td>
<td>4</td>
<td>9</td>
<td>1328</td>
<td>21.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHK</td>
<td>2.00</td>
<td>20</td>
<td>25</td>
<td>2</td>
<td>9</td>
<td>1328</td>
<td>22.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHL</td>
<td>0.00</td>
<td>10</td>
<td>14</td>
<td>0</td>
<td>9</td>
<td>1328</td>
<td>18.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHM</td>
<td>2.00</td>
<td>20</td>
<td>29</td>
<td>2</td>
<td>9</td>
<td>1328</td>
<td>21.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHO</td>
<td>0.00</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>479</td>
<td>17.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHR</td>
<td>0.00</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>9</td>
<td>1328</td>
<td>10.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHU</td>
<td>0.00</td>
<td>10</td>
<td>-7</td>
<td>0</td>
<td>9</td>
<td>479</td>
<td>9.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHV</td>
<td>0.00</td>
<td>15</td>
<td>18</td>
<td>0</td>
<td>9</td>
<td>1410</td>
<td>13.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHW</td>
<td>0.00</td>
<td>19</td>
<td>15</td>
<td>0</td>
<td>9</td>
<td>1410</td>
<td>14.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHX</td>
<td>0.00</td>
<td>10</td>
<td>22</td>
<td>0</td>
<td>9</td>
<td>1410</td>
<td>4.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHY</td>
<td>0.00</td>
<td>2</td>
<td>18</td>
<td>0</td>
<td>9</td>
<td>1410</td>
<td>14.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHZ</td>
<td>0.00</td>
<td>15</td>
<td>25</td>
<td>0</td>
<td>9</td>
<td>1973</td>
<td>8.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>OHAA</td>
<td>0.00</td>
<td>22</td>
<td>11</td>
<td>0</td>
<td>9</td>
<td>1973</td>
<td>20.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>SDA</td>
<td>0.00</td>
<td>8</td>
<td>-8</td>
<td>0</td>
<td>53</td>
<td>1877</td>
<td>16.2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDB</td>
<td>0.00</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>53</td>
<td>811</td>
<td>1.9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDC</td>
<td>0.00</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>53</td>
<td>811</td>
<td>15.1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDD</td>
<td>0.00</td>
<td>9</td>
<td>6</td>
<td>0</td>
<td>53</td>
<td>811</td>
<td>10.5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDE</td>
<td>0.00</td>
<td>13</td>
<td>17</td>
<td>0</td>
<td>53</td>
<td>811</td>
<td>12.7</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDF</td>
<td>0.00</td>
<td>5</td>
<td>12</td>
<td>0</td>
<td>53</td>
<td>1877</td>
<td>20.1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDG</td>
<td>0.00</td>
<td>12</td>
<td>8</td>
<td>0</td>
<td>53</td>
<td>1877</td>
<td>24.6</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDH</td>
<td>2.40</td>
<td>20</td>
<td>23</td>
<td>3</td>
<td>53</td>
<td>3046</td>
<td>14.1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDI</td>
<td>2.00</td>
<td>10</td>
<td>24</td>
<td>3</td>
<td>53</td>
<td>3046</td>
<td>14.7</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDJ</td>
<td>0.00</td>
<td>10</td>
<td>-6</td>
<td>0</td>
<td>53</td>
<td>1877</td>
<td>18.0</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDK</td>
<td>0.00</td>
<td>10</td>
<td>-8</td>
<td>0</td>
<td>53</td>
<td>1877</td>
<td>16.9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>SDL</td>
<td>2.20</td>
<td>18</td>
<td>23</td>
<td>5</td>
<td>53</td>
<td>3046</td>
<td>25.2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>KSA</td>
<td>0.00</td>
<td>7</td>
<td>13</td>
<td>0</td>
<td>48</td>
<td>2670</td>
<td>1.3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Value</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
<td>M</td>
<td>Width</td>
<td>Height</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSB</td>
<td>0.00</td>
<td>7</td>
<td>-3</td>
<td>0</td>
<td>48</td>
<td>2670</td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSC</td>
<td>0.00</td>
<td>15</td>
<td>-6</td>
<td>0</td>
<td>48</td>
<td>302</td>
<td>14.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSD</td>
<td>2.00</td>
<td>23</td>
<td>15</td>
<td>2</td>
<td>48</td>
<td>302</td>
<td>24.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSE</td>
<td>0.00</td>
<td>8</td>
<td>18</td>
<td>0</td>
<td>48</td>
<td>2670</td>
<td>6.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSG</td>
<td>0.00</td>
<td>5</td>
<td>13</td>
<td>0</td>
<td>48</td>
<td>2119</td>
<td>3.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSH</td>
<td>0.00</td>
<td>20</td>
<td>-1</td>
<td>0</td>
<td>48</td>
<td>1638</td>
<td>8.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSI</td>
<td>0.00</td>
<td>15</td>
<td>-3</td>
<td>0</td>
<td>48</td>
<td>302</td>
<td>21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSJ</td>
<td>0.00</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>48</td>
<td>302</td>
<td>22.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSK</td>
<td>0.00</td>
<td>8</td>
<td>-7</td>
<td>0</td>
<td>48</td>
<td>747</td>
<td>10.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSL</td>
<td>0.00</td>
<td>15</td>
<td>-8</td>
<td>0</td>
<td>48</td>
<td>2119</td>
<td>12.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSM</td>
<td>4.00</td>
<td>17</td>
<td>10</td>
<td>4</td>
<td>48</td>
<td>1376</td>
<td>16.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSN</td>
<td>0.00</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>48</td>
<td>1376</td>
<td>15.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSO</td>
<td>3.00</td>
<td>15</td>
<td>24</td>
<td>3</td>
<td>48</td>
<td>302</td>
<td>18.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSP</td>
<td>0.00</td>
<td>10</td>
<td>-4</td>
<td>0</td>
<td>48</td>
<td>1376</td>
<td>16.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KSQ</td>
<td>3.00</td>
<td>16</td>
<td>11</td>
<td>3</td>
<td>48</td>
<td>1376</td>
<td>18.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXA</td>
<td>2.91</td>
<td>20</td>
<td>16</td>
<td>4</td>
<td>24</td>
<td>2253</td>
<td>21.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXB</td>
<td>2.00</td>
<td>15</td>
<td>17</td>
<td>2</td>
<td>24</td>
<td>2253</td>
<td>20.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXC</td>
<td>0.00</td>
<td>13</td>
<td>10</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>15.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXD</td>
<td>0.00</td>
<td>2</td>
<td>18</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>14.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXE</td>
<td>0.00</td>
<td>15</td>
<td>11</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>20.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXF</td>
<td>0.00</td>
<td>10</td>
<td>18</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>20.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXG</td>
<td>0.00</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>16.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXH</td>
<td>2.33</td>
<td>30</td>
<td>21</td>
<td>4</td>
<td>24</td>
<td>2253</td>
<td>16.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXI</td>
<td>0.00</td>
<td>15</td>
<td>4</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>20.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXJ</td>
<td>1.50</td>
<td>18</td>
<td>14</td>
<td>2</td>
<td>24</td>
<td>536</td>
<td>23.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXK</td>
<td>0.00</td>
<td>13</td>
<td>9</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>11.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXL</td>
<td>0.00</td>
<td>17</td>
<td>13</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>13.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXM</td>
<td>0.00</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>6.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXN</td>
<td>0.00</td>
<td>10</td>
<td>7</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>17.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXO</td>
<td>0.00</td>
<td>15</td>
<td>14</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>10.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TXP</td>
<td>0.00</td>
<td>13</td>
<td>7</td>
<td>0</td>
<td>24</td>
<td>2253</td>
<td>22.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCA</td>
<td>3.33</td>
<td>10</td>
<td>42</td>
<td>5</td>
<td>26</td>
<td>445</td>
<td>23.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCB</td>
<td>1.67</td>
<td>15</td>
<td>13</td>
<td>2</td>
<td>26</td>
<td>728</td>
<td>33.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCC</td>
<td>0.00</td>
<td>5</td>
<td>11</td>
<td>0</td>
<td>26</td>
<td>445</td>
<td>14.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCD</td>
<td>0.00</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>26</td>
<td>445</td>
<td>23.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCE</td>
<td>0.00</td>
<td>5</td>
<td>-5</td>
<td>0</td>
<td>26</td>
<td>926</td>
<td>21.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCF</td>
<td>0.00</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>26</td>
<td>547</td>
<td>22.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCG</td>
<td>0.00</td>
<td>5</td>
<td>-6</td>
<td>0</td>
<td>26</td>
<td>445</td>
<td>14.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCH</td>
<td>2.00</td>
<td>6</td>
<td>15</td>
<td>2</td>
<td>26</td>
<td>445</td>
<td>20.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCN</td>
<td>0.00</td>
<td>10</td>
<td>-4</td>
<td>0</td>
<td>26</td>
<td>445</td>
<td>25.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCJ</td>
<td>0.00</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>26</td>
<td>547</td>
<td>18.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCK</td>
<td>0.00</td>
<td>10</td>
<td>4</td>
<td>0</td>
<td>26</td>
<td>728</td>
<td>15.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Count</td>
<td>Day</td>
<td>Time</td>
<td>GT</td>
<td>Type</td>
<td>Value</td>
<td>Code</td>
<td>Count</td>
<td>Day</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>-----</td>
<td>------</td>
<td>----</td>
<td>------</td>
<td>-------</td>
<td>------</td>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>NCL</td>
<td>0.00</td>
<td>10</td>
<td>0</td>
<td>26</td>
<td>728</td>
<td>19.5</td>
<td>9</td>
<td>NCM</td>
<td>0.00</td>
</tr>
<tr>
<td>ALA</td>
<td>0.00</td>
<td>10</td>
<td>08</td>
<td>0</td>
<td>45</td>
<td>2433</td>
<td>10.2</td>
<td>10</td>
<td>ALB</td>
</tr>
<tr>
<td>ALC</td>
<td>0.00</td>
<td>03</td>
<td>11</td>
<td>0</td>
<td>45</td>
<td>1174</td>
<td>10.9</td>
<td>10</td>
<td>ALD</td>
</tr>
<tr>
<td>ALE</td>
<td>0.00</td>
<td>15</td>
<td>07</td>
<td>0</td>
<td>45</td>
<td>2433</td>
<td>19.7</td>
<td>10</td>
<td>ALF</td>
</tr>
<tr>
<td>ALG</td>
<td>0.00</td>
<td>08</td>
<td>02</td>
<td>0</td>
<td>45</td>
<td>2104</td>
<td>15.9</td>
<td>10</td>
<td>ALH</td>
</tr>
<tr>
<td>ALI</td>
<td>0.00</td>
<td>05</td>
<td>11</td>
<td>0</td>
<td>45</td>
<td>2433</td>
<td>16.0</td>
<td>10</td>
<td>ALJ</td>
</tr>
<tr>
<td>ALL</td>
<td>0.00</td>
<td>08</td>
<td>28</td>
<td>0</td>
<td>45</td>
<td>1229</td>
<td>16.2</td>
<td>10</td>
<td>ALN</td>
</tr>
<tr>
<td>ALO</td>
<td>3.00</td>
<td>15</td>
<td>14</td>
<td>5</td>
<td>45</td>
<td>1229</td>
<td>16.7</td>
<td>10</td>
<td>ALP</td>
</tr>
<tr>
<td>ALQ</td>
<td>0.00</td>
<td>05</td>
<td>-1</td>
<td>0</td>
<td>45</td>
<td>1229</td>
<td>20.5</td>
<td>10</td>
<td>ALR</td>
</tr>
<tr>
<td>ALS</td>
<td>2.00</td>
<td>13</td>
<td>18</td>
<td>2</td>
<td>45</td>
<td>1628</td>
<td>18.8</td>
<td>10</td>
<td>ALT</td>
</tr>
<tr>
<td>ALU</td>
<td>1.00</td>
<td>10</td>
<td>14</td>
<td>1</td>
<td>45</td>
<td>1229</td>
<td>22.5</td>
<td>10</td>
<td>ALV</td>
</tr>
<tr>
<td>ALW</td>
<td>3.00</td>
<td>23</td>
<td>09</td>
<td>3</td>
<td>45</td>
<td>1628</td>
<td>17.0</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>
Bibliography


Monahan, J.F. (1989), Evaluating the Smirnov distribution function, *Institute of*


