



Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

**THE INVERTED BIVARIATE AND MULTIVARIATE LOSS
FUNCTIONS: PROPERTIES AND APPLICATIONS**

TSOI, SZE LEONG

M.Phil

The Hong Kong Polytechnic University

2013

The Hong Kong Polytechnic University

Department of Applied Mathematics

**The Inverted Bivariate and Multivariate Loss Functions:
Properties and Applications**

Tsoi, Sze Leong

A thesis submitted in partial fulfillment of the requirements for the degree of Master
of Philosophy

December 2011

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in text.

Tsoi, Sze Leong

ABSTRACT

The purpose of this research project is to develop a loss function by inverting the Dirichlet distribution. The motive is to explore the possibility of developing loss functions of higher dimension. Spiring (1993), Spiring and Yeung (1998) and Spiring and Leung (2002) developed a series of Inverted Probability Loss Functions (IPLFs), while most of them are univariate. A multivariate IPLF, which in theory, should have the capability of depicting financial loss when more than one quality characteristics are concerned. The Dirichlet distribution is chosen because of its additiveness of random variables in nature and its ease to include any number of variables.

To depict the long term financial loss of manufacturing a product, it is desired to calculate the expected loss by $E[L] = \int_{\Omega} L(x)g(x)dx$. $L(x)$ is the loss function and $g(x)$ is the conjugate function, which describes the measurement of the quality characteristic which the manufacturer is interested in. This research project focuses on developing a multivariate inverted probability loss function and some discussion will be made about the conjugate function. The prime objective of this research is to develop an inverted multivariate probability loss function. For this purpose, extending the conjugate function into higher dimension will also be discussed.

Although it is highly desired that real data can be collected to form the conjugate function, sometimes it is not feasible to do so and simulation may be the solution to this situation. Three common random variate generation methods, namely the

conditional approach, acceptance and rejection approach and gamma factors approach, are used to invert the Dirichlet distribution. The major distinction between the three aforementioned approaches lays in the method in obtaining the estimators, respectively, the method of moments, maximum likelihood estimation and a modified percentile matching method. The theory and the procedure underlying the methods in obtaining estimators are discussed in details and the sets of estimates obtained are then evaluated by the Chi-square goodness-of-fit test to compare the acceptability of the three estimation techniques. Cells arrangement is also discussed and verified by Peacock's two-dimensional Kolmogorov-Smirnov goodness-of-fit test, which is known for its conservativeness. Properties and various probable conjugate distributions are examined to compare the univariate IPLF with the multivariate IPLF. At last, the bivariate inverted normal loss function, introduced by Spiring (1993) and the inverted Dirichlet loss function are applied to data collected in solder paste composition and dimensions of fasteners, as an application of the proposed technique in depicting the financial loss due to deviation in production, and ultimately, to show the feasibility of such bivariate loss functions.

ACKNOWLEDGEMENT

This thesis represents the work done in the past four years in the area of loss function. Through this journey, I have received lots of help and kind advice from many different friends and colleagues. I would like to thank my supervisor, Dr. P.K. Leung, for his guidance, patience and support. Not only did he teach me the essence of statistics, he also taught me how to become a better person. I am truly grateful for all his help in this period of tough time. Dr. Joseph Lee and Dr. Chi Kin Chan also provided extremely useful advice and suggestion for this piece of work and their kind help are very much appreciated.

Aitken Enterprises, a local trading company, has lent me a generous helping hand with samples of fasteners from a Germany manufacturer, which they act in agent for. Mr. Daniel Sy, Mr. Pogy Tsoi, Mr. Andy Chan and Mr. Yiu have provided me valuable insights in manufacturing and quality control, and certainly their precise products, which I gathered information from to demonstrate the feasibility of the proposed multivariate inverted probability loss functions.

My parents and my wife gave me unlimited support, not only during my study, but throughout my life journey. I am not the most reasonable person, and they accept me and embrace all my shortcomings with patience and love. They give me all the strength I need to survive through harsh time and frustration. I cannot show enough of my gratitude. I love you all.

TABLE OF CONTENTS

Certificate of Originality	i
Abstract	ii
Acknowledgement	iv
Table of Contents	v
List of Tables	ix
List of Figures	xi
CHAPTER 1 Introduction	1
1.1 Background	1
1.2 Objectives	3
1.3 Outline of Thesis	5
CHAPTER 2 Literature Review	8
2.1 Goalpost Philosophy	9
2.2 Taguchi's Philosophy	13
2.3 Inverted Probability Loss Function (IPLF)	19
2.4 Modified Reflected Normal Loss Function	22
2.5 Revised Inverted Normal Loss Function	25
2.6 IPLF base on other distributions	26
2.7 Properties of the Inverted Probability Loss Functions	35
2.8 The Bivariate Inverted Normal Loss Function	40

CHAPTER 3	Dirichlet Distribution	44
3.1	Setup and General Characteristics of the Dirichlet Distribution.....	45
3.2	Conditional Density.....	49
3.3	Random Number Generation.....	50
3.3.1	Conditional Approach.....	51
3.3.2	Acceptance and Rejection Approach.....	56
3.3.3	Gamma Factors Approach.....	61
3.4	Parameter Estimations.....	63
3.4.1	Method of Moments Estimation.....	64
3.4.2	Maximum Likelihood Estimation.....	68
3.4.3	Percentile Matching Method.....	71
3.5	Biasedness of the Estimation Methods.....	75
3.5.1	Method of Moments Estimation.....	76
3.5.2	Maximum Likelihood Estimation.....	76
3.5.3	Percentile Matching Method.....	77
3.6	Goodness-of-Fit Test.....	78
3.7	Comparison among different methods and approaches.....	84
3.7.1	Discussion on the Two Tests.....	94
3.7.2	Discussion on the Three Methods of Estimation.....	97
3.8	Discussion and Suggestion on Simulating Techniques and Estimating Methods.....	98
CHAPTER 4	Inverted Dirichlet Loss Function.....	100

4.1	Spiring's Inverted bivariate Normal Loss Function	101
4.1.1	Procedure of Building a Bivariate INLF	102
4.1.2	Relationship of a Bivariate INLF and its Marginal Loss Functions ..	113
4.2	Inverted Dirichlet Loss Function.....	117
4.2.1	Bivariate Inverted Dirichlet Loss Function.....	120
4.3	Conjugate Functions of the Inverted Dirichlet Loss Function	122
4.3.1	Uniform Distribution.....	124
4.3.2	Inverted Dirichlet Distribution.....	125
4.3.3	Bivariate Gamma Distribution	127
4.3.4	Bivariate Exponential Distribution	130
4.3.5	Bivariate Normal Distribution.....	132
4.3.6	Section Summary	135
4.4	Properties of Multivariate Inverted Probability Loss Functions	136
4.4.1	Properties of the Loss Inversion Ratio (LIR).....	136
4.5	Fitting an Inverted Probability Loss Function.....	143
CHAPTER 5	Applications.....	145
5.1	Wave Soldering Process	146
5.1.1	Maximum Likelihood Estimators	149
5.1.2	Percentile Matching Method	149
5.1.3	Chi-Square Goodness-of-Fit Test	151
5.1.4	Expected Loss	156

5.2	Fasteners	163
5.2.1	Self-tapping Threaded Insert.....	163
5.2.2	Pressed-in Threaded Inserts	177
5.3	Choosing of Inverted Probability Loss Functions and Conjugate Functions	
	199	
CHAPTER 6	Conclusion.....	201
Appendix	212

LIST OF TABLES

Table 3-1 Computational time for the three simulation methods	63
Table 3-2 The first two moments of $\text{Beta}(\theta_1, \theta_2, \theta_3)$	65
Table 3-3 The first two sample moments of empirical data.....	65
Table 3-4 Table of the first two moments based on 10000 generated random vectors	66
Table 3-5 Observed versus expected frequency for simulated data.....	83
Table 3-6 Tests results of random variates simulated by conditional approach with [2 3 4]	85
Table 3-7 Tests results of random variates simulated by conditional approach with [8 6 4]	86
Table 3-8 Tests results of random variates simulated by conditional approach with [1.5 1.5 1.5].....	87
Table 3-9 Tests results of random variates simulated by gamma factors approach with [2 3 4].....	88
Table 3-10 Tests results of random variates simulated by gamma factors approach with [8 6 4].....	89
Table 3-11 Tests results of random variates simulated by gamma factors approach with [1.5 1.5 1.5].....	90
Table 3-12 Tests results of random variates simulated by acceptance and rejection approach with [2 3 4]	91
Table 3-13 Tests results of random variates simulated by acceptance and rejection approach with [8 6 4]	92

Table 3-14 Tests results of random variates simulated by acceptance and rejection approach with [1.5 1.5 1.5]	93
Table 3-15 Relationship between maximum difference and p -value for the 2D K-S test	95
Table 5-1 Collected data of solder paste	147
Table 5-2 Observed versus expected frequency for the solder paste example with maximum likelihood estimates	153
Table 5-3 Observed versus expected frequency for the solder paste example with estimators obtained by the modified percentile matching method.....	154
Table 5-4 Table of estimators of the conjugate function and the IbNLF.....	160
Table 5-5 Collected data of the self-tapping threaded inserts in mm	165
Table 5-6 Observed versus expected frequency with the two estimation methods .	169
Table 5-7 Calculation of Chi-square goodness-of-fit test statistic.....	172
Table 5-8 Estimates of the conjugate function and IbNLF	176
Table 5-9 Collected data on the pressed-in threaded inserts in mm	178
Table 5-10 Observed versus expected frequency for the three estimation methods	183
Table 5-11 Calculation for Peacock's 2D K-S test	186
Table 5-12 Expected loss depicted by the three estimation methods.....	187
Table 5-13 Estimators for the conjugate function and the IDLF	191
Table 5-14 Table for the estimators obtained by MLE.....	192
Table 5-15 Expected loss depicted with three different scenarios.....	193
Table 5-16 Estimators of the transformed lengths obtained by method of moments and MLE	195
Table 5-17 Expected loss depicted by MME and MLE for the two individual quality characteristic	197

LIST OF FIGURES

Figure 2-1 Goalpost theorem when limits are 1 and 2	10
Figure 2-2 Taguchi's quadratic loss function (T=2.5).....	14
Figure 2-3 Modified quadratic loss function for asymmetric tolerances (m=2)	17
Figure 2-4 Taguchi's loss function and Spiring's INLF	20
Figure 2-5 Spiring's piecewise reflected normal loss function	22
Figure 2-6 Various Spiring's INLF	24
Figure 2-7 An example of Pan's loss function	26
Figure 2-8 Various Inverted Gamma Loss Functions	29
Figure 2-9 Illustration of IGLF's tails	29
Figure 2-10 Inverted Tukey Loss Function ($\lambda = 1.5$ (above), $\lambda = 2.5$ (below)).....	31
Figure 2-11 Various Inverted Beta Loss Functions	34
Figure 2-12 Illustration of two IBLF with different scales	39
Figure 2-13 A bivariate INLF	43
Figure 3-1 Scatter plot of 10000 sample points generated by the conditional approach	54
Figure 3-2 Plot of Beta(2,7) pdf.....	55
Figure 3-3 Plot of the conditional distribution, Beta(3,4).....	55
Figure 3-4 Illustration of acceptance and rejection approach	58
Figure 3-5 Scatter plot of data generated by the acceptance and rejection approach	60
Figure 3-6 Scatter plot of data generated by gamma factors	62
Figure 3-7 Illustration of the setup of the three regions.....	72
Figure 3-8 Illustration of integrating a triangular region with Riemann sums concept	82

Figure 3-9 Illustration of cells dissection.....	83
Figure 4-1 $g(\sigma_x, \sigma_y)$ with $UL_x=10, UL_y=2$	107
Figure 4-2 IbNLF with $T_x = T_y = 0, \sigma_x = \sigma_y = 1, \rho = 0$	109
Figure 4-3 IbNLF with $T_x = T_y = 0, \sigma_x = 2, \sigma_y = 3, \rho = 0$	110
Figure 4-4 IbNLF with $T_x = T_y = 0, \sigma_x = \sigma_y = 2, \rho = 0.5$	110
Figure 4-5 IbNLF with $T_x = T_y = 0, \sigma_x = \sigma_y = 2, \rho = 0.25$	111
Figure 4-6 Contour plot of two different IbNLFs (above: $\rho = 0.5$, below: $\rho = 0.9$)	112
Figure 4-7 Surface of an IbDLF with a conjugate distribution of Inverted Dirichlet	127
Figure 4-8 Surface of McKay's bivariate gamma[2, 3, 4].....	128
Figure 4-9 Surface of an IbDLF with a conjugate distribution of gamma[2, 3, 4]..	129
Figure 4-10 Plot of McKay's bivariate gamma distribution.....	130
Figure 4-11 Surface of an IbDLF with a conjugate distribution of Exp[1]	131
Figure 4-12 Plot of a bivariate Normal Distribution.....	133
Figure 4-13 Surface of an IbDLF with a conjugate distribution of Normal[.5, .5, 1, 1]	134
Figure 5-1 Scatter plot of solder paste data.....	148
Figure 5-2 Illustration of regions used in the modified percentile matching method	150
Figure 5-3 Illustration of cells dissection in the Chi-square goodness-of-fit test	152
Figure 5-4 Conjugate function based on the maximum likelihood method.....	155
Figure 5-5 Conjugate function based on the modified percentile matching method	155
Figure 5-6 Three-dimensional plot of $g(s_x, s_y)$ in solder paste example.....	157
Figure 5-7 Contour plot of $g(s_x, s_y)$ in solder paste example.....	158

Figure 5-8 Plot of IbNLF when correlation is assumed to be zero	159
Figure 5-9 Plot of IbNLF when correlation is assumed to be -0.95.....	160
Figure 5-10 Plot of volume of IbNLF versus correlation	161
Figure 5-11 Plot of E[L] versus correlation	162
Figure 5-12 Design of the Self-tapping Threaded Insert	164
Figure 5-13 Scatter plot of Self-tapping Threaded Inserts data in mm.....	166
Figure 5-14 Conjugate function of self-tapping threaded insert by mle	167
Figure 5-15 Conjugate function of self-tapping threaded insert by percentile matching method.....	168
Figure 5-16 Illustration of cells dissection in the Self-tapping Threaded Inserts example	169
Figure 5-17 Illustration of cells distribution when number of cells are assumed to be 8.....	171
Figure 5-18 Design diagram of the pressed-in threaded insert	177
Figure 5-19 Scatter plot of the original measurements	179
Figure 5-20 Scatter plot of length of head vs length of body.....	180
Figure 5-21 Scatter plot of transformed data collected from pressed-in threaded inserts	182
Figure 5-22 Illustration of cell dissections for the Chi-square goodness-of-fit test.	183
Figure 5-23 Conjugate function built by method of moments estimates.....	184
Figure 5-24 Conjugate function built by mle estimates	185
Figure 5-25 Conjugate function built by percentile matching estimates	185
Figure 5-26 Expected loss vs. values of parameters	188
Figure 5-27 Relationship between the estimator and loss.....	190
Figure 5-28 Plot of IDLF with respect to the two transformed measurements.....	191

Figure 5-29 Plot of loss versus values of estimates when correlation is assumed to be 0.....	194
Figure 5-30 The pdf with the estimates obtained.....	194
Figure 5-31 Relationship between loss and values of estimators	196

CHAPTER 1 INTRODUCTION

1.1 Background

In a flattening world, most industries nowadays are facing fierce competition from rivals around the world. Quality control and re-engineering are now crucial to survival. A product with poor quality leads to consequences like short life cycle, customers complaints, goods return, law suit due to malfunction, damage to company's reputation, losses to both manufacturer, buyers, and as well to the society.

A product with excellent quality, on the other hand, would help in increasing market share, reducing scrapping, reworking and repairing works, leading to higher production efficiency and eventually higher profit, leaving the company high flexibility in pricing and increasing its competitiveness. Moreover, a product with excellent quality would theoretically have a longer life cycle, which would in turn conserve social resources, for example in the case of public infrastructures. Thus, producing a product right at the first time becomes an important issue in manufacturing.

In quality control literature for example, Crosby (1979), quality is related to the loss caused during the life cycle of a product. A high quality product will have a minimal loss to the society as it goes through its life cycle. This loss being considered is not just the loss imposed to the manufacturer, but also to the customers. For manufacturers, a poor quality product will lead to higher warranty cost in replacement, return and repairing; quality check will have to be performed more frequently and demand additional resources. On the customer's side, excessive energy may be needed for the product, inconvenience will be caused when the product is imperfect, and pollution including heat and noise could also arise. Consider a customer purchasing a heater with poor quality; this heater may use excessive energy, producing less heat and noise as a by-product. In the most unfortunate scenario, it may explode.

Loss functions are used to measure the financial loss caused by a product. To be specific, costs arise when the measurements of a product deviate from the optimal values, and loss functions depict the monetary costs according to the deviation.

Hazewinkel, M. (1988) defined loss function as: in a problem of a statistical decision making, a non-negative function indicating the loss to an experimenter

given a particular state of the world and a particular decision. In manufacturing, given the measurement of a product, financial loss is indicated by the loss function.

Loss functions provide useful information about the overall quality of a batch of products. With sampling, an overall financial loss can be estimated. The decision on whether machines or performance of a particular factory needs to be upgraded can be based on the estimated loss indicated by the loss function.

1.2 Objectives

Through the development of loss functions, economic loss relating to several parameters can be developed. The objectives of this research project are listed as follows:

1. Develop a loss function based on the inversion of a bivariate probability density functions (IbPLF);
2. Derive the properties linked with such the loss function in objective 1;
3. Develop a loss function based on the inversion of multivariate probability density functions (ImPLF);

4. Examine the relationship between loss functions based on inverting univariate and bivariate probability density functions – the dependence and independence cases;
5. Demonstrate applications of various IbPLFs.

This research will develop loss functions, an IbPLF and an ImPLF, based on the inversion of a Dirichlet distribution, where, one of its forms is the commonly known as the Beta distribution. Intuitively, more parameters should be able to give a multi-dimensional picture on how deviations in different aspects would cause the financial losses to grow; although, complexity of the function and its calculations will increase substantially.

Spiring (1993) introduced a bivariate version of his inverted normal loss function (INLF), which is the only bivariate IPLF developed so far. In this research project, Spiring's IbNLF will be revisited and discussed. In addition, an IPLF based on the Dirichlet distribution will be developed. The Dirichlet distribution has a domain of $x_1 + x_2 + \dots + x_k \leq 1$. In other words, k quality characteristics can be included into the depiction of financial loss during the manufacturing process of a certain product. Each quality characteristic is related to other quality characteristics and the sum of

all quality characteristics must be smaller than one. This may be useful when a manufacturer is keen to find out the long run financial loss of his product when the quality characteristics are additive in nature.

1.3 Outline of Thesis

Chapter 2 consists of extensive literature review in loss functions built upon various theorems: including the traditional loss function, which is commonly known as the goalpost theorem, Taguchi's quadratic loss function, Spiring's inverted normal loss function, and various inverted probability loss functions developed under the influence of Spiring. Advantages and constraints of these aforementioned loss functions were compared.

Since this research project focuses in developing an inverted probability loss function based on the Dirichlet distribution. Chapter 3 first starts with a revision on the fundamentals of the Dirichlet distribution. Ideally, raw data is most welcomed in constructing the risk function, which is also known as the conjugate. However, in case of absence of raw data, simulated data may be used. Therefore, various

simulation techniques will also discussed in chapter 3. Modeling or construction of the risk function is needed after raw data is collected or data is being simulated. Modeling requires a statistician to estimate the population parameters. Therefore, in this project, various estimation techniques have been used including the method of moments and the method of maximum likelihood estimation. A modified percentile matching method is also considered and discussed. Chi-square goodness-of-fit test is then used to justify the proposed distributions estimated by various estimators mentioned previously.

The bivariate inverted normal loss function, introduced by Spiring (1993), is revisited in Chapter 4. The major contribution of Chapter 4 lies in the development of an inverted Dirichlet loss function, the general form of multi-variates inverted Dirichlet loss function, as well as the properties of bivariate inverted probability loss functions. Issues about the conjugate function will also be discussed in Chapter 4.

Applications of the suggested loss functions are included in Chapter 5. Data of solder paste collected by students and samples of fasteners are used to develop loss functions and depict financial losses. Comparison of depicted financial losses is done in one of the sample with loss functions built by bivariate normal, bivariate

Dirichlet and univariate Dirichlet distributions. In Chapter 5, result shows that the bivariate Chi-square goodness-of-fit test can be applied in justifying a model with very vague rules on cells division. A bivariate Kolmogorov-Smirnov test will be used to determine whether these rules are appropriate or not. Finally, conclusion will be drawn in Chapter 6 and potential future extensions to this research will be discussed.

CHAPTER 2 LITERATURE REVIEW

In production nowadays, with fierce competition, quality control is becoming more and more important for a company to survive (Taguchi (1986), Grosby (1979)).

High quality products would not just attract more customers; the higher the quality and consistency, manufacturer will save more money and lead to a higher profit.

The concept to produce a high quality product right in the first time becomes an essential notion in manufacturing.

In quality control literature, quality is related to the loss caused by a product during its life cycle. A truly high quality product will have a minimal loss to society as it goes through this life cycle. Loss is imposed to both the manufacturer and the customers. For manufacturers, a poor quality product will involve higher cost in quality check as excessive man power is needed, and warranty cost, including inspecting, repairing or replacement. From the standpoint of customers, loss may include impairment in energy consumption, inconvenience when the product is not working, or even side effects like pollution, noise and heat.

Loss functions are used to measure financial loss caused by a product. In this report, theorems regarding quality control are reviewed and different loss functions are examined. At last, updated series of inverted probability loss functions, based on various kinds of probability distributions will be discussed. Statistical properties of this IbBLF will be explored. Real data is collected to demonstrate the feasibility of using such an IbBLF.

2.1 Goalpost Philosophy

The traditional view of loss, which is also known as the goalpost philosophy, is based on tolerancing. Tolerancing was based on the experience of designers regarding what has worked for them satisfactorily. Experienced designers have documented the tolerances that have worked well for different manufacturing processes, and assume these tolerances will be adequate and appropriate to be used for a new product if the production machinery has the same capability as those used in obtaining the standard tolerances.

Capability measures the variability about a desired target with which a machine or a process is capable of producing, and this capability is usually measured in terms of standard deviation, σ . In general, these tolerances are measured by the target T plus or minus 3σ . A product is assumed to be satisfactory if all its measurements lie inside this tolerance of $T \pm 3\sigma$. In Crosby (1979), he supports this view by claiming as long as a product is made according to the product engineer, within this permitted tolerance, this product is of good quality.

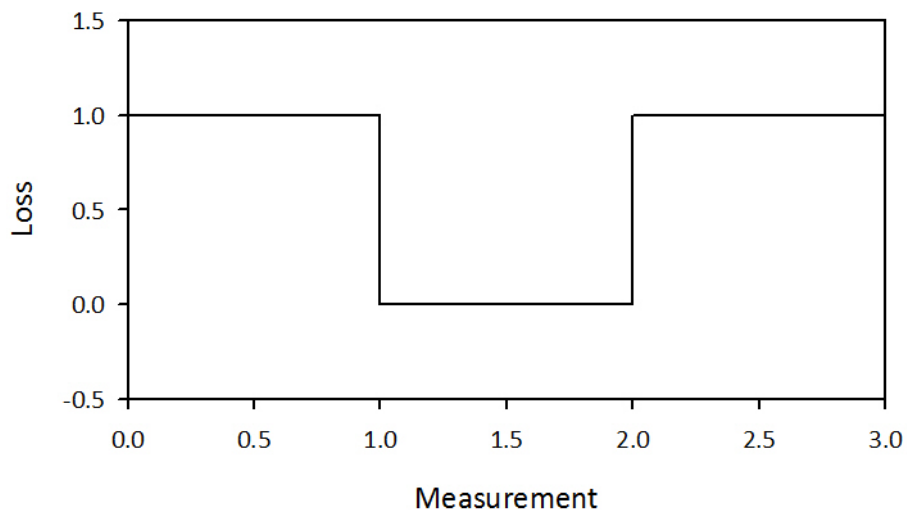


Figure 2-1 Goalpost theorem when limits are 1 and 2

Figure 2-1 above is an illustration of the loss function proposed by the goalpost theorem. It is very easy to understand and generally welcomed by manufacturers and product designers. This view is very strict, within a permitted tolerance interval designed by the product designer; the product is therefore satisfactory and not

satisfactory elsewhere. Graphically, it looks like a discrete function with jumps. Assume Δ_1 and Δ_2 are the lower and upper limit of the permitted tolerance, in this example, 1 and 2. For values outside this permitted tolerance, the loss will be the cost of rework or scratch since those products are not satisfactory and cannot be shipped and sold to the customers. In this case, there is only one possible amount of loss, and that is the cost of rework.

There are a few drawbacks about this goalpost philosophy. The most obvious is that, this goalpost philosophy is manufacturer-oriented. A product is classified as satisfactory just by examining whether it lies within the tolerance interval. It does not consider the financial loss which may be imposed to the customers at all.

Another criticism is that, this philosophy does not value the desired target. During product design, the target is supposed to be the measurement that would give the best quality over the product life cycle. The goalpost theorem concerns about whether the measurement is found within the acceptance interval. In other words, the manufacturer is indifferent whether the measurement is close to the target value or not, given it is found inside the tolerance limit. In the customers' point of view, this is another story. Customers want to purchase a product with the best quality

available, and products with small deviation yield a better quality than others. This is a point which is ignored by the goalpost philosophy.

One last criticism lies at the calculation of the tolerance interval $T \pm 3\sigma$. The standard deviation σ is estimated by the process standard deviation during the manufacturing process by the production machinery. In other words, this tolerance interval may vary when different machinery is used. In addition, it does not guarantee the quality of the product. If a production machine is not performing well, and it has been producing products with poor quality, under the calculation of the tolerance interval, many products with unsatisfactory quality will be accepted since the tolerance interval will be enlarged due to the larger standard deviation caused by the poor quality, or in other words, the large deviation apart from the desired nominal value. Base on the fact that the tolerance interval is calculated upon the performance of an individual machine, the decision on whether a machine should be replaced becomes inconclusive.

2.2 Taguchi's Philosophy

Dr. Genichi Taguchi (1986) believes that any deviation away from the desired target will lead to financial loss. His point of view is supported by the desire of customers whom purchase products which have high quality and consistency.

Ross (1996) pointed out, as a customer, the closer the amount of closing force is to the target value, the happier the customer is. If the product is a little off from the target, the customer would sense some loss. Even the product is classified as satisfactory in the goalpost theorem; the customer may still find the quality of the product to be poor, if the deviation is large. Thus, he suggested using the quadratic function to represent the actual loss function.

Taguchi defined his quadratic loss function as:

$$L(y) = B(y - T)^2, \quad (2.1)$$

where B is a proportionality constant and T is the desired target.

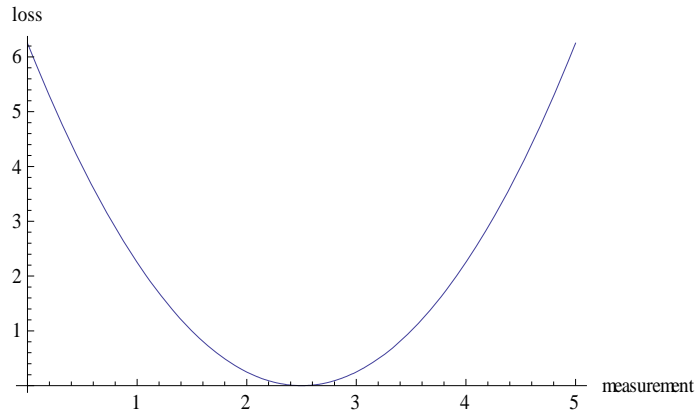


Figure 2-2 Taguchi's quadratic loss function ($T=2.5$)

The idea behind Taguchi's quadratic loss function is easily understood, as shown in Figure 2-2 above, as the measurement y gets farther away from the desired target, there will be a larger financial loss. If y is close to the nominal value, denoted by T , there will be little or no loss. B is used to scale how big or small the financial loss should be related to the square of variation.

Another admirable advantage of the quadratic loss function is that it does not depend on the production machineries. With the tradition loss functions, the tolerance interval is calculated as $T \pm 3\sigma$, while the standard deviation depends on how well or how poorly the machine works. In Taguchi's loss function, it does not matter; only the measurement between the product's value and the nominal value will of our concern. This will help in determining whether production machines need to be upgraded as the performance of each individual machine can be analyzed and

compared. A large expected financial loss indicates the poor performance of a certain machine and the manufacturer can use this information to decide whether or not this machine needs to be upgraded or replaced.

As pointed out by Leon and Wu (1992) and Spiring (1993), there are a few restrictions on the quadratic loss function. First of all, the quadratic loss function fails to provide a maximum possible loss. In other words, as the deviation gets bigger in either direction, the loss will be approaching infinity, while, in reality, the financial loss imposed by a product is usually finite, especially after product screening and unsatisfied products are very likely to be screened out and not shipped to the customers. With the help of quality check system, it is not likely that an unsatisfied product will be shipped and thus, the financial loss imposed by a satisfactory, but poor quality product should not be infinite, as it should have passed the basic safety requirement already.

In order to resolve the unreasonable infinite loss over the far ends of specified limits, Taguchi (1989) modified the quadratic loss function by truncating the quadratic loss function at the points where the function intersects the maximum loss. The general form of this modified quadratic loss function is

$$L(y) = \begin{cases} B(y-T)^2, & |y-T| \leq \sqrt{K/B} \\ K, & |y-T| > \sqrt{K/B} \end{cases}, \quad (2.2)$$

where K is the maximum loss in monetary terms. Although this modification gives a maximum loss for the quadratic loss function, the function itself is not smooth. It looks like the combination of the loss function based on the goalpost theory and Taguchi's loss function.

Another restriction the quadratic loss function has, lies in the fact that it is quadratic, making it hard to model complicated data by a simple function in the quadratic functional form. It is difficult to measure the financial loss when the behaviour of loss does not have a quadratic relationship with the measurement. For example, if the relationship between the loss and the deviation is linear, the quadratic loss function will not be accurate. It is obvious that, a simple quadratic function is difficult to model complicated data. It is also clear that, some amendments or innovations should be done to improve the quadratic loss function.

Leon and Wu (1992) suggested, for many products and manufacturing processes, performance can be conveniently measured in terms of a dispersion measurement because it is often easy to centre output around the target once dispersion has been reduced. They suggested when the loss function is non-quadratic, general dispersion,

location and off-target measures can to be used. Expected values and variances are used to build loss functions. In the loss function proposed by Leon and Wu (1992), the data no longer needs to be quadratic, and their suggestion is more flexible than the quadratic loss function developed by Taguchi (1989).

Taguchi (1989) also introduced loss function for asymmetric tolerances by setting two separate functions connecting at the nominal value, m . It is presented as follow:

$$L(y) = \begin{cases} \frac{B_1}{\Delta_1^2}(y-m)^2 & y \leq m \\ \frac{B_2}{\Delta_2^2}(y-m)^2 & y > m \end{cases} \quad (2.3)$$

Δ_1 and Δ_2 stand for the lower and upper tolerance limits respectively. B_1 and B_2 represent the proportionate coefficients for the two sides of the loss function.

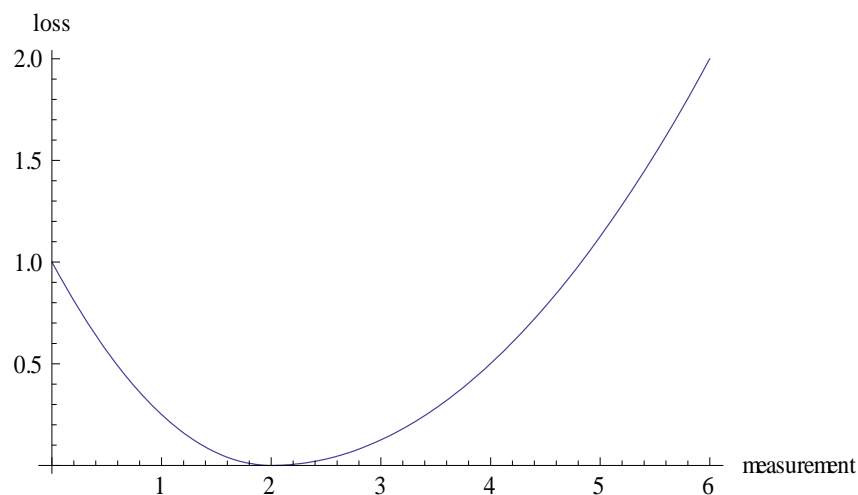


Figure 2-3 Modified quadratic loss function for asymmetric tolerances (m=2)

Figure 2-3 gives the graphical presentation of the asymmetric loss function introduced by Taguchi. This modification is more flexible than the original consideration; however, the problem about non-quadratic data remains unsolved.

In his book, Taguchi (1989) introduced The-Smaller-The-Better (S type) and The-Larger-The-Better (L type) tolerances, with the desired value to be $m=0$ and $m=+\infty$ respectively. An S type tolerance involves a nonnegative characteristic, whose ideal value is zero. The S type tolerance can be applied to wear, shrinkage, deterioration, and noise levels. In this type of tolerance, the upper limit is defined and the loss function will be small if the measurements are close to 0. In contrast, the L type tolerance describes characteristics such as the strength of material, fuel efficiency, and heat resistance for fire resistance material. The followings are the loss functions for these two types:

a. The-Smaller-The-Better type

$$L = \frac{B}{\Delta^2} \left(\frac{y_1^2 + y_2^2 + \dots + y_n^2}{n} \right) \quad (2.4)$$

where B is the proportionate coefficient and Δ is the tolerance.

b. The-Larger-The-Better type

$$L = \frac{B}{(1/\Delta)^2} \left(\frac{1}{y} \right)^2 \quad (2.5)$$

where B is the proportionate coefficient and Δ is the tolerance.

From the equations of the two loss functions, the financial losses are measured by the measurements of some particular characteristic. In equation (2.4) and (2.5), the financial loss is estimated by the actual measurement, instead of the deviation of the measurements away from the target value. All of these modifications give the quadratic loss function more flexibility and feasibility; however, its potential problems of an unbounded maximum loss and the possibility of non-quadratic data remain largely unresolved.

2.3 Inverted Probability Loss Function (IPLF)

Spiring (1993) introduced the idea of inverting a normal probability density function to build a loss function. The shape of inverting a normal density function looks almost the same as the quadratic function, except the tails of the reflected normal

loss function converge to some particular value, representing the maximum possible loss. In addition, with the normal distribution, there are two parameters, namely μ and σ . With two parameters, the model should be more flexible and data can be represented more accurately.

Spiring (1993) defined his reflected normal loss function as:

$$L(y) = K \left\{ 1 - \exp\left(-\frac{(y-T)^2}{2\gamma^2}\right) \right\} \quad (2.6)$$

where y represents the quality measurement, T the target value, γ a shape parameter (or standard deviation), and K the maximum-loss parameter.

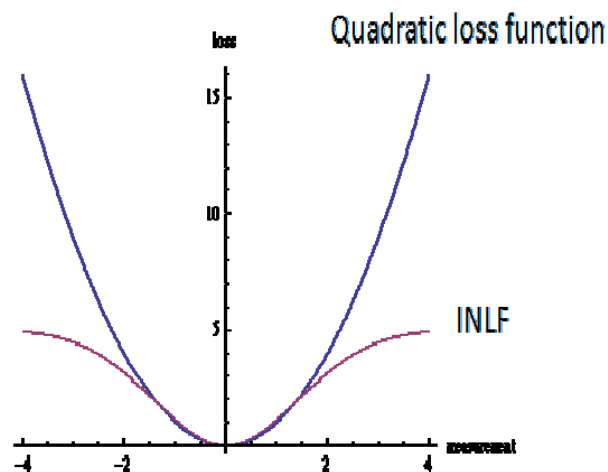


Figure 2-4 Taguchi's loss function and Spiring's INLF

In the Figure 2-4 above, both the quadratic loss function by Taguchi and the reflected normal loss function are illustrated. The blue line represents Taguchi's quadratic loss function and the red curve represents the INLF introduced by Siring. The above graph illustrated that, with the quadratic loss function, as the measurement deviates farther away from the target, the financial loss tends to infinity, while the reflected normal loss function depicts the financial loss approaching to the maximum loss. In addition, the reflected normal loss distribution performs better in the sense that, the magnitude of losses associated with extreme deviations from the target is smoother and tends to the maximum loss smoothly.

In comparison to Taguchi's modified quadratic loss functions, Siring also introduced an asymmetric loss function, to match the situation when the plus and minus tolerances are not equivalent. With the piecewise function, more parameters give more flexibility and freedom to the loss function. In his paper, he used the following example:

$$L(y) = \begin{cases} 5 \left\{ 1 - \exp\left(-\frac{y^2}{2}\right) \right\}, & y \leq 0 \\ 3 \left\{ 1 - \exp\left(-\frac{y^2}{8}\right) \right\}, & y > 0 \end{cases} \quad (2.7)$$

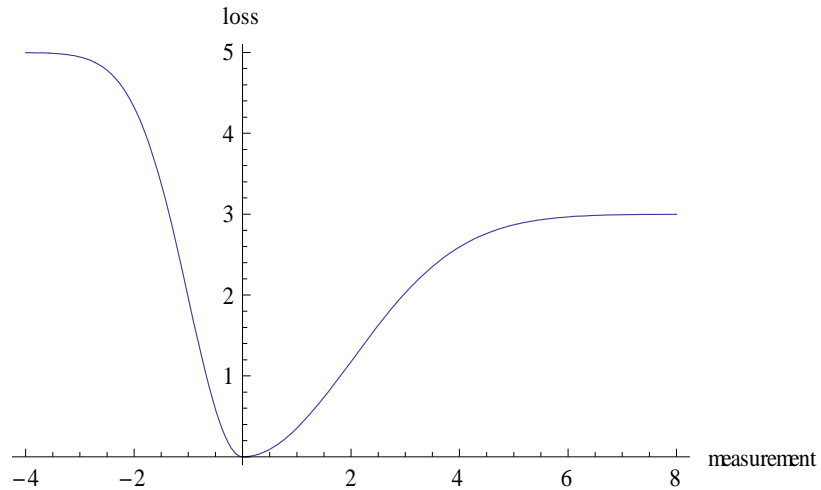


Figure 2-5 Spiring's piecewise reflected normal loss function

As illustrated in the example above (Figure 2-5), this piecewise reflected normal loss function is more satisfactory than the piecewise quadratic loss function. The shape of the function can be amended easily and so is the maximum loss.

2.4 Modified Reflected Normal Loss Function

Sun et al. (1996) proposed a modified reflected normal loss function enhancing the fundamental contributions of Taguchi and Spiring on quality loss function. They further pointed out that there are two major problems that have impeded the acceptance of loss functions as a fundamental means for expressing this important

element of quality costs, and they are: difficult to realistically access the loss to society corresponding to a specified deviation of the characteristic from the target, and the apparent absence of easy-to-use family of flexible loss functions that accurately represent the actual loss.

They introduced this modified reflected normal (MRN) loss function aiming to measure the loss to society more realistically. Instead of setting a maximum financial loss, they tried to measure the loss given a particular distance from the target, and use that measurement and distance to set up this MRN loss function.

In their paper, they defined the modified reflected normal loss function as:

$$L(y) = \frac{K_{\Delta}}{1 - \exp\left\{-1/2(\Delta/\gamma)^2\right\}} \left[1 - \exp\left\{-\frac{1}{2}\left(\frac{y-T}{\gamma}\right)^2\right\} \right] \quad (2.8)$$

New parameters are introduced. K_{Δ} is no longer the maximum financial loss, it is now defined as the value of loss at a specific distance Δ from the target, and the shape parameter γ is defined as Δ divided by some number. In the Figure 2-6 below, a family of curves for γ ranging from $\Delta/0.1$ to $\Delta/5$ are plotted. Most shape parameters give a shape very close to the original reflected normal loss function.

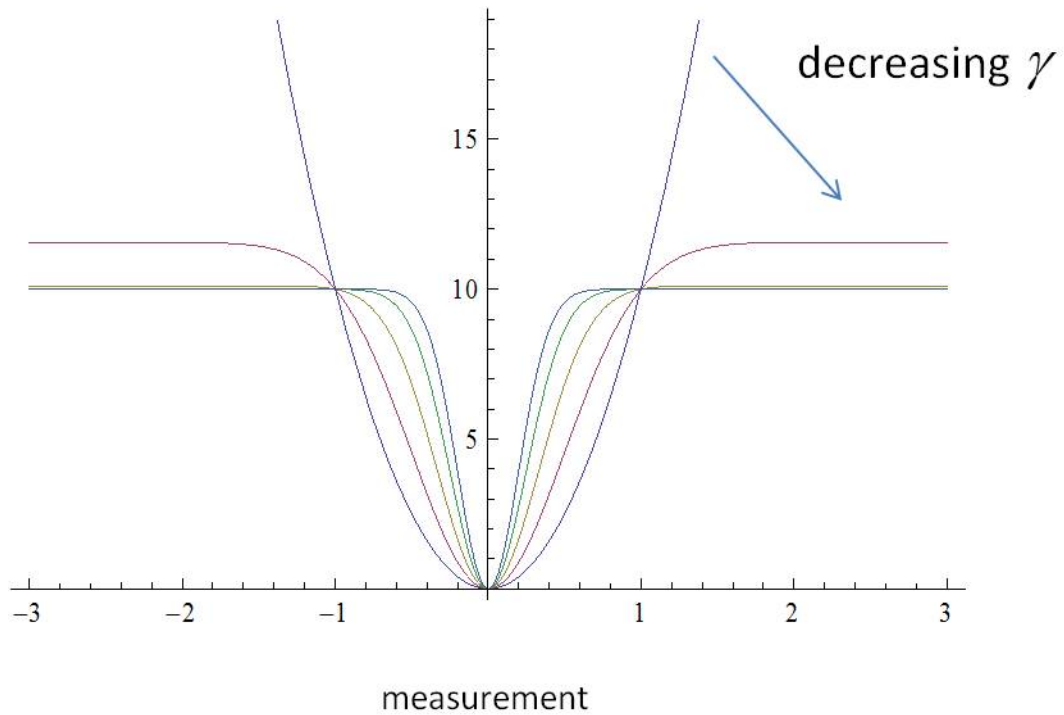


Figure 2-6 Various Spiring's INLF

For $\Delta / 0.1$, it looks just like the quadratic loss function. The other curves, $\Delta / 2$, $\Delta / 3$, $\Delta / 4$ and $\Delta / 5$, look like the reflected normal loss function as proposed by Spiring. In fact, significant difference can only be found in the curve of $\Delta / 2$. The advantage of the MRN loss function is that no maximum financial loss is defined or needed. The MRN loss function is built in a fashion that the maximum loss is caused by the maximum allowed deviation. Observed from figure 2-6, most shape parameters give the same shape as the original reflected normal loss function. It may be interpreted that the MRN is a RNLF with a greater maximum loss.

2.5 Revised Inverted Normal Loss Function

Pan (2000) suggested another modified version of INLF. His view based on a claim that with measurement close to the desired target, there is no financial loss. Hence, his proposed loss function is a mixture of goalpost theorem and Taguchi's loss function, based on the technique developed by Spiring.

Similar to the goalpost theorem, Pan suggests an upper and lower acceptable range of a quality characteristic. Within this interval between U and L , abbreviated for the upper and lower bound, no quality loss is recorded. Outside the interval of U and L , loss is depicted by an inverted normal loss function. Mathematically, Pan's loss function is given by:

$$L(y) = \begin{cases} K_1 \left\{ 1 - \exp \left[-\frac{(y-L)^2}{2\sigma_{L_1}^2} \right] \right\} & y < L \\ 0 & L \leq y \leq U \\ K_2 \left\{ 1 - \exp \left[-\frac{(y-U)^2}{2\sigma_{L_2}^2} \right] \right\} & y > U \end{cases} \quad (2.9)$$

An example of the loss function is plotted in Figure 2-7 underneath, where $L = 1$ and $U = 2$. The variances are set to be 1 and $K_1 = 5$, $K_2 = 8$ representing the maximum possible loss of such a product.

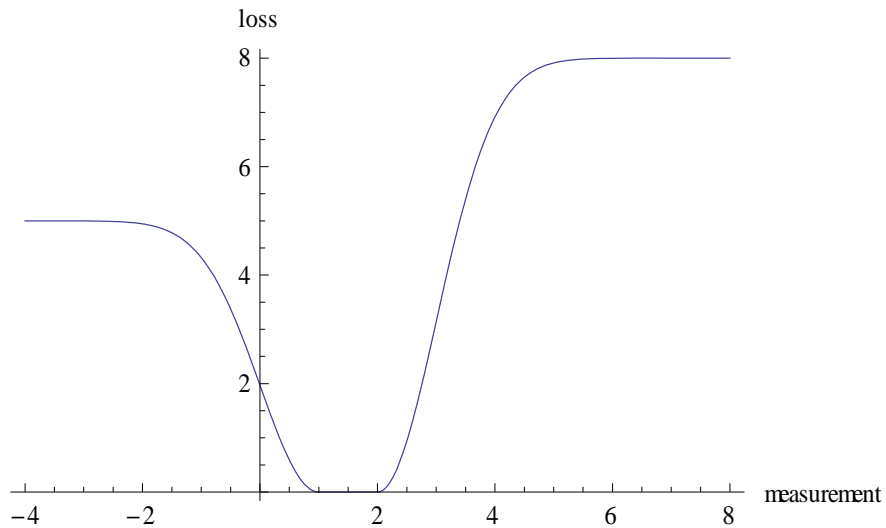


Figure 2-7 An example of Pan's loss function

This loss function is comprehensive in terms of tackling asymmetric loss. Due to the ease in assigning the limiting loss on both sides, financial loss of a quality characteristic which is “nominal-the-best”, “smaller-the-better” or “larger-the-better” can be derived easily. This revised INLF is modifying the INLF derived by Spiring with an interval indicating zero financial loss.

2.6 IPLF base on other distributions

In response to the problem of lack of family of flexible and easy-to-use loss functions, Spiring and Yeung (1998) proposed a general class of loss functions. The

general class is based on the inversions of common probability density functions.

This class of loss functions satisfies the criteria of a loss function that the loss must always be positive, it is minimized at the target value, it is monotonically increasing as the process deviates from target, and it reaches a quantifiable maximum.

Let $g(x, T)$ be a probability density function, which bears a unique maximum, used in creating the loss function for the process of interest, such that

$$\sup_{x \in \Omega} g(x, T) = m \quad (2.10)$$

where $\sup_{x \in \Omega} g(x, T)$ denotes the supremum of $g(x, T)$ for the measurement space, Ω .

The value of m is then the maximum attainable value of $g(x, T)$. The general form of the Inverted Probability Loss Function (IPLF) is defined to be

$$L(x, T) = K \left(1 - \frac{g(x, T)}{m} \right), \quad \forall x \in \Omega \quad (2.11)$$

where x denotes the process measurements, K is the maximum loss and T is the process target. Letting $f(x, \tilde{\theta})$ denote the probability density function associated with the behaviour of process measurement x , the general form of the expected loss function associated will be

$$E\{L(x, T)\} = \int_{\Omega} K \left(1 - \frac{g(x, T)}{m} \right) f(x, \tilde{\theta}) dx \quad (2.12)$$

With the proposed general class of IPLF, Spiring and Yeung (1998) developed the Inverted Gamma Loss Function and the Inverted Tukey Loss Function. In Spiring's reflected normal loss function, the inverted "bell" shape allow practitioners to customize a loss function in order to accurately depict losses associated with process departures from the target by varying the shape parameter and the targeted value. For asymmetric data, the Inverted Gamma Loss Function works better. In their paper, the Inverted Gamma Loss Function is defined to be

$$L(x,T) = K \left(1 - \left(\frac{x \exp\left\{1 - \frac{x}{T}\right\}}{T} \right)^{\alpha-1} \right). \quad (2.13)$$

In Figure 2-8 below, it is assumed that $T = 2$, $\alpha = 1.5, 1.6, 1.7, 1.8$ and $K = 0.4$. From the various IGLFs plotted in Figure 2-8, it is obvious that they share the advantages of an INLF, in the sense that a maximum loss can be reached and the shape of the loss function can be adjusted easily by altering the shape parameter.

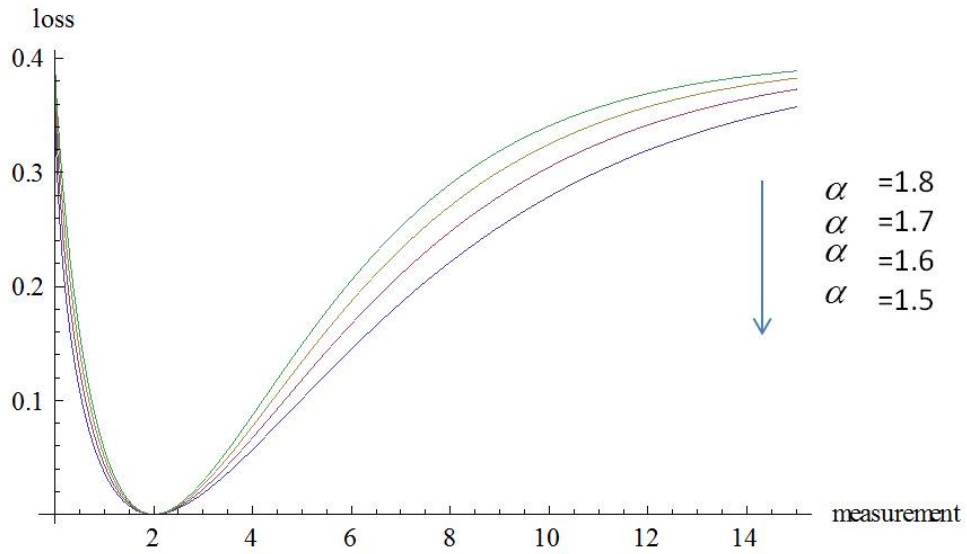


Figure 2-8 Various Inverted Gamma Loss Functions

From figure 2-8, it is observed that the Inverted Gamma Loss Function works very well when there is a small distance between 0 and the target value. The longer tail, on the other hand, may have significantly different shape, depends on the shape parameter.

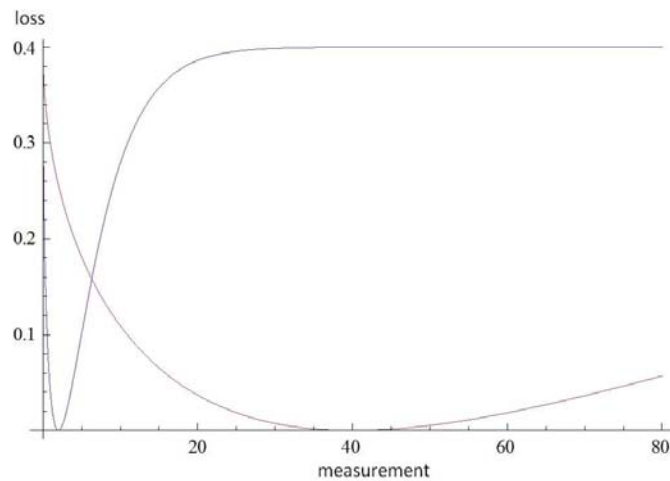


Figure 2-9 Illustration of IGLF's tails

Two more IGLFs are plotted in Figure 2-9, one with the assumption of $T = 2$ and one with $T = 40$. As the target value gets bigger, the loss function looks more or less like a quadratic loss function, with a very long tail on the right side of the target value.

Spiring and Yeung (1998) also built an inverted loss function based on the family of distributions defined by the transformation

$$R(p, \lambda) = \frac{p^\lambda - (1-p)^\lambda}{\lambda} \quad (2.14)$$

where p is bounded between 0 and 1. This family of distributions was introduced by Tukey (1960) and is known as Tukey's symmetric lambda distribution, where $p^\lambda - (1-p)^\lambda$ is a percentile function. In their paper, Spiring and Yeung defined the

characteristic of interest to be $x(p) = T + \frac{p^\lambda - (1-p)^\lambda}{\lambda}$ and the distribution itself is $g(x(p), T) = \frac{1}{p^{\lambda-1} + (1-p)^{\lambda-1}}$, for some λ greater than 0, with target value T for all

p is bounded between 0 and 1. Due to the symmetry of $R(p, \lambda)$,

$$\sup_{x \in \Omega} g(x(p), T) = \begin{cases} 2^{\lambda-2} & \lambda < 1 \text{ or } \lambda > 2 \\ 1 & 1 \leq \lambda \leq 2 \end{cases} \quad (2.15)$$

and the IPLF based on the lambda distribution can be rewritten as

$$L(x(p), T) = \begin{cases} K \left(1 - \frac{2^{2-\lambda}}{p^{\lambda-1} + (1-p)^{\lambda-1}} \right) & \text{otherwise} \\ K \left(1 - \frac{1}{p^{\lambda-1} + (1-p)^{\lambda-1}} \right) & 1 < \lambda < 2 \end{cases} \quad (2.16)$$

The followings are two plots of the IPLF of Tukey distribution according to different

values of λ :

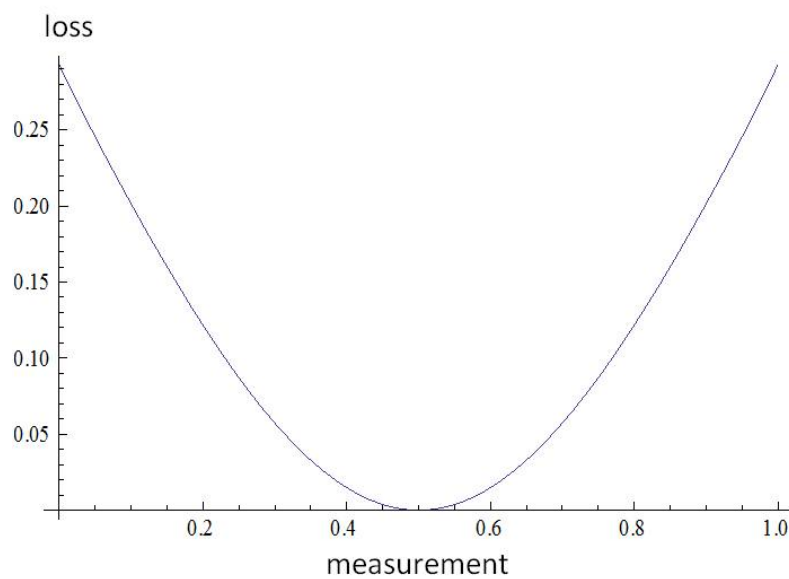
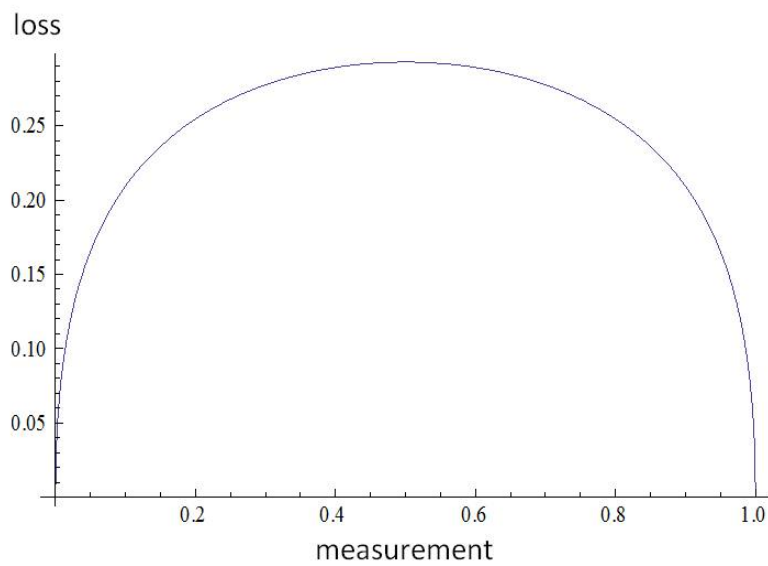


Figure 2-10 Inverted Tukey Loss Function ($\lambda = 1.5$ (above), $\lambda = 2.5$ (below))

The λ of the two plots are 1.5 and 2.5 respectively. There is an anomaly occurs for $1 < \lambda < 2$, where the resulting loss function has maximum loss at the target. Tukey (1977) has found that this family of distributions can be used to give useful approximations to the percentage points of the normal and the Student's t distributions. Since the Tukey lambda distribution is approximately equivalent to a normal distribution and it is considered to be complicated, the inverted probability loss function of Tukey distribution is not widely used.

Leung and Spiring (2002) developed a family of symmetric and asymmetric loss functions based on the Beta distribution, the Inverted Beta Loss Function (IBLF). The most desirable property of the IBLF is that the shape of the IBLF can be modified easily to suit the practitioner's needs, while providing all the properties of the other inverted probability loss functions. Unlike the Inverted Gamma Loss Function, the IBLF can have its shorter tail on either side of the target value, which gives a lot more flexibility to practitioners in developing a loss function.

Given a standard beta distribution with $\alpha > 0$ and $\beta > 0$ having the functional form,

$$f(x) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, 0 < x < 1 \quad (2.17)$$

the modal point and the maxima occurs at

$$T = \frac{\alpha - 1}{\alpha + \beta - 2} \quad (2.18)$$

Assume T , the target value, to be fixed. Using the unique maximum conditions associated with the beta distribution, a linear relationship can be established between α and β through T . The relationship can be rewritten as

$$\alpha = \frac{T}{1-T} \beta + \frac{1-2T}{1-T} \quad (2.19)$$

Based on the technique of developing Inverted Probability Loss Functions, the IBLF can be written as:

$$L(x, T) = K \left\{ 1 - C \left[x(1-x)^{(1-T)/T} \right]^{\alpha-1} \right\} \quad (2.20)$$

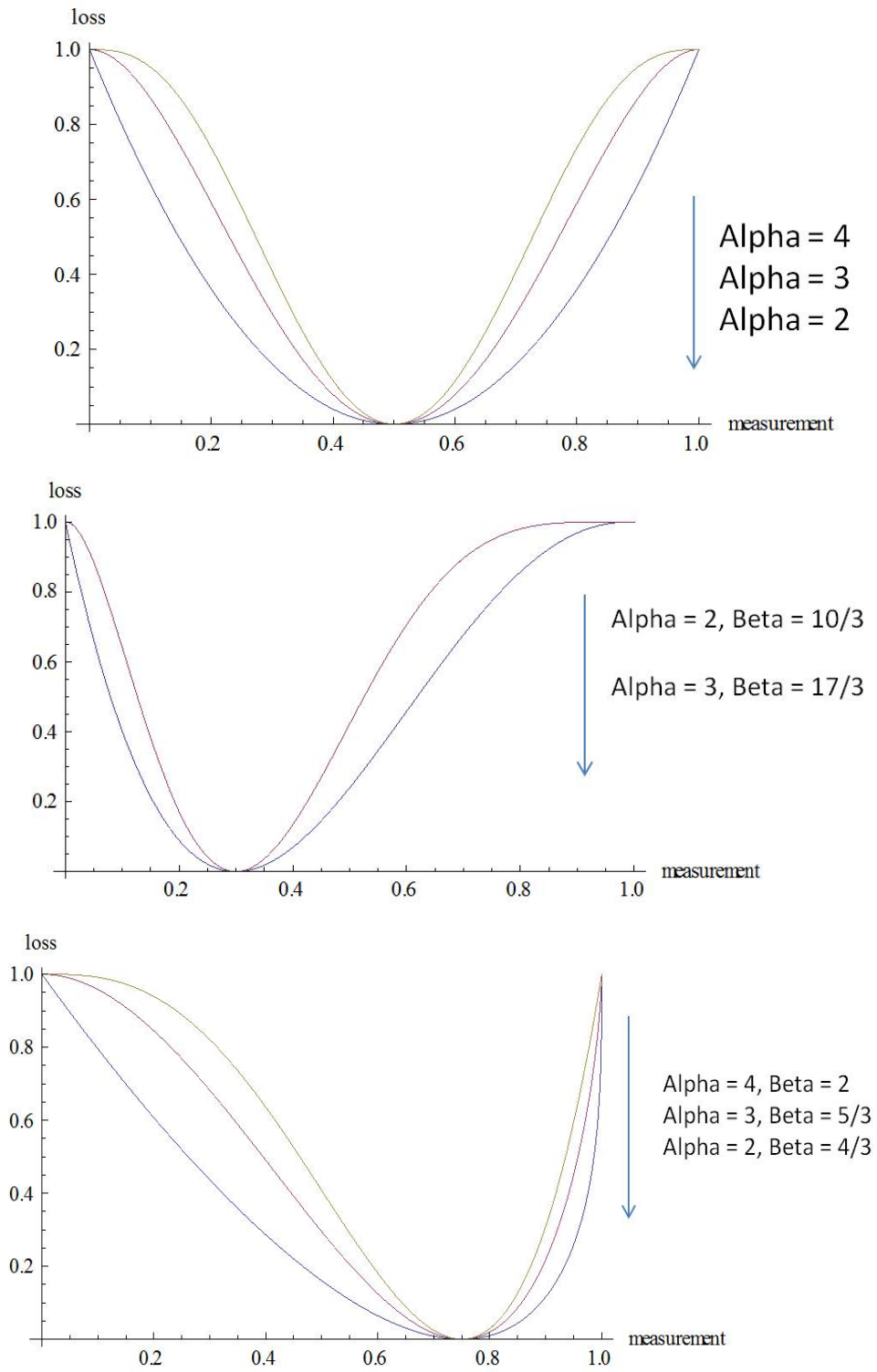


Figure 2-11 Various Inverted Beta Loss Functions

A few IBLFs are plotted above with a fixed target value and various shape parameters. The flexibility of the IBLF is shown in the sense that simple adjustment in the shape parameters can compose different loss functions with different shapes.

The tail of an IBLF could be adjusted to look like a quadratic loss function if $\alpha - 1$ or $\beta - 1$ is less than or equal to 1. From the Figure 2-11 above, the functions with tails which look like quadratic functions are set to have either values of α or β to be bounded between 1 and 2. One characteristic which may worth mentioning is that the right tail of the loss function must be defined straightly, since the tail may have funny looks outside the valid range.

2.7 Properties of the Inverted Probability Loss Functions

Leung and Spiring (2004) established some properties for the loss inversion ratio (LIR) and the family of inverted probability loss functions. Let $f(x)$ be a probability density function possessing a unique maximum at x . Let $T = x$ be the value at which the pdf attains its unique maximum, where T is the target value. Let $\pi(x, T) = f(x)$ and $m = \sup_{x \in X} f(x) = f(T)$, where $\pi(x, T)$ is the probability density function of X in

terms of the target value, T and m is the maximum of $f_x(x)$. The IPLF takes the form of $L(x, T) = K[1 - \pi(x, T)/m]$, where K is the maximum loss incurred when the target is not attained and $\pi(x, T)/m$ is the loss inversion ratio.

Properties of the LIR follow directly from its definition, they include:

1. The LIR represents the relative proportional loss associated with any x value and hence $1 - \pi(x, T)/m$ represents the proportional penalty paid at any value of x .
2. The r -th moment of the LIR is bounded by 0 and 1 for $r \geq 1$. The expectation equals to 1 if $r = 0$.
3. The variance of the LIR is:

$$V[\pi(X, T)/m] = E\{[\pi(X, T)/m]^2\} - \{E[\pi(X, T)/m]\}^2$$

the variance shown above is bounded by 0 and 1.

Properties of IPLF listed in Leung and Spiring (2004) are stated as the following:

1. The general form of the risk function for IPLFs is

$$E[L(X, T)] = E\{K[1 - \pi(X, T)/m]\} = K\{1 - E[\pi(X, T)/m]\}$$
2. The IPLF is bounded between 0 and K , where K is the maximum loss incurred when the target is missed.

3. The variance of $L(X, T)$ arises from using $f_R(x)$ as the process characteristic distribution given by

$$V[L(X, T)] = K^2 \left\{ E\left\{ \left[\frac{\pi(X, T)}{m} \right]^2 \right\} - \left\{ E\left[\frac{\pi(X, T)}{m} \right] \right\}^2 \right\}.$$

4. The variance of an IPLF is K^2 times the variance of LIR. Hence it is bounded between 0 and K^2 .
5. The loss inversion ratio is scale invariant under linear transformation.
6. The shape of IPLF is scale invariant under linear transformation.
7. The loss function is scale invariant under linear transformation.
8. The risk function is scale invariant under linear transformation.

The authors gave proofs for most properties listed above. Out of all these properties for the family of IPLFs, scale invariant under a linear transformation enables practitioners to model financial loss with actual number and scale of a particular characteristic of quality.

Assume IBLF which has unique maximum conditions, then a transformation of the form $y = a + bx$ results in an IBLF with a similar shape but a different scale and/or target. Leung and Spiring (2002) illustrated this property by plots of two IBLF whereas one is based on the transformation of $y = 20 + 20x$.

The two plots in Figure 2-11 below have the same shape. In fact, they share the same scale parameter of $K = 10$ and shape parameters of $\alpha = 2$ and $\beta = 1.54$.

There is only a small adjustment in terms of the loss function itself. For the first

IBLF, the IBLF is defined to be $L(x,T) = 10 \left[1 - \frac{x}{0.65} \left(\frac{1-x}{1-0.65} \right)^{0.54} \right]$, and the second

IBLF, based on linear transformation, is defined to be

$L'(x,T) = 10 \left[1 - \frac{x-20}{0.65(20)} \left(\frac{40-x}{0.35(20)} \right)^{0.54} \right]$. For the second IBLF, the general form of

Beta distribution is used and the minimum and maximum of x is set to be 20 and 40

respectively. In other words, for the generalized beta distribution,

$g(y) = \frac{1}{B(\alpha, \beta)(q-p)} \left(\frac{y-p}{q-p} \right)^{\alpha-1} \left(\frac{q-y}{q-p} \right)^{\beta-1}$, assuming $p = 20$ and $q = 40$, the same

shape is obtained.

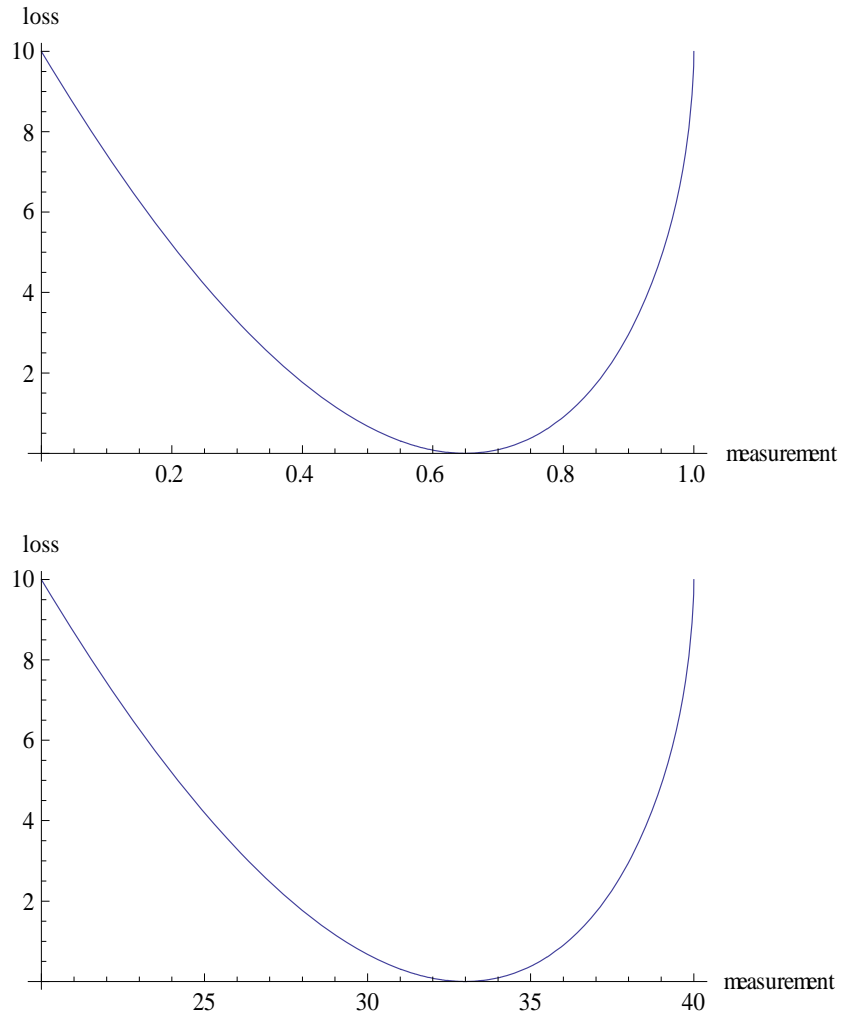


Figure 2-12 Illustration of two IBLF with different scales

As illustrated above, under a linear transformation, the IBLF shares the same shape as the IBLF associated with the standard beta probability density function. This follows directly from the result that a generalized beta distribution can be transformed to a standard beta distribution. If IBLF is used to assess the loss of a process having a normal distribution with mean μ and standard deviation σ , then it is possible to generalize the IBLF with $p = \mu - k\sigma$ and $q = \mu + k\sigma$, with $k \geq 3.5$,

such that it can be used to cover most of the range of the process measurement. Leung and Spiring (2002) also demonstrated the feasibility of IBLF when the process characteristic follows a gamma distribution in their paper. In general, the scale invariant nature of the IBLF under linear transformation holds for any distribution having a unique maximum. With this property, the IBLF can be used to describe financial loss even if the process characteristic follows another distribution.

There are some limitations of the IBLF when the unique maximum conditions do not hold. For example, if both α and β are assumed to equal to one with any target value, the loss will be zero over the range (0, 1) when standard beta is the underlying distribution, whereas zero loss between the two specification limits is unrealistic.

2.8 The Bivariate Inverted Normal Loss Function

In his first paper, Spiring (1993) regarding the inverted normal loss function, Spiring showed the possible extension to the case where two characteristics are of interest in assessing loss. The general form of this bivariate loss function is:

$$L(\mathbf{y}) = K \{1 - \exp[-(\mathbf{y} - \mathbf{T})^T \Gamma^{-1}(\mathbf{y} - \mathbf{T})]\} \quad (2.21)$$

where $\mathbf{y} = \begin{bmatrix} x \\ y \end{bmatrix}$ denotes the vector of quality characteristics, K represents the maximum loss and $\mathbf{T} = \begin{bmatrix} T_x \\ T_y \end{bmatrix}$ indicates the vector of target values. Γ is a matrix

with the form $\Gamma = \begin{bmatrix} \gamma_1 & \gamma_{12} \\ \gamma_{12} & \gamma_2 \end{bmatrix}$, where $\gamma_1 = \frac{\Delta_1}{4}$, $\gamma_2 = \frac{\Delta_2}{4}$, $\gamma_{12} = \frac{\beta\Delta_2^2}{16}$. Δ_1 and Δ_2 are the

Euclidean distance from \mathbf{T} to the point where maximum loss occurs along the principal and the secondary axis respectively. β represents the slope of the principle axis. He derived the loss function based on two values, the target, where zero financial loss should occur and delta, which is the distance between the target and the point where maximum financial cost is obtained. In Leung and Spiring (2004), careful work has been done to illustrate the steps to develop an IPLF, with only the parameters of a given distribution. An inverted bivariate normal loss function can be reworked with the same procedures.

Consider the probability density function of a bivariate normal distribution:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \quad (2.22)$$

and follow the steps similar to other loss functions developed by Leung and Spiring

(2004):

$$\begin{aligned}
f(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \\
\ln f &= \ln \left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \right) - \frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \\
\frac{\delta \ln f}{\delta x} &= -\frac{1}{2(1-\rho^2)} \left[2 \left(\frac{x-\mu_x}{\sigma_x} \right) \frac{1}{\sigma_x} - 2\rho \left(\frac{y-\mu_y}{\sigma_y} \right) \frac{1}{\sigma_x} \right] = 0 \\
\frac{\delta \ln f}{\delta y} &= -\frac{1}{2(1-\rho^2)} \left[2 \left(\frac{y-\mu_y}{\sigma_y} \right) \frac{1}{\sigma_y} - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \frac{1}{\sigma_y} \right] = 0 \\
x^* &= \mu_x = T_x \\
y^* &= \mu_y = T_y \\
\pi(x, y, T_x, T_y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \\
m(T_x, T_y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{\mu_x - \mu_x}{\sigma_x} \right)^2 + \left(\frac{\mu_y - \mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{\mu_x - \mu_x}{\sigma_x} \right) \left(\frac{\mu_y - \mu_y}{\sigma_y} \right) \right] \right\} \\
&= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \\
L(x, y) &= K \left\{ 1 - \frac{\pi(x, y, T_x, T_y)}{m(T_x, T_y)} \right\} \\
&= K \left[1 - \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \right]
\end{aligned}$$

(2.23)

If it is assumed that $\mu_x = \mu_y = 0, \rho = 0, \sigma_x = 1, \sigma_y = 2, K = 1$, the loss function will

have the form: $L(x, y) = K \left\{ 1 - \exp \left[-\frac{1}{2} \left(x^2 + \frac{y^2}{4} \right) \right] \right\}$, and the 3-dimensional plot

will look like:

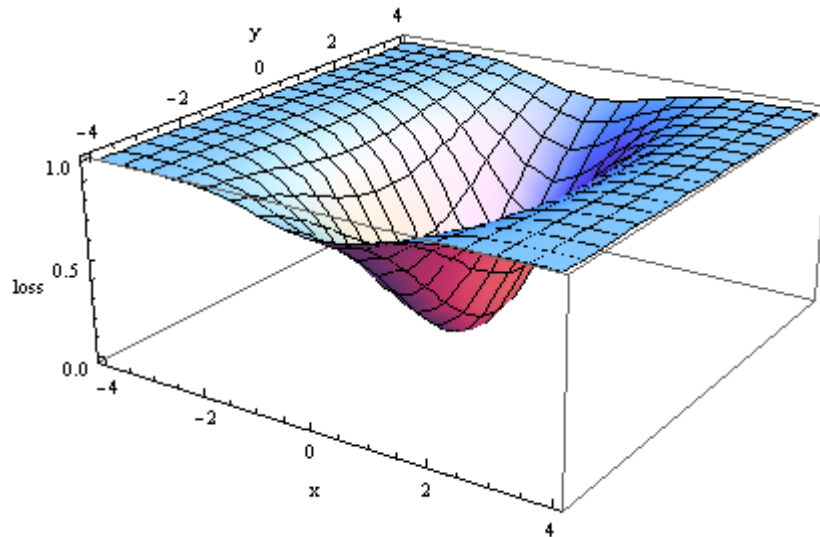


Figure 2-13 A bivariate INLF

According to figure 2-13, it can be observed that the INLF converges to the maximum financial loss; however, with different assumption, for example, coefficient of correlation and standard deviations, attention must be given to make sure the tolerance limits yield the maximum financial loss. In Chapter 4, the author will discuss the selection of parameters for an IbNLF and IDLF.

CHAPTER 3 DIRICHLET DISTRIBUTION

This chapter begins with a brief review of the characteristics of the Dirichlet distribution and works done on simulating random variates for the Dirichlet distribution. While the Dirichlet distribution is a natural choice for the conjugate function when the Inverted Dirichlet Loss Function (IDLF) is applied, its parameters need to be estimated. This chapter will perform the parameter estimation by the method of moments, method of maximum likelihood and the method of percentile matching, and verify the estimates by the Chi-square goodness-of-fit test and a Peacock's two-dimensional Kolmogorov-Smirnov test.

Simulation proves to be a valuable tool when sufficient real data is difficult to be collected. It provides data in this chapter to illustrate how the method of estimation and the two tests are applied. Various simulation techniques have been developed, the three techniques which will be studied in this thesis are: 1) conditional approach, 2) accept-reject method and 3) the method of gamma variates.

3.1 Setup and General Characteristics of the Dirichlet Distribution

Kotz et al. (2000) described the set up of a Dirichlet Distribution. When there are X_0, X_1, \dots, X_m independent random variables, where X_j is a Chi-square distribution with v_j degrees of freedom for $j=0,1,\dots,m$ the joint distribution of Y_1, Y_2, \dots, Y_m ,

where $Y_k = \frac{X_k}{\sum_{i=0}^m X_i}$, $k=1,2,\dots,m$ and $Y_0 = \sum_{i=0}^m X_i$ can be found as:

$$P_{Y_0, Y_1, \dots, Y_m}(y_0, y_1, \dots, y_m) = \left[2^{\frac{1}{2} \sum_{j=0}^m v_j} \prod_{j=0}^m \Gamma\left(\frac{1}{2}(v_j)\right) \right]^{-1} \left[\prod_{j=0}^m y_j^{\left(\frac{v_j}{2}-1\right)} \right] y_0^{\left(\frac{1}{2} \sum_{j=0}^m v_j - 1\right)} \left(1 - \sum_{j=1}^m y_j\right)^{\left(\frac{v_0}{2}-1\right)} e^{-\frac{y_0}{2}} \quad (3.1)$$

defined over $w(\mathbf{y}) = \left\{ (y_0, y_1, \dots, y_m) \mid y_j \geq 0; j=0,1,\dots,m; \sum_{j=1}^m y_j \leq 1 \right\}$.

Integrating out y_0 , and the joint density of Y_1, Y_2, \dots, Y_m becomes:

$$P_{Y_1, \dots, Y_m}(y_1, \dots, y_m) = \frac{\Gamma\left(\frac{1}{2} \sum_{j=0}^m v_j\right)}{\prod_{j=0}^m \Gamma\left(\frac{1}{2} v_j\right)} \left[\prod_{j=1}^m y_j^{\left(\frac{v_j}{2}-1\right)} \right] \left(1 - \sum_{j=1}^m y_j\right)^{\frac{v_0}{2}-1} \quad (3.2)$$

which is also defined over $w(\mathbf{y}) = \left\{ (y_1, y_2, \dots, y_m) \mid y_j \geq 0; j = 1, 2, \dots, m; \sum_{j=1}^m y_j \leq 1 \right\}$. The

above density is often known as the Dirichlet distribution. The standard Dirichlet

distribution is obtained by letting $\frac{1}{2} v_j = \theta_j$, and the density will then be given by:

$$P_{y_1, \dots, y_m}(y_1, \dots, y_m) = \frac{\Gamma\left(\sum_{j=0}^m \theta_j\right)}{\prod_{j=0}^m \Gamma(\theta_j)} \left(1 - \sum_{j=1}^m y_j\right)^{\theta_0 - 1} \left[\prod_{j=1}^m y_j^{\theta_j - 1}\right] \quad (3.3)$$

under the constraints of $\left(0 \leq y_j; j = 1, \dots, m; \sum_{j=1}^m y_j \leq 1\right)$. Because of the simplicity of

this standard Dirichlet distribution, it will be referred as the Dirichlet distribution.

Some authors prefer to write θ_{m+1} instead of θ_0 . This is only a matter of preference

and does not change the Dirichlet distribution.

The mixed moment, μ'_{r_1, \dots, r_m} , of the Dirichlet distribution can be calculated easily and

found its result to be:

$$\begin{aligned}
\mu_{r_1, \dots, r_m} &= E \left[\prod_{j=1}^m Y_j^{r_j} \right] \\
&= \frac{\Gamma \left(\sum_{j=0}^m \theta_j \right)}{\prod_{j=0}^m \Gamma(\theta_j)} \iint \cdots \int \left(1 - \sum_{j=1}^m y_j \right)^{\theta_0 - 1} \prod_{j=1}^m y_j^{\theta_j + r_j - 1} d\mathbf{y} \\
&= \frac{\Gamma \left(\sum_{j=0}^m \theta_j \right) \prod_{j=0}^m \Gamma(\theta_j + r_j)}{\prod_{j=0}^m \Gamma(\theta_j) \Gamma \left(\sum_{j=0}^m (\theta_j + r_j) \right)} \\
&= \frac{\prod_{j=0}^m \theta_j^{r_j}}{\left\{ \sum_{j=0}^m \theta_j \right\}^{\sum_{j=0}^m r_j}}.
\end{aligned} \tag{3.4}$$

If $\Theta = \sum_{j=0}^m \theta_j$, the following results are obtained:

$$E[Y_i] = \frac{\theta_i}{\Theta}, \tag{3.5}$$

$$\text{Var}(Y_i) = \frac{\theta_i(\Theta - \theta_i)}{\Theta^2(\Theta + 1)}, \tag{3.6}$$

$$\text{corr}(Y_i, Y_j) = -\sqrt{\frac{\theta_i \theta_j}{(\Theta - \theta_i)(\Theta - \theta_j)}}. \tag{3.7}$$

The mode of a standard Dirichlet distribution can be found by the first derivative test.

Due to the difficulty of differentiating the joint density, it is easier to differentiate the

log of the joint density.

$$\log p = \log \Gamma \left(\sum_{j=1}^m \theta_j \right) - \sum_{j=1}^m \log \Gamma(\theta_j) + \sum_{j=1}^m (\theta_j - 1) \log y_j + (\theta_0 - 1) \log \left(1 - \sum_{j=1}^m y_j \right) \tag{3.8}$$

Take derivatives with respect to various variables:

$$\begin{aligned}
\frac{\partial \log p}{\partial y_1} &= \frac{\theta_1 - 1}{y_1} - \frac{\theta_0 - 1}{1 - \sum_{j=1}^m y_j}, \\
\frac{\partial \log p}{\partial y_2} &= \frac{\theta_2 - 1}{y_2} - \frac{\theta_0 - 1}{1 - \sum_{j=1}^m y_j}, \\
&\vdots \\
\frac{\partial \log p}{\partial y_m} &= \frac{\theta_m - 1}{y_m} - \frac{\theta_0 - 1}{1 - \sum_{j=1}^m y_j}.
\end{aligned} \tag{3.9}$$

By setting the above equations equal to 0, it can be shown that, the modes, with respect to each variable, are:

$$\begin{aligned}
y_1^* &= \frac{\theta_1 - 1}{\theta_0 + \theta_1 + \dots + \theta_m - m}, \\
y_2^* &= \frac{\theta_2 - 1}{\theta_0 + \theta_1 + \dots + \theta_m - m}, \\
&\vdots \\
y_m^* &= \frac{\theta_m - 1}{\theta_0 + \theta_1 + \dots + \theta_m - m}.
\end{aligned} \tag{3.10}$$

It may be worth noting that, when there is only one variable, the standard Dirichlet distribution, in this case, is also known as the Beta distribution:

$$p_Y(y) = \frac{\Gamma(\theta_0 + \theta_1)}{\Gamma(\theta_0)\Gamma(\theta_1)} y^{\theta_1 - 1} (1 - y)^{\theta_0 - 1} \tag{3.11}$$

where y must be bounded between 0 and 1 and the parameters θ_0, θ_1 must be greater than or equal to 0.

3.2 Conditional Density

Recall the standard Dirichlet distribution, it can be shown that, if y_{s+1}, \dots, y_m are integrated out, the remaining, that is, the variables Y_1, Y_2, \dots, Y_s ($s < m$) have a

standard Dirichlet distribution with parameters $\theta_1, \theta_2, \dots, \theta_s; \Theta - \sum_{j=1}^s \theta_j$.

Based on the result shown above, the conditional joint distribution of

$Y'_j = \frac{Y_j}{1 - \sum_{i=1}^s Y_i}$, ($j = s+1, \dots, m$), given Y_1, Y_2, \dots, Y_s , is a standard Dirichlet

distribution with parameters $\theta_{s+1}, \dots, \theta_m, \theta_0$.

As an illustration, consider a standard Dirichlet distribution with two variables,

X_1 and X_2 . The joint density is:

$$p(x_1, x_2) = \frac{\Gamma(\theta_1 + \theta_2 + \theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2)\Gamma(\theta_3)} x_1^{\theta_1-1} x_2^{\theta_2-1} (1 - x_1 - x_2)^{\theta_3-1}. \quad (3.12)$$

This density is defined in the region of $0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1, 0 \leq x_1 + x_2 \leq 1$. Based on

the bivariate standard Dirichlet distribution, the marginal distributions are:

$$\begin{aligned} p(x_1) &= \frac{\Gamma(\theta_1 + \theta_2 + \theta_3)}{\Gamma(\theta_1)\Gamma(\theta_2 + \theta_3)} x_1^{\theta_1-1} (1 - x_1)^{\theta_2 + \theta_3 - 1} = \text{Beta}(\theta_1, \theta_2 + \theta_3), \\ p(x_2) &= \frac{\Gamma(\theta_1 + \theta_2 + \theta_3)}{\Gamma(\theta_2)\Gamma(\theta_1 + \theta_3)} x_2^{\theta_2-1} (1 - x_2)^{\theta_1 + \theta_3 - 1} = \text{Beta}(\theta_2, \theta_1 + \theta_3). \end{aligned} \quad (3.13)$$

The conditional probability of X_1 given $X_2 = x_2$ is, therefore, a standard Dirichlet distribution with parameters θ_1 and θ_3 :

$$p(x_1 | x_2) = \frac{\Gamma(\theta_1 + \theta_3)}{\Gamma(\theta_1)\Gamma(\theta_3)} \left(\frac{x_1}{1-x_2}\right)^{\theta_1-1} \left(\frac{1-x_1-x_2}{1-x_2}\right)^{\theta_3-1} \frac{1}{1-x_2}, 0 \leq x_1 \leq 1-x_2, 0 \leq x_2 \leq 1. \quad (3.14)$$

3.3 Random Number Generation

Assume one random number is to be generated, instead of going through complex algorithm; mathematics software, like Mathematica and Matlab, can be used to generate a random number easily according to specific probability distributions including the normal distribution, uniform distribution and distributions of gamma family. In the case of a univariate Dirichlet distribution, which is simply the Beta distribution, most mathematical software provides a built-in function to generate a random number handily.

Random number generation technique becomes more complicated when higher dimension is desired. Suppose a vector $x = [x_1 \ x_2 \ \dots \ x_n]^T$ is to be randomly generated, where this n -variable vector is distributed by an n -dimensional probability

distribution. Due to the problem of dependence between various components of this random vector, generation of a random vector is often a non-trivial task and built-in functions are not available. Based on the nature of a standard Dirichlet distribution, it is obvious that there are some sorts of dependence among its variables, as the domain of the distribution is given by $0 \leq x_1, x_2, \dots, x_n \leq 1$ and $\sum_{j=1}^n x_j \leq 1$ when there are n variables in total.

In this thesis, three different random number generation techniques are being considered: 1) conditional approach; 2) acceptance and rejection approach; and 3) gamma factors approach. Johnson (1987) laid clear steps on the first two approaches and the third approach was based on the idea that a Dirichlet variable could be computed by a Chi-square variable divided by the sum of Chi-square variables.

3.3.1 Conditional Approach

The conditional approach provides a mechanism to generate a series of n univariate random numbers to form a random vector. The general process is listed below:

1. Simulate x_1 from a $\text{Beta}(\theta_1, \sum_{j=2}^n \theta_j)$.
2. For $j = 2, 3, \dots, n-1$, simulate ϕ_j from a $\text{Beta}(\theta_j, \sum_{i=j+1}^{n-1} \theta_i)$.
3. For $j = 2, 3, \dots, n-1$, obtain x_j by $x_j = \left(1 - \sum_{i=1}^{j-1} x_i\right) \phi_j$.
4. $x_n = 1 - \sum_{i=1}^{n-1} x_i$.

The process will be continued until all n variants are generated and the random vector is filled. This process is relatively easy to perform and the vast body of techniques in generating univariate random variant can be utilized.

Assuming a random vector of Dirichlet distribution of two components $x = [x_1 \ x_2]^T$ with parameters $[\theta_1 \ \theta_2 \ \theta_3]^T = [2 \ 3 \ 4]^T$ is desired.

The joint distribution $p(x_1, x_2)$ is given by:

$$p(x_1, x_2) = \frac{\Gamma(2+3+4)}{\Gamma(2)\Gamma(3)\Gamma(4)} x_1 x_2^2 (1-x_1-x_2)^3. \quad (3.15)$$

With marginal distribution $p(x_1)$:

$$p(x_1) = \int_0^{1-x_1} \frac{\Gamma(9)}{\Gamma(2)\Gamma(3)\Gamma(4)} x_1 x_2^2 (1-x_1-x_2)^3 dx_2. \quad (3.16)$$

Let $u = \frac{1-x_1-x_2}{1-x_1}$, $x_2 = (1-x_1)(1-u)$, $du = -\frac{dx_2}{1-x_1}$, (3.16) becomes:

$$\begin{aligned}
 p(x_1) &= -\int_1^0 \frac{\Gamma(9)}{\Gamma(2)\Gamma(3)\Gamma(4)} x_1 (1-x_1)^2 (1-u)^2 u^3 (1-x_1) du \\
 &= \frac{\Gamma(9)}{\Gamma(2)\Gamma(7)} x_1 (1-x_1)^6 \int_0^1 \frac{\Gamma(7)}{\Gamma(3)\Gamma(4)} u^3 (1-u)^2 du \\
 &= \frac{\Gamma(9)}{\Gamma(2)\Gamma(7)} x_1 (1-x_1)^6 \\
 &= \text{Beta}(2,7).
 \end{aligned} \tag{3.17}$$

Following the procedure, condition distribution of X_2 given $X_1 = x_1$:

$$\begin{aligned}
 p(x_2 | X_1 = x_1) &= \frac{p(x_1, x_2)}{p(x_1)} \\
 &= \frac{\frac{\Gamma(9)}{\Gamma(2)\Gamma(3)\Gamma(4)} x_1 x_2^2 (1-x_1-x_2)^3}{\frac{\Gamma(9)}{\Gamma(2)\Gamma(7)} x_1 (1-x_1)^6} \\
 &= \frac{\Gamma(7)}{\Gamma(3)\Gamma(4)} \frac{1}{1-x_1} \left(\frac{x_2}{1-x_1} \right)^2 \left(\frac{1-x_1-x_2}{1-x_1} \right), 0 \leq x_2 \leq 1-x_1, 0 \leq x_1 \leq 1.
 \end{aligned} \tag{3.18}$$

Let $u = \frac{x_2}{1-x_1}$, $\left| \frac{du}{dx_2} \right| = \frac{1}{1-x_1}$,

$$\begin{aligned}
 p(u | X_1 = x_1) &= \frac{\Gamma(7)}{\Gamma(3)\Gamma(4)} u^2 (1-u)^3 \\
 &= \text{Beta}(3,4), 0 \leq u \leq 1, 0 \leq x_1 \leq 1.
 \end{aligned} \tag{3.19}$$

The marginal distribution of X_1 based on such setting gives an univariate distribution of Beta(2, 7). The following result is obtained:

1. Generate x_1 based on Beta(2,7). By the `RandomReal[BetaDistribution[2,7]]`

command of Mathematica, 0.173531 is obtained.

2. Generate ϕ_2 based on Beta(3, 4). With the RandomReal command in Mathematica, 0.486983 is obtained.
3. x_2 is then obtained by $x_2 = (1 - 0.173531)0.486983 = 0.402476$.

$x = [0.173531 \quad 0.402476]^T$ becomes the random vector as requested for a bivariate Dirichlet distribution.

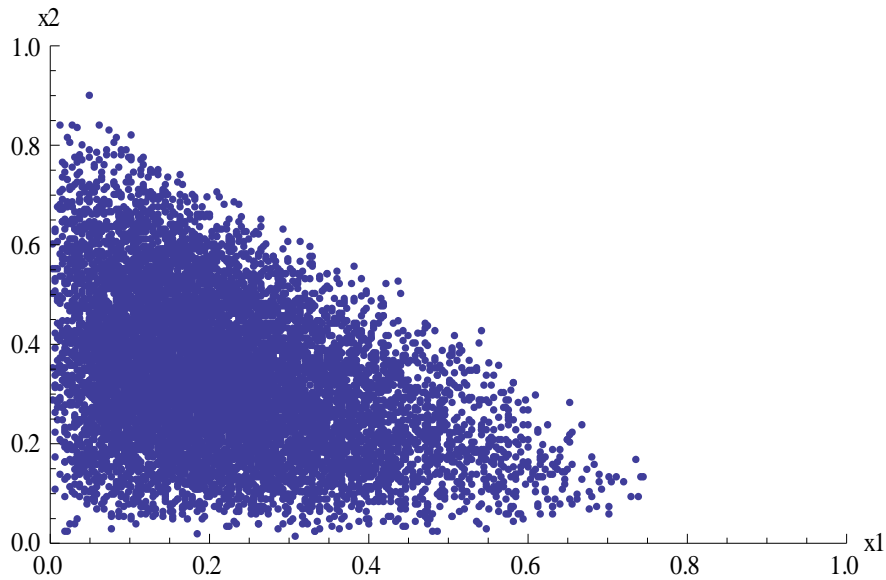


Figure 3-1 Scatter plot of 10000 sample points generated by the conditional approach

In Figure 3-1 above, ten thousand random variates are generated and plotted according to the procedure mentioned previously. The horizontal axis represents X_1 , and the vertical axis represents X_2 . Very small amount of data points are observed when x_1 is greater than 0.7. This can be explained by studying the marginal and

conditional probability density functions. The plot below shows the marginal distribution used in this example (i.e., equation (3.17)).

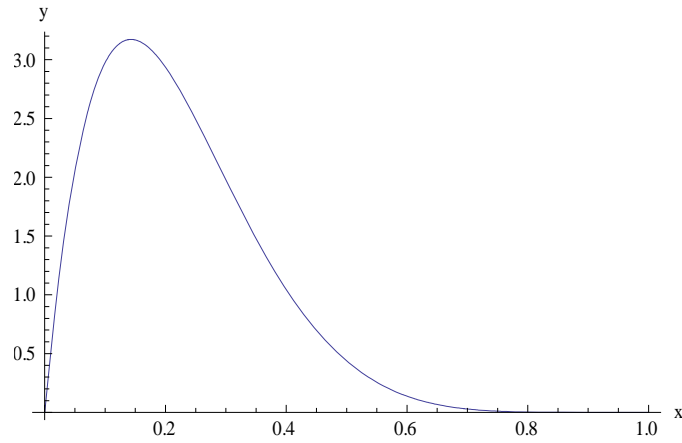


Figure 3-2 Plot of Beta(2,7) pdf

The marginal density gives a rightly skewed distribution, in other words, a thin right tail, and very little weight is found when x is greater than or equal to 0.7, which reflects on the scatter plot of 10000 random vectors.

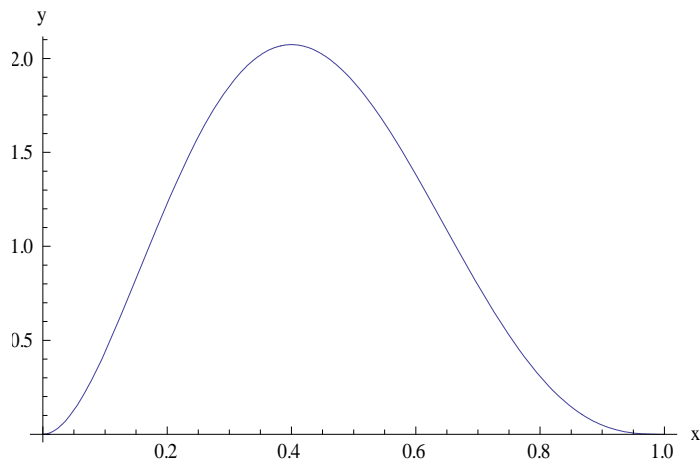


Figure 3-3 Plot of the conditional distribution, Beta(3,4)

The conditional distribution is Beta(3,4), as illustrated in Figure 3-2 above. It is obvious that the curve looks more like a normal curve and weights are split almost equally. In the scatter plot of 10000 normal vectors, one can see that the data on the vertical axis are spread evenly, just like the plot of its probability density function. Although some characteristics between the scatter plot and probability distribution function plot match well, further analysis is required to test whether the random number generation is satisfactory, and this will be addressed in the subsequent sections.

3.3.2 Acceptance and Rejection Approach

If the conditional density or an inverse cumulative distribution function does not exist or is not readily available, the conditional approach or the transformation approach may not be applicable. Acceptance and rejection approach is a better alternative, requires neither the conditional density nor the inverse cumulative distribution function.

The acceptance and rejection approach is being used frequently due to its ease of application. For example, one would like to generate a random number from an unknown or complex probability density function $f(x)$. Suppose, there is a known pdf $g(x)$ in which there is an algorithm to generate random number easily, and contains a maximum value similar to $f(x)$. This method requires two random numbers, one from $g(x)$ and one from the uniform distribution, they are for the domain and the range of $f(x)$. A general procedure is outlined below:

1. Generate a random number x distributed by $g(x)$.
2. Generate a random number u from the uniform distribution within the interval of 0 and 1.
3. If $u \leq \frac{f(x)}{cg(x)}$, then accept x , otherwise, repeat steps 1 and 2.

c is a constant, which is hoped to be close to 1. If $c = 1$, the maximum value of

$u = \frac{f(x)}{cg(x)}$ equals to 1 is attained when $f(x) = g(x)$. The minimum value of u is 0

when $f(x) = 0$. In other words, $0 \leq u \leq 1$, which allows the generation of uniform distribution between 0 and 1.

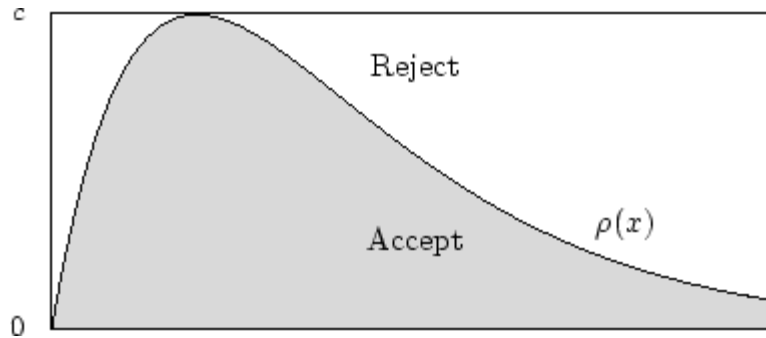


Figure 3-4 Illustration of acceptance and rejection approach

Figure 3-4 above display the idea of the acceptance and rejection approach clearly, where $\rho(x) = \frac{f(x)}{g(x)}$. For an x , generated by $g(x)$, will be compared with the random number, u , generated by the uniform distribution and the value given by $\rho(x)$, if $c = 1$. If u is less than the value of $\rho(x)$, x will be accepted; otherwise, the procedure will be repeated.

In light of generating a random vector of a standard Dirichlet distribution, this procedure of generating a univariate variable must be modified. Since it is difficult to find a function which has a similar shape to a standard Dirichlet distribution, uniform distribution will be used instead. The amended procedure, for generating a random vector with two components (bivariate variate), is stated as follow:

1. Generate a random number x as distributed by the Uniform distribution $(0,1)$.
2. Generate a random number y as distributed by the Uniform distribution $(0, 1-x)$.

3. Generate a random number z as distributed by the Uniform distribution $(0, max)$, where max is the functional value given at the mode of the standard Dirichlet distribution.
4. If z is less than $f(x, y)$, then (x, y) is accepted, otherwise, repeat steps 1 to 3.

For illustration, consider the generation of a random vector with two components,

which has a standard Dirichlet distribution with $[\theta_1 \ \theta_2 \ \theta_3]^T = [2 \ 3 \ 4]^T$. The

mode can be found easily as $T_x = \frac{\theta_1 - 1}{\theta_1 + \theta_2 + \theta_3 - 3} = \frac{1}{6}$ and $T_y = \frac{\theta_2 - 1}{\theta_1 + \theta_2 + \theta_3 - 3} = \frac{1}{3}$,

where T_x and T_y represent the mode of X and Y respectively.

$f(T_x, T_y) = \frac{\Gamma(9)}{\Gamma(2)\Gamma(3)\Gamma(4)} T_x T_y^2 (1 - T_x - T_y)^3 = \frac{70}{9}$ gives the maximum value given by

the specific standard Dirichlet distribution at the mode. The procedure of generating

a random vector under this specific setting is listed as follows:

1. Generate a random number x which is Uniform(0,1). Using `RandomReal[UniformDistribution[{0,1}]]` in Mathematica gives an $x = 0.136482$.
2. Generate a random number y which is Uniform(0, 0.863518) and 0.251819 is returned.
3. Generate a random number z as distributed by Uniform(0, 70/9), where 6.4903 is returned.

4. Calculate $f(0.136482, 0.251819) = 6.65587$. Since $6.4903 < 6.65587$, then $(0.136482, 0.251819)$ is accepted.

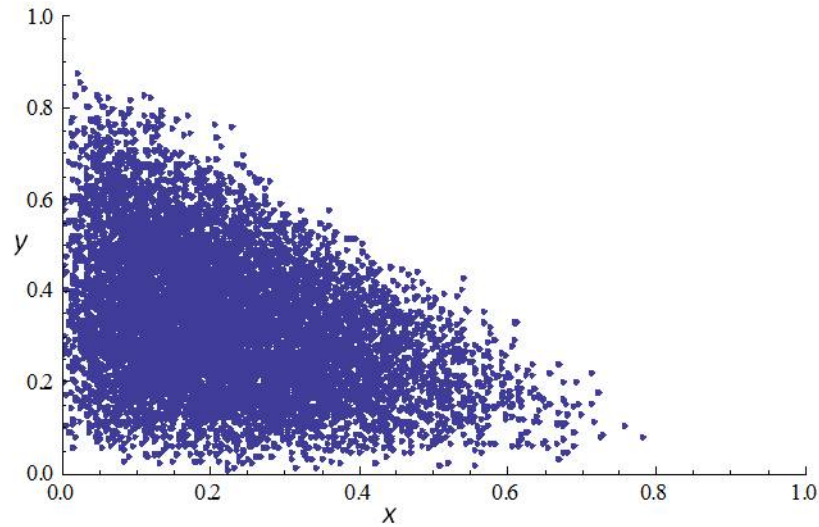


Figure 3-5 Scatter plot of data generated by the acceptance and rejection approach

A similar plot to the one plotted by the conditional approach is attached here, when 10000 random vectors are generated. Although specific tool has not been set up to compare these two methods, visual inspection shows that the data points with the acceptance and rejection approach appears closer to the origin than its outer bound, which is the line segment formed by the function $x + y = 1$.

3.3.3 Gamma Factors Approach

Luc Devroye (1986) pointed out that, based on the relationship between a standard Dirichlet distribution and a gamma distribution, a Dirichlet random variable X_j can

be expressed as $\frac{Y_j}{\sum_{j=1}^m Y_j}$, $j = 1, 2, \dots, m$, where each random variable Y_j is distributed by

Chi-square distribution with ν_j degrees of freedom, $j = 1, 2, \dots, m$. Since Chi-square distribution is a special case of gamma distribution, therefore, an n -variate standard Dirichlet distribution can be expressed by n gamma factors. The idea of this method

is to generate n random variables, each by $\text{gamma}(\alpha_j, 1)$ and $X_j = \frac{Y_j}{\sum_{i=1}^n Y_i}$.

For example, a random vector with two components which has a standard Dirichlet distribution with $[\theta_1 \ \theta_2 \ \theta_3]^T = [2 \ 3 \ 4]^T$ is desired. The generating procedure is listed below:

1. Generate a random number y_1 which is $\text{gamma}(2, 1)$. Using the command `RandomReal[GammaDistribution[2,1]]` in Mathematica gives a random real number of 3.28525.
2. Generate a random number y_2 which is $\text{gamma}(3, 1)$. Mathematica gives a number of 5.83422.

3. Generate a random number y_3 which is gamma(4, 1). 4.1743 is returned by

Mathematica.

4. x_1 is then equated to $\frac{3.28525}{3.28525 + 5.83422 + 4.1743} = 0.247127$ and

$$x_2 = \frac{5.83422}{3.28525 + 5.83422 + 4.1743} = 0.438869.$$

Figure 3-6 is the scatter plot of 10000 random vectors generated by gamma factors.

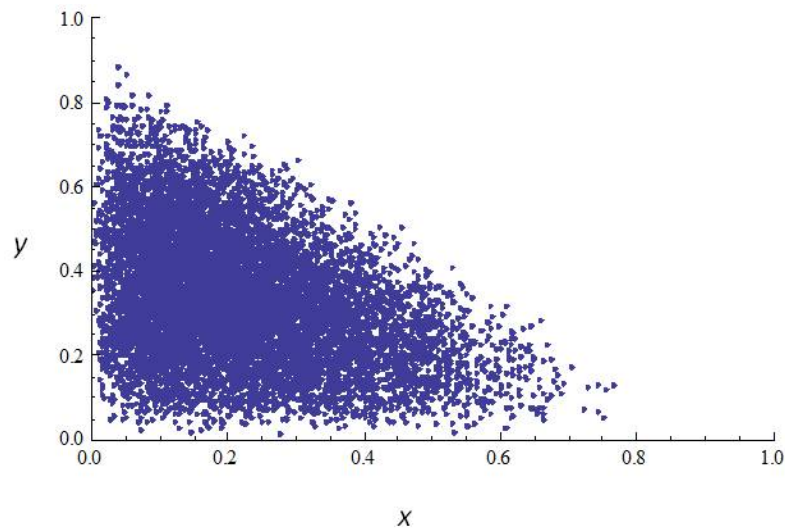


Figure 3-6 Scatter plot of data generated by gamma factors

The scatter plot obtained by the gamma factors approach (Figure 3-6) looks almost identical to the plot given by the acceptance and rejection approach (Figure 3-4). A goodness-of-fit test needs to be performed to compare all three methods.

In Mathematica, there is a built-in function Timing[], where the time used in carrying out a command can be timed. The following table shows the time used in one trial of 10000 random vectors generated by each methods.

Table 3-1 Computational time for the three simulation methods

	Conditional approach	Acceptance and rejection approach	Gamma factors approach
Time used (seconds)	0.78	2.262	2.277

If cpu time is the only concern for generating random numbers with a standard Dirichlet distribution, the conditional approach seems to be the best. As mentioned before, these approaches will be tested and compared by a goodness-of-fit test. Before a goodness-of-fit test can be carried out, estimations on the parameters must be done.

3.4 Parameter Estimations

The Dirichlet distribution plays an important role in Bayesian statistics as one of the popular choices of prior distribution. Due to its association with the gamma functions, parameter estimations have not been easy for the standard Dirichlet distribution. Three common methods of estimating parameters have been studied,

and they are: 1) method of moments; 2) method of maximum likelihood estimation and 3) percentile matching method.

3.4.1 Method of Moments Estimation

Recall the standard bivariate Dirichlet distribution, three parameters, $\theta_1, \theta_2, \theta_3$, are involved. For method of moments with bivariate distributions, $p(x, y)$, the first and second moments of both variables are required to estimate the method of moments estimators. Johnson and Kotz (2002) gave a brief walkthrough with the method.

Population moments are:

$$E[X_i] = \frac{\theta_i}{\Theta}. \quad (3.20)$$

$$E[X_i^2] = \frac{\theta_i(\theta_i + 1)}{\Theta(\Theta + 1)}. \quad (3.21)$$

where $\Theta = \sum_{i=1}^k \theta_i$. Consider a k -variate Dirichlet distribution, $k-1$ first moments and $k-1$ second moments are available, which leads to $\binom{2k-2}{k}$ possible combinations for estimating the k parameters. Since method of moments estimation is not unique, different combination of moments gives different estimation. Although calculations are not difficult to handle, the non-uniqueness in estimation is not satisfactory.

Consider a list of random vectors, being generated by a standard bivariate Dirichlet distribution with parameters $[\theta_1 \ \theta_2 \ \theta_3]^T = [2 \ 3 \ 4]^T$, by one of the three methods discussed in Section 3.3. With a bivariate Dirichlet distribution, three estimators are needed and the first moments and second moments, expressed by the parameters are listed below:

Table 3-2 The first two moments of Beta($\theta_1, \theta_2, \theta_3$)

	First Moment	Second Moment
X_1	$\frac{\theta_1}{\theta_1 + \theta_2 + \theta_3}$	$\frac{\theta_1(\theta_1 + 1)}{(\theta_1 + \theta_2 + \theta_3)(\theta_1 + \theta_2 + \theta_3 + 1)}$
X_2	$\frac{\theta_2}{\theta_1 + \theta_2 + \theta_3}$	$\frac{\theta_2(\theta_2 + 1)}{(\theta_1 + \theta_2 + \theta_3)(\theta_1 + \theta_2 + \theta_3 + 1)}$

A table of sample moments are listed below:

Table 3-3 The first two sample moments of empirical data

	First Moment	Second Moment
X_1	$M'_{11} = \frac{1}{n} \sum_{i=1}^n X_{i1}$	$M'_{21} = \frac{1}{n} \sum_{i=1}^n X_{i1}^2$
X_2	$M'_{12} = \frac{1}{n} \sum_{i=1}^n X_{i2}$	$M'_{22} = \frac{1}{n} \sum_{i=1}^n X_{i2}^2$

The idea of method of moments is to match the sample moments to the population moments, in order to estimate the unknown parameters. As mentioned previously, four equations are available while only three are needed.

In the previous example, a standard Dirichlet distribution is chosen as the joint distribution function with parameters $[2 \ 3 \ 4]^T$. Ten thousand random vectors have been generated and moments are calculated as follows:

Table 3-4 Table of the first two moments based on 10000 generated random vectors

	First Moment	Second Moment
X_1	0.2246	0.0675
X_2	0.3325	0.1328

If the first moments of X_1 and X_2 and the second moment of X_1 are used to estimate the parameters, the following equations are obtained:

$$\begin{aligned}\hat{\theta}_1 &= \frac{(M'_{11} - M'_{21})M'_{11}}{M'_{21} - (M'_{11})^2} = \frac{(0.2246 - 0.0675)0.2246}{0.0675 - 0.2246^2} = 2.0669 \\ \hat{\theta}_2 &= \frac{(M'_{11} - M'_{21})M'_{12}}{M'_{21} - (M'_{11})^2} = \frac{(0.2246 - 0.0675)0.3325}{0.0675 - 0.2246^2} = 3.0602 \\ \hat{\theta}_3 &= \frac{(M'_{11} - M'_{21})(1 - M'_{11} - M'_{21})}{M'_{21} - (M'_{11})^2} \\ &= \frac{(0.2246 - 0.0675)(1 - 0.2246 - 0.3325)}{0.0675 - 0.2246^2} = 4.0761\end{aligned}$$

To illustrate the difference, the parameters are estimated again with the first moments and the second moment of X_2 , the followings are obtained:

$$\begin{aligned}\hat{\theta}_1 &= \frac{(M'_{12} - M'_{22})M'_{11}}{M'_{22} - (M'_{12})^2} = \frac{(0.3325 - 0.1328)0.2246}{0.1328 - 0.3325^2} = 2.01641 \\ \hat{\theta}_2 &= \frac{(M'_{12} - M'_{22})M'_{12}}{M'_{22} - (M'_{12})^2} = \frac{(0.3325 - 0.1328)0.3325}{0.1328 - 0.3325^2} = 2.98512 \\ \hat{\theta}_3 &= \frac{(M'_{12} - M'_{22})(1 - M'_{11} - M'_{12})}{M'_{22} - (M'_{12})^2} \\ &= \frac{(0.3325 - 0.1328)(1 - 0.2246 - 0.3325)}{0.1328 - 0.3325^2} = 3.97627\end{aligned}$$

By substituting the equation involving the second moment of X_2 into the one with the second moment of X_1 , it is observed that the difference of $\hat{\theta}_3$ is as large as 0.1 or approximately 2.5%.

The above example concludes that different choice of equations will definitely affect the result in the estimation using method of moments. Since there is no rule derived in which equations should be chosen, a more scientific method should be used to determine a unique set of estimators for the standard bivariate Dirichlet distribution.

3.4.2 Maximum Likelihood Estimation

Fisher (1920) introduced the maximum likelihood estimator and it became one of the most popular statistical techniques to fit data to a model, giving estimates as distribution parameters. The method of maximum likelihood estimation often gives a unique set of answer when differentiation is applicable. Consider the standard Dirichlet distribution:

$$p_{X_1, \dots, X_{k-1}}(x_1, \dots, x_{k-1}) = \frac{\Gamma\left(\sum_{j=1}^k \theta_j\right)}{\prod_{j=1}^k \Gamma(\theta_j)} \left(1 - \sum_{j=1}^{k-1} x_j\right)^{\theta_k - 1} \left[\prod_{j=1}^k x_j^{\theta_j - 1}\right]. \quad (3.22)$$

The likelihood function, L , is given by, when there are n random vectors of X :

$$L = \left[\frac{\Gamma\left(\sum_{j=1}^k \theta_j\right)}{\prod_{j=1}^k \Gamma(\theta_j)} \right]^n \prod_{i=1}^n \left(1 - \sum_{j=1}^{k-1} x_j\right)^{\theta_k - 1} \prod_{j=1}^k \left(\prod_{i=1}^n x_{ij}\right)^{\theta_j - 1}. \quad (3.23)$$

The log-likelihood and its partial derivatives are therefore:

$$\begin{aligned}
\log L &= n \log \Gamma \left(\sum_{j=1}^k \theta_j \right) - n \log \prod_{j=1}^k \Gamma(\theta_j) + (\theta_k - 1) \sum_{i=1}^n \log \left(1 - \sum_{j=1}^{k-1} x_{ij} \right) + \sum_{j=1}^{k-1} (\theta_j - 1) \log \prod_{i=1}^n x_{ij} \\
\frac{\partial \log L}{\partial \theta_j} &= n \frac{\partial \log \Gamma \left(\sum_{j=1}^k \theta_j \right)}{\partial \theta_j} - n \frac{\partial \log \prod_{j=1}^k \Gamma(\theta_j)}{\partial \theta_j} + \log \prod_{i=1}^n x_{ij} \\
&= n \Psi \left(\sum_{j=1}^k \theta_j \right) - n \Psi(\theta_j) + \log \prod_{i=1}^n x_{ij} \\
\frac{\partial \log L}{\partial \theta_k} &= n \frac{\partial \log \Gamma \left(\sum_{j=1}^k \theta_j \right)}{\partial \theta_k} - n \frac{\partial \log \Gamma(\theta_k)}{\partial \theta_k} + \sum_{i=1}^n \log \left(1 - \sum_{j=1}^{k-1} x_{ij} \right) \\
&= n \Psi \left(\sum_{j=1}^k \theta_j \right) - n \Psi(\theta_k) + \sum_{i=1}^n \log \left(1 - \sum_{j=1}^{k-1} x_{ij} \right).
\end{aligned} \tag{3.24}$$

Ψ is known as the digamma function. Due to the difficulty of solving the above partial derivatives when they are set equal to 0, numerical techniques are used to obtain the maximum likelihood estimators. Narayanan (1991) applied the Newton-Raphson method and Ronning (1989) suggested to set all $\theta_j = \min_i X_{ij}$. Johnson,

Kotz and Balakrishnan (2000) used Fisher's scoring method to perform iterations:

$$\begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix}_{(i)} = \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix}_{(i-1)} + \begin{bmatrix} \text{var}(\hat{\theta}_1) & \cdots & \text{cov}(\hat{\theta}_1, \hat{\theta}_k) \\ \vdots & \ddots & \vdots \\ \text{cov}(\hat{\theta}_k, \hat{\theta}_1) & \cdots & \text{var}(\hat{\theta}_k) \end{bmatrix}_{(i-1)} \begin{bmatrix} g(\hat{\theta}_1) \\ \vdots \\ g(\hat{\theta}_k) \end{bmatrix}_{(i-1)}. \tag{3.25}$$

The variance-covariance matrix can be found by the inverse of the Fisher

Information matrix, I , where $I = \{I_{ij}\} = -E \left[\frac{\partial^2 \log L}{\partial \theta_i \partial \theta_j} \right]$ and

$g_k(\hat{\theta}) = n \Psi(\sum \theta_j) - n \Psi(\theta_k) + \sum_{i=1}^n \log x_{ik}$. The initial guess of $\hat{\theta}$ will be based on the

method of moments estimators. Although many authors have written different

algorithm to find the maximum likelihood estimators, the method mentioned above by Johnson, Kotz and Balakrishnan (2000) provided convergent estimation very quickly.

For illustration, ten thousand data points are generated by a Dirichlet distribution with parameters 2, 3 and 4. Using the values estimated by the method of moments as the initial trial, therefore, $[\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3]_0^T = [2.0155 \ 3.0087 \ 3.9981]^T$ and the calculation after one trial is:

$$\begin{bmatrix} 2.0277 \\ 3.0329 \\ 4.0165 \end{bmatrix} = \begin{bmatrix} 2.0155 \\ 3.0087 \\ 3.9981 \end{bmatrix} + \begin{bmatrix} 0.0004 & 0.0004 & 0.0006 \\ 0.0004 & 0.0010 & 0.0010 \\ 0.0006 & 0.0010 & 0.0017 \end{bmatrix} \begin{bmatrix} 13.3011 \\ 30.9829 \\ -11.9365 \end{bmatrix}$$

Another trial results in $[\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3]_2^T = [2.0277 \ 3.0329 \ 4.0165]^T$ and one more run results in $[\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3]_3^T = [2.0277 \ 3.0329 \ 4.0165]^T$, which confirms convergence. Solution usually converges after three trial runs.

3.4.3 Percentile Matching Method

Percentile matching method can also be used in parameter estimation even though it is not as popular as the method of moments and the maximum likelihood estimation method. The idea of the method is to match the percentile points of the empirical and population distribution at two or more particular points, usually 0.05 and 0.95 or 0.10 and 0.90. Klugman, Panjer and Willmot (1998) briefly described the method by saying that a percentile matching estimate of θ in an univariate distribution is any solution of the p equations $\pi_{g_k}(\theta) = \hat{\pi}_{g_k}$, $k = 1, 2, \dots, p$, where g_1, g_2, \dots, g_p are p arbitrarily chosen percentiles. The extension to a bivariate setting is being explored in this research project.

The most obvious problem spanning from the percentile matching problem is the arbitrarily chosen percentiles mentioned above. There is no rule in choosing a percentile. In this study, it is necessary to fit vectors to multivariate Dirichlet distribution. The percentiles used in determining the estimators are arbitrary and there is more than one point with the same “percentile.” Consider a standard bivariate Dirichlet distribution, which has a triangular domain. There exists various vectors of (x_1, x_2) which satisfy $\Pr(X_1 < x_1, X_2 < x_2) \leq 0.10$. This set of vectors is

used to estimate the parameters under percentile matching estimation. In actuarial science, since insurance companies are particularly interested in the tail behavior, where the probability of a large amount of claim may happen, 0.10 and 0.90 or 0.05 and 0.95 are often being considered.

Similar idea is being studied in this project. Since there are three parameters to be estimated for the bivariate Dirichlet distribution, three equations have to be formed.

In other words, three arbitrary points must be pre-determined.

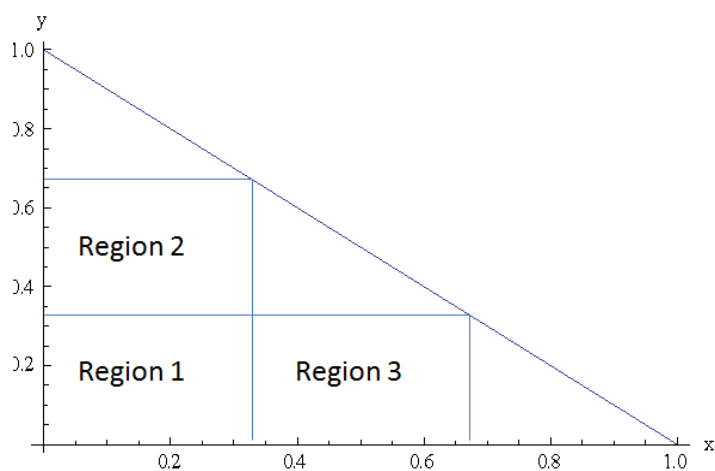


Figure 3-7 Illustration of the setup of the three regions

Consider a smaller triangle with vertices located at $(1/3, 1/3)$, $(1/3, 2/3)$ and $(2/3, 1/3)$.

Three regions are formed as indicated in the above figure. The following procedures may be used:

1. Count the number of vectors less than those three points in both the x and y directions respectively.
2. Find a set of estimators, which gives the same frequency as counted in step 1.

For illustration, 10000 random vectors are simulated using the acceptance and rejection approach. Using the aforementioned setting, it is found that the observed frequencies in the three regions are 3773, 7919 and 5222 respectively. The remaining question is to find the set of $[\hat{\theta}_1 \quad \hat{\theta}_2 \quad \hat{\theta}_3]^T$, such that the following system of equations are satisfied,

$$\left\{ \begin{array}{l} \int_0^{1/3} \int_0^{1/3} \frac{\Gamma(\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3)}{\Gamma(\hat{\theta}_1)\Gamma(\hat{\theta}_2)\Gamma(\hat{\theta}_3)} x^{\hat{\theta}_1-1} y^{\hat{\theta}_2-1} (1-x-y)^{\hat{\theta}_3-1} dx dy = 0.3773 \\ \int_0^{1/3} \int_0^{2/3} \frac{\Gamma(\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3)}{\Gamma(\hat{\theta}_1)\Gamma(\hat{\theta}_2)\Gamma(\hat{\theta}_3)} x^{\hat{\theta}_1-1} y^{\hat{\theta}_2-1} (1-x-y)^{\hat{\theta}_3-1} dx dy = 0.7919 . \\ \int_0^{2/3} \int_0^{1/3} \frac{\Gamma(\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3)}{\Gamma(\hat{\theta}_1)\Gamma(\hat{\theta}_2)\Gamma(\hat{\theta}_3)} x^{\hat{\theta}_1-1} y^{\hat{\theta}_2-1} (1-x-y)^{\hat{\theta}_3-1} dx dy = 0.5222 \end{array} \right.$$

The built-in function of “Solve” and “FindRoot” in Mathematica cannot be used in this application, due to the complexity of numerical integration in the above system. To tackle this problem, consider a cube of $2.00 \pm 0.02, 3.00 \pm 0.02, 4.00 \pm 0.02$ and dissected into 64000 points. The point of (2.00, 3.00, 4.00) is estimated by the maximum likelihood estimation method. For the 64000 points, each point will be

evaluated by numerical integration within the three sub-domain listed above. The point, which gives the desired set of density, will be the estimated parameters obtained by the percentile matching method.

Finding a set of estimated parameters which exactly satisfies the desired density may not be possible, the set with the minimum distance from our target is selected as the set of estimate. In this procedure, a set of estimate $[1.96 \ 3.02 \ 4.01]^T$ is obtained.

Unlike the method of moments, the percentile matching method does not require differentiation techniques and approximation technique like the scoring method; however, approximation needed, due to the inaccuracy of numerical integrations, an exact set of estimates which satisfies the frequencies given by the three pre-selected points are often impossible. This method of percentile matching offers a lot of flexibility to statisticians, as mentioned previously, an unlimited number of set of estimates can be found with different selection of points to be estimated.

Recall the generated random vectors and estimates according to the maximum likelihood estimation, method of moments and the percentile matching method previously, the differences between the three sets of estimates are quite small. All

methods, as discussed, have their respective advantages and disadvantages. To examine these three methods, a test must be employed to determine which method gives the better estimates.

3.5 Biasedness of the Estimation Methods

Biasedness is an interesting topic in studying an estimator. An unbiased estimator is often preferable. Statistically speaking, if $\hat{\theta}$ is an estimator of a parameter θ , then $E[\hat{\theta}] = \theta + bias$ in general. Consider the bivariate Dirichlet distribution being studied; three parameters are to be estimated. Assume the three parameters of the bivariate Dirichlet distribution are $\alpha_1, \alpha_2, \alpha_3$ and their corresponding estimators are a_1, a_2, a_3 .

3.5.1 Method of Moments Estimation

Assume the estimators obtained by method of moments depending on the first moments of X and Y and the second moment of X , where X and Y are the two variables of a bivariate Dirichlet distribution. The estimators are:

$$a_1 = \frac{\sum x_i (\sum x_i^2 - \sum x_i)}{(\sum x_i)^2 - n \sum x_i^2}$$
$$a_2 = \frac{\sum y_i (\sum x_i^2 - \sum x_i)}{(\sum x_i)^2 - n \sum x_i^2}$$
$$a_3 = \frac{n \sum x - \sum x^2 - n \sum x^2 + \sum x \sum x^2 - \sum x \sum y + \sum x^2 \sum y}{n \sum x_i^2 - (\sum x_i)^2}$$

Expected values of a_1, a_2, a_3 are difficult to obtain.

3.5.2 Maximum Likelihood Estimation

To study the biasedness of maximum likelihood estimators, formulation of maximum likelihood estimators must be obtained. The following procedures are necessary to derive the maximum likelihood estimators:

$$\begin{aligned}
f &= \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} x^{\alpha_1-1} y^{\alpha_2-1} (1-x-y)^{\alpha_3-1} \\
L &= \left[\frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \right]^n \left(\prod_{i=1}^n x_i \right)^{\alpha_1-1} \left(\prod_{i=1}^n y_i \right)^{\alpha_2-1} \left[\prod_{i=1}^n (1-x_i-y_i) \right]^{\alpha_3-1} \\
l &= \ln L = n \ln \Gamma(\alpha_1 + \alpha_2 + \alpha_3) - n \ln \Gamma(\alpha_1) - n \ln \Gamma(\alpha_2) - n \ln \Gamma(\alpha_3) \\
&\quad + (\alpha_1 - 1) \sum_{i=1}^n \ln x_i + (\alpha_2 - 1) \sum_{i=1}^n \ln y_i + (\alpha_3 - 1) \sum_{i=1}^n \ln(1-x_i-y_i) \\
\frac{\partial l}{\partial \alpha_1} &= n\psi(\alpha_1 + \alpha_2 + \alpha_3) - n\psi(\alpha_1) + \sum_{i=1}^n \ln x_i = 0 \\
\frac{\partial l}{\partial \alpha_2} &= n\psi(\alpha_1 + \alpha_2 + \alpha_3) - n\psi(\alpha_2) + \sum_{i=1}^n \ln y_i = 0 \\
\frac{\partial l}{\partial \alpha_3} &= n\psi(\alpha_1 + \alpha_2 + \alpha_3) - n\psi(\alpha_3) + \sum_{i=1}^n \ln(1-x_i-y_i) = 0,
\end{aligned}$$

where $\psi(\cdot)$ is the digamma function. The maximum likelihood estimators of $\alpha_1, \alpha_2, \alpha_3$ are the solution to the system of partial derivatives. Unfortunately, for the parameters lie within the digamma functions, it cannot be solved explicitly. In other words, there is no closed form for the estimation of $\alpha_1, \alpha_2, \alpha_3$. Hence, biasedness cannot be evaluated easily for the maximum likelihood estimators.

3.5.3 Percentile Matching Method

Statisticians have done very little in studying the biasedness of percentile matching method. As shown in Section 3.4.3, there is no clear expression of the estimators

estimated by the percentile matching method. Thus, biasedness and other characteristics including consistency cannot be evaluated.

3.6 Goodness-of-Fit Test

Different methods of random vector generation and parameter estimations have been discussed in previous sections, a test is needed to examine how well these random vectors agree with the standard Dirichlet distribution having the parameter estimates obtained. Although, as D'Agostino and Stephens (1986) pointed out, most goodness-of-fit techniques are aimed at univariate data and techniques for multivariate data are much less developed. In this study, it is necessary to employ the Chi-Square goodness-of-fit test to test the generated random vectors with the estimated standard Dirichlet distribution in a bivariate setting.

In the classical univariate Chi-square goodness-of-fit test, to test the hypothesis that a random sample X_1, X_2, \dots, X_n has the distribution function $F(x)$, Pearson (1900) partitioned the range of X into M cells. If N_1, N_2, \dots, N_M are the observed number of X_j 's in these cells, then N_j has the binomial distribution with parameters n and

$\Pr(X_j \text{ falls in } E_i) = \int_{E_i} dF(x)$ when the null hypothesis is true. Pearson reasoned that the difference between the observed and expected cell frequencies express lack of fit of the data to the distribution F , which is being tested. The Chi-square test statistic is therefore $\chi^2 = \sum_{i=1}^M \frac{(N_i - np_i)^2}{np_i}$, which approximately follows the $\chi^2_{M-1-\# \text{ of parameters estimated}}$ distribution. X_j may also be tested whether it is a member of a parametric family $\{F(\cdot|\theta): \theta \text{ in } \Omega\}$, where Ω is a p -dimensional parameter space. The estimate $\tilde{\theta}_n$ may be used to estimate θ and test the fit to the distribution $F(\cdot|\tilde{\theta}_n)$. Thus, the estimated cell frequencies become $n\Pr(X_j \text{ falls in } E_i) = np_i(\tilde{\theta}_n) = n \int_{E_i} dF(x|\tilde{\theta}_n)$ and the Chi-square test statistic becomes $\chi^2(\tilde{\theta}_n) = \sum_{i=1}^M \frac{(N_i - np_i(\tilde{\theta}_n))^2}{np_i(\tilde{\theta}_n)}$. This Chi-square goodness-of-fit test possesses great flexibility and it can be applied to univariate or multivariate data and any distribution, although it is usually not the most powerful test.

An objection to this test has been the arbitrariness of cells division. Mann and Wald (1942) recommended that the number M of equi-probable cells to be used. They found that for a sample of size n , which is significantly large, and at significance level α , one should use approximately $M = 4 \left(\frac{2n^2}{c^2(\alpha)} \right)^{1/5}$, where $c(\alpha)$ is the upper α -point of the standard normal distribution. Schorr (1974) reworked the calculation by Mann and Wald (1942), with better approximations, confirmed that the optimal M

was actually smaller than the one given by Mann and Wald. A more convenient value of M was suggested to be $M \cong 2n^{2/5}$.

In the example, where 10000 random vectors are generated, using the formula by Schorr, it is found that the optimal $M = 2 * 10000^{2/5} \doteq 39.8$. The value given by the formula suggested by Schorr does not work, especially when the sub-regions are kept equally large or equi-probable. Due to the triangular shape of domain, sub-region with equivalent size means that each sub-region must also be triangles. The remaining problem involves, how many sub-regions, or how many triangles should be used. If there are approximately 40 sub-regions, each equal side of isosceles triangle would have a length of $1/6$ and results in 36 sub-regions. Due to the three parameters being estimated, the degree of freedom for the Chi-square distribution becomes 32. If the level of significance is 0.05, the critical value will be 46.1943. This critical value seems large enough for any well distributed random vectors; however, according to the example, 36 sub-regions seem too many, as the Pearson statistic increases exponentially as number of sub-regions increases. This is once again, an evidence of opposition against the Chi-square goodness-of-fit test, as the choice of number of cells dominates the decision of the test.

In the example of 10000 random vectors generated by the method of gamma factors, if 9 sub-regions are desired, degrees of freedom equals to 5. The Chi-square goodness-of-fit test is used to verify the maximum likelihood estimates of the parameters of the bivariate Dirichlet distribution with values of $[\hat{\theta}_1 \quad \hat{\theta}_2 \quad \hat{\theta}_3]^T = [2.02766 \quad 3.03293 \quad 4.01645]^T$.

The Chi-square goodness-of-fit test requires the expected frequency within each cell, and integration must be employed to obtain the desired values. Because of the complexity in integration, the built-in function of “NIntegrate” in Mathematica is used; however, this function can only evaluate square or rectangular regions, evaluation of the density when a function of, for example, $y = \frac{1}{3} - x$ is not allowed. For the proposed cells division method, the idea of Riemann sum is needed. Within each triangle, the density is evaluated by dissecting into 100 rectangles with a width of 1/300 and height of 1/300 more or less than the one besides. The idea can be illustrated in the following plot.

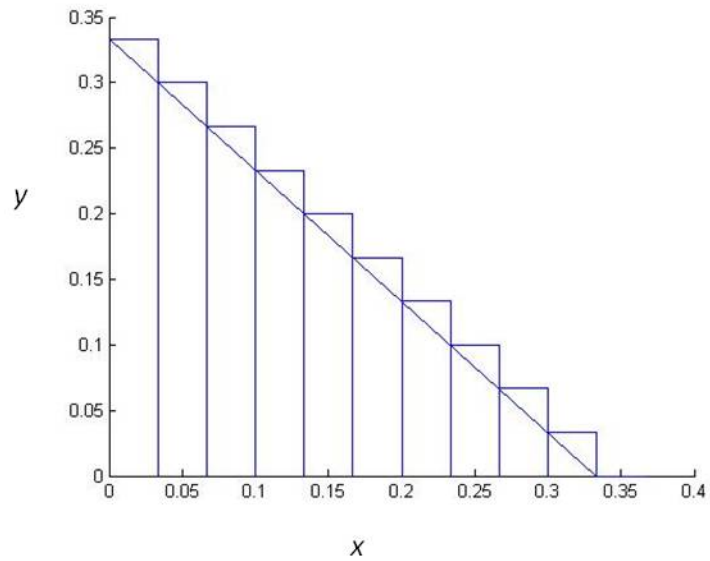


Figure 3-8 Illustration of integrating a triangular region with Riemann sums concept

In the plot above, to serve as a means to illustrate, 10 rectangles are formed instead of 100, which is the actual number of rectangles used in calculation. The idea is that, if the width of each rectangle is small enough, the error incurred will be minimized and an accurate estimation of expected frequency can be obtained. The detail of sub-regions can be seen in the following plot and for the ease of programming and presenting, the cells are named in the following manner:

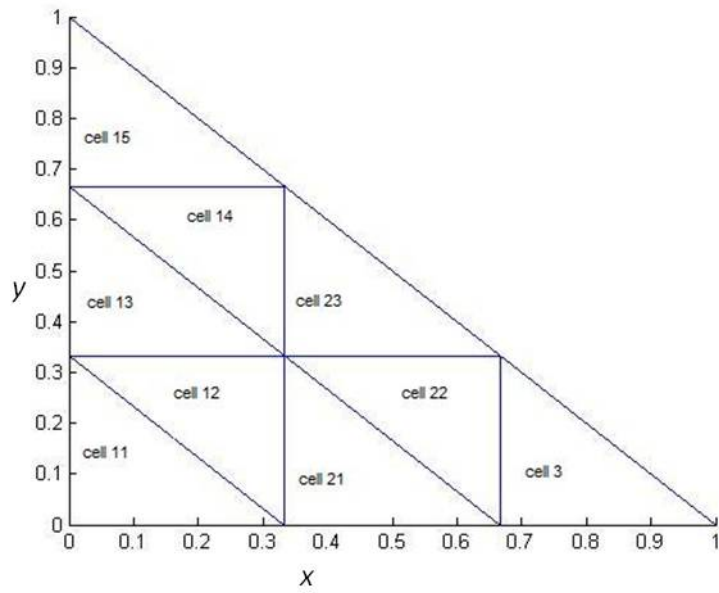


Figure 3-9 Illustration of cells dissection

The result between observed and expected frequencies is found to be:

Table 3-5 Observed versus expected frequency for simulated data

	Observed	Expected		Observed	Expected
Cell 11	864	833.97	$X < 1/3$	8076	8032.34
Cell 12	2944	2911.12	$1/3 < X < 2/3$	1893	1942.02
Cell 13	2710	2689.72	$X > 2/3$	31	25.64
Cell 14	1357	1402.55	$Y < 1/3$	5305	5294.98
Cell 15	201	194.98	$1/3 < Y < 2/3$	4494	4510.04
Cell 21	872	913.32	$Y > 2/3$	201	194.98
Cell 22	594	610.93			
Cell 23	427	417.77			
Cell 3	31	25.64			

Refer to Table 3-5, the value of Pearson statistic obtained is 6.93411 and the critical value for 5 degrees of freedom at $\alpha = 0.05$ is 11.07 and the critical value at $\alpha = 0.10$ is 9.23636, where the null hypothesis, under both levels of significance, is not rejected.

3.7 Comparison among different methods and approaches

In order to compare the result of the three simulation techniques and the three estimation methods, three sets of parameters are considered, namely [2 3 4], [8 6 4] and [1.5 1.5 1.5]. The three sets of parameters cover both ascending and descending order of the three parameters and equivalent parameters, while both integers and non-integers are considered. Simulation of 1000 random variates by each simulation method and each set of parameters aforementioned will be performed. This experiment will be repeated five times and the results will be compared by both Chi-square goodness-of-fit test and Peacock's two-dimensional Kolmogorov-Smirnov test. Readers may be alerted that for the following tables, the p-value calculated by Peacock's two-dimensional Kolmogorov-Smirnov test may be greater than 1 and this will be explained in Subsection 3.7.1.

Table 3-6 Tests results of random variates simulated by conditional approach with [2 3 4]

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	p-value	Max. Diff.	Zn	p-value
Trial 1	MME	2.0616	3.11	3.9959	5.4276	0.2462	0.0326	1.0316	1.1338
	MLE	2.0646	3.1306	4.0229	5.7101	0.2219	0.0307	0.9713	1.2802
	PM	2.2116	3.208	4.1339	4.6734	0.3225	0.056	1.7711	0.0782
Trial 2	MME	2.1359	3.2759	4.3405	3.6118	0.4611	0.0299	0.9462	1.3408
	MLE	2.0223	3.0869	4.079	4.3223	0.3641	0.0365	1.1531	0.8494
	PM	2.0619	3.1999	4.3205	4.6329	0.3271	0.043	1.36	0.4534
Trial 3	MME	2.076	3.2372	4.1896	2.5108	0.6427	0.0371	1.1745	0.8024
	MLE	2.0848	3.2279	4.186	2.4252	0.6581	0.036	1.1389	0.8814
	PM	2.06	3.1412	4.0496	2.0792	0.7212	0.0446	1.4111	0.3782
Trial 4	MME	2.086	3.236	4.308	4.3829	0.3567	0.0289	0.9141	1.417
	MLE	2.0203	3.1217	4.1278	4.6356	0.3268	0.0273	0.8648	1.5307
	PM	1.948	3.122	4.088	5.371	0.2513	0.0354	1.1206	0.9231
Trial 5	MME	1.8758	2.8334	3.8025	5.3758	0.2509	0.04	1.2659	0.6161
	MLE	1.9906	3.0185	4.0361	2.9668	0.5634	0.0338	1.0696	1.0425
	PM	1.8698	2.8894	3.9005	5.299	0.258	0.0431	1.3623	0.4499

Table 3-7 Tests results of random variates simulated by conditional approach with [8 6 4]

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	<i>p</i> -value	Max. Diff.	Zn	<i>p</i> -value
Trial 1	MME	8.2445	6.2247	4.1338	1.5928	0.661	0.0277	0.8756	1.5062
	MLE	8.2913	6.2661	4.1576	1.4123	0.933	0.0268	0.8488	1.5661
	PM	8.4025	6.3187	4.2438	1.3117	0.7264	0.0272	0.859	1.5436
Trial 2	MME	7.5139	5.6462	3.7287	3.3892	0.3354	0.0316	1.0001	1.2102
	MLE	7.7711	5.8486	3.8593	4.523	0.6629	0.0316	0.9992	1.2125
	PM	7.2959	5.4262	3.5427	2.6749	0.4445	0.0453	1.4314	0.3508
Trial 3	MME	7.6938	5.8583	3.9379	3.8419	0.2791	0.0393	1.243	0.6604
	MLE	7.5953	5.7756	3.8729	3.5561	0.9964	0.0393	1.243	0.7147
	PM	7.4738	5.6383	3.8239	2.5064	0.4741	0.0345	1.092	0.9895
Trial 4	MME	8.2183	6.0333	4.072	1.6238	0.654	0.0278	0.8777	1.5015
	MLE	7.9436	5.8134	3.9427	1.4799	0.9977	0.0285	0.9025	1.4442
	PM	8.0583	5.8733	3.882	1.6873	0.6398	0.0349	1.1032	0.9633
Trial 5	MME	7.7121	5.8855	3.8691	1.4743	0.6882	0.0432	1.3646	0.4462
	MLE	7.9438	6.0753	3.9913	1.956	0.968	0.0374	1.1829	0.7843
	PM	7.8901	6.0175	4.0191	1.6349	0.6515	0.0383	1.2111	0.7247

Table 3-8 Tests results of random variates simulated by conditional approach with [1.5 1.5 1.5]

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	<i>p</i> -value	Max. Diff.	Zn	<i>p</i> -value
Trial 1	MME	1.5213	1.4968	1.4684	7.1467	0.21	0.0339	1.0711	1.0389
	MLE	1.58	1.5643	1.5447	6.3819	0.2708	0.0354	1.1182	0.9286
	PM	1.6993	1.6448	1.6164	9.8331	0.0801	0.0502	1.588	0.1861
Trial 2	MME	1.466	1.4513	1.4503	2.2027	0.8204	0.0295	0.9328	1.3728
	MLE	1.4881	1.4689	1.4817	2.0311	0.8448	0.03	0.9475	1.3376
	PM	1.624	1.6453	1.5663	5.7006	0.3365	0.0481	1.5197	0.2483
Trial 3	MME	1.5223	1.4576	1.4776	3.4716	0.6277	0.0273	0.8625	1.5358
	MLE	1.4909	1.4328	1.4814	4.0135	0.5475	0.0325	1.0269	1.1451
	PM	1.4923	1.4556	1.4676	4.1436	0.5289	0.0287	0.9083	1.4308
Trial 4	MME	1.6101	1.6929	1.6429	10.0479	0.0739	0.0354	1.1191	0.9266
	MLE	1.5138	1.5779	1.5472	8.2206	0.1445	0.0321	1.0151	1.1739
	PM	1.7841	1.7909	1.7309	20.0164	0.0012	0.056	1.772	0.0779
Trial 5	MME	1.5212	1.4007	1.4731	2.8829	0.718	0.0287	0.9072	1.4332
	MLE	1.5107	1.3894	1.4625	3.2142	0.667	0.0275	0.8697	1.5197
	PM	1.4592	1.3387	1.4471	5.5458	0.353	0.0297	0.94	1.3556

Table 3-9 Tests results of random variates simulated by gamma factors approach with [2 3 4]

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	<i>p</i> -value	Max. Diff.	Zn	<i>p</i> -value
Trial 1	MME	1.9939	2.9334	3.8905	5.8526	0.2104	0.0289	0.9151	1.4148
	MLE	2.0468	3.0382	4.0055	6.8996	0.1413	0.0322	1.0186	1.1653
	PM	2.1219	3.0294	3.9485	6.9242	0.14	0.0441	1.3955	0.4002
Trial 2	MME	2.067	3.0951	4.1856	10.1999	0.0372	0.0388	1.2278	0.6907
	MLE	1.9605	2.9244	3.9349	6.6366	0.1564	0.0331	1.0481	1.0941
	PM	1.877	2.9171	3.9656	3.7103	0.4466	0.0519	1.6416	0.1464
Trial 3	MME	2.1632	3.3672	4.4039	9.0513	0.0598	0.032	1.013	1.179
	MLE	2.0414	3.1569	4.1279	4.7497	0.314	0.0305	0.9657	1.2938
	PM	2.0752	3.1472	4.2439	6.5665	0.1606	0.0331	1.0479	1.0945
Trial 4	MME	1.9325	2.873	3.8835	3.3085	0.5076	0.0313	0.9895	1.2359
	MLE	1.999	2.9818	3.9994	2.6004	0.6268	0.031	0.9817	1.255
	PM	1.8065	2.805	3.8515	4.9056	0.2971	0.05	1.5823	0.1907
Trial 5	MME	2.0689	3.104	4.1428	6.1423	0.1888	0.0339	1.0729	1.0346
	MLE	1.9859	2.9963	3.9839	4.2258	0.3763	0.0416	1.3153	0.5268
	PM	1.9589	2.992	3.9228	3.9334	0.4151	0.0434	1.3718	0.4352

Table 3-10 Tests results of random variates simulated by gamma factors approach with [8 6 4]

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	<i>p</i> -value	Max. Diff.	Zn	<i>p</i> -value
Trial 1	MME	8.2305	6.2775	4.2718	5.0329	0.1694	0.0299	0.9459	1.3414
	MLE	8.3917	6.4062	4.3549	5.3702	0.1466	0.028	0.8846	1.4856
	PM	8.0165	6.1395	4.1318	4.3801	0.2232	0.037	1.1698	0.8128
Trial 2	MME	8.5646	6.3519	4.2142	1.8941	0.5947	0.0333	1.0537	1.0807
	MLE	8.3757	6.2051	4.1078	1.4824	0.6863	0.0308	0.9726	1.277
	PM	8.4946	6.2859	4.2482	0.745	0.8626	0.0294	0.9284	1.3831
Trial 3	MME	8.2954	6.1063	4.2243	2.393	0.4949	0.0342	1.0827	1.0114
	MLE	8.2038	6.028	4.1775	2.4217	0.4896	0.0324	1.0259	1.1477
	PM	8.4734	6.2963	4.3323	4.1943	0.2412	0.0418	1.3233	0.5132
Trial 4	MME	7.2701	5.5762	3.7771	1.3097	0.7268	0.0343	1.0862	1.0032
	MLE	7.2252	5.5365	3.7472	1.3764	0.7111	0.0333	1.0536	1.0808
	PM	7.4681	5.7722	3.9511	1.5669	0.6669	0.0394	1.2453	0.656
Trial 5	MME	7.6823	5.8069	3.7849	11.863	0.0079	0.0367	1.1595	0.8354
	MLE	7.8336	5.9155	3.8837	11.3113	0.0102	0.042	1.3287	0.5041
	PM	7.6003	5.7229	3.9549	15.3432	0.0015	0.0634	2.0037	0.0215

Table 3-11 Tests results of random variates simulated by gamma factors approach with [1.5 1.5 1.5]

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	p-value	Max. Diff.	Zn	p-value
Trial 1	MME	1.5305	1.4445	1.4663	6.3589	0.2728	0.0308	0.9753	1.2704
	MLE	1.5524	1.4857	1.5112	6.0882	0.2977	0.0291	0.921	1.4008
	PM	1.4925	1.4085	1.4363	7.7396	0.1712	0.0334	1.0556	1.0761
Trial 2	MME	1.5919	1.6063	1.6121	8.2361	0.1437	0.0267	0.8431	1.5786
	MLE	1.5431	1.5602	1.5541	6.68	0.2455	0.0277	0.8762	1.505
	PM	1.4719	1.4343	1.4561	7.2094	0.2055	0.0404	1.2773	0.5949
Trial 3	MME	1.5052	1.6205	1.5165	15.5917	0.0081	0.0346	1.0956	0.9811
	MLE	1.4981	1.6191	1.5383	15.3461	0.009	0.0371	1.1744	0.8027
	PM	1.6992	1.7385	1.5665	26.2427	0.0001	0.0663	2.096	0.0121
Trial 4	MME	1.5718	1.5527	1.5495	1.1336	0.9511	0.0226	0.714	1.8238
	MLE	1.5884	1.5701	1.5748	0.9267	0.9683	0.0218	0.6897	1.86
	PM	1.5238	1.5367	1.5615	1.7694	0.88	0.0274	0.8667	1.5263
Trial 5	MME	1.4935	1.4829	1.5144	2.257	0.8126	0.0249	0.788	1.6927
	MLE	1.4836	1.4679	1.5085	2.496	0.7771	0.0233	0.7371	1.786
	PM	1.4455	1.4749	1.4804	2.4252	0.7877	0.0233	0.7371	1.786

**Table 3-12 Tests results of random variates simulated by acceptance and rejection approach
with [2 3 4]**

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	p-value	Max. Diff.	Zn	p-value
Trial 1	MME	2.1836	3.2246	4.1157	4.6674	0.3232	0.0252	0.7984	1.6721
	MLE	2.0965	3.0795	3.9183	1.4183	0.841	0.0266	0.842	1.581
	PM	2.1276	3.0546	3.8957	1.4691	0.8321	0.0378	1.1959	0.7565
Trial 2	MME	1.9999	2.9618	3.9037	2.2334	0.6929	0.0243	0.7699	1.7273
	MLE	1.9968	2.9409	3.8825	2.3273	0.6758	0.0254	0.8029	1.6629
	PM	1.9439	2.9058	3.7837	2.8529	0.5827	0.031	0.9805	1.2578
Trial 3	MME	1.9475	2.9225	3.8314	0.8745	0.9282	0.0349	1.1039	0.9616
	MLE	1.9526	2.9275	3.8706	1.2344	0.8724	0.0357	1.1285	0.9049
	PM	1.8555	2.7645	3.6714	1.7415	0.7832	0.0353	1.1163	0.9329
Trial 4	MME	1.9204	2.9671	4.0045	2.6395	0.6198	0.0288	0.9117	1.4226
	MLE	1.9231	2.9767	4.0362	2.267	0.6868	0.0291	0.9199	1.4035
	PM	1.9744	3.1211	4.1425	1.3844	0.8469	0.0414	1.3101	0.5358
Trial 5	MME	2.0167	2.8951	4.0245	7.8972	0.0954	0.0336	1.0636	1.0569
	MLE	2.1084	3.0424	4.2148	5.4452	0.2446	0.0317	1.0017	1.2065
	PM	2.0627	3.0491	4.2225	6.0181	0.1978	0.0348	1.1002	0.9704

**Table 3-13 Tests results of random variates simulated by acceptance and rejection approach
with [8 6 4]**

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	p-value	Max. Diff.	Zn	p-value
Trial 1	MME	8.2007	6.0425	4.0832	1.2883	0.7319	0.0392	1.2403	0.6658
	MLE	8.2275	6.053	4.0949	1.4757	0.6879	0.0386	1.2218	0.7029
	PM	8.1987	6.1545	4.1532	0.5905	0.8986	0.0418	1.3222	0.5151
Trial 2	MME	8.5659	6.2581	4.416	10.2116	0.0169	0.0345	1.0917	0.9904
	MLE	8.251	6.0172	4.2366	8.1127	0.0437	0.0322	1.017	1.1692
	PM	8.4059	6.1201	4.482	5.523	0.1373	0.051	1.6128	0.1668
Trial 3	MME	7.5085	5.566	3.784	7.2311	0.0649	0.034	1.0767	1.0257
	MLE	7.7185	5.7466	3.8796	8.0083	0.0458	0.0308	0.9747	1.2718
	PM	7.2925	5.506	3.704	6.4465	0.0918	0.0451	1.4273	0.3562
Trial 4	MME	8.0078	5.9434	4.0014	1.2536	0.7402	0.0328	1.0359	1.1236
	MLE	8.0334	5.9675	4.0093	1.4597	0.6916	0.0329	1.0408	1.1117
	PM	7.9578	5.8554	4.0334	0.6425	0.8866	0.036	1.1387	0.8818
Trial 5	MME	8.2946	6.3467	4.1551	2.9226	0.4037	0.043	1.3598	0.4537
	MLE	8.0077	6.0958	4.0064	3.0183	0.3888	0.0447	1.4125	0.3762
	PM	8.1766	6.1867	3.9411	4.0143	0.2599	0.0602	1.9027	0.0386

Table 3-14 Tests results of random variates simulated by acceptance and rejection approach with [1.5 1.5 1.5]

	Methods	a1	a2	a3	χ^2 goodness-of-fit test		Peacock's 2D K-S test		
					test stat.	<i>p</i> -value	Max. Diff.	Zn	<i>p</i> -value
Trial 1	MME	1.5461	1.5302	1.5182	5.9209	0.314	0.0303	0.959	1.31
	MLE	1.5765	1.5844	1.5732	5.2158	0.3901	0.0333	1.0524	1.0837
	PM	1.7441	1.7262	1.6522	10.1572	0.0709	0.0612	1.9349	0.0322
Trial 2	MME	1.5128	1.5542	1.4968	2.0757	0.8386	0.028	0.8852	1.4843
	MLE	1.4921	1.5141	1.4715	2.3644	0.7968	0.0261	0.8243	1.6188
	PM	1.5542	1.5048	1.7108	1.9973	0.8495	0.0264	0.8352	1.5955
Trial 3	MME	1.5312	1.6597	1.6719	6.6986	0.244	0.0289	0.9142	1.4169
	MLE	1.5509	1.6812	1.6844	6.5978	0.2523	0.0292	0.9226	1.3969
	PM	1.5412	1.6317	1.6659	7.3135	0.1984	0.0373	1.178	0.7948
Trial 4	MME	1.4666	1.5234	1.5047	6.2153	0.2858	0.026	0.8229	1.6216
	MLE	1.4442	1.5023	1.4736	7.0022	0.2205	0.0255	0.8072	1.6543
	PM	1.6346	1.7154	1.6827	11.279	0.0461	0.0475	1.5025	0.2663
Trial 5	MME	1.5511	1.6072	1.5157	5.4878	0.3593	0.0274	0.8669	1.526
	MLE	1.5067	1.5524	1.4896	6.6831	0.2453	0.0294	0.9283	1.3835
	PM	1.7351	1.8052	1.6217	10.071	0.0732	0.059	1.8657	0.0475

In this simulation exercise, one thousand random variates are generated for each trial.

The three estimating techniques are then applied to each trial and the Chi-square goodness-of-fit test and Peacock's two-dimensional Kolmogorov-Smirnov test (2D K-S test) are applied to obtain the *p*-value. Assuming a 0.10 level of significance, thus a *p*-value being less than 0.10 would result in rejection of the proposed model.

The aforementioned procedure would be carried out for five times for each random seed. The pre-selected random seeds are [2, 3, 4], [1.5, 1.5, 1.5] and [8, 6, 4] respectively. The selection of random seeds includes scenarios of: $a_1 < a_2 < a_3$, $a_1 = a_2 = a_3$ and $a_1 > a_2 > a_3$. There are three simulating techniques being considered in this exercise and they are: 1) acceptance and rejection approach, 2) conditional approach and 3) gamma factors approach. In short, there are a total of 45 trials, 135 sets of estimates and 270 tests being conducted.

3.7.1 Discussion on the Two Tests

When reading the results above, one may concern about the last column of in table 3-6 to 3-14, where there are some p -values greater than 1. This can be explained by the formula given by Peacock (1983), the author who designed this two-dimensional Kolmogorov-Smirnov test. Peacock's procedure was to:

- i) Obtain D_n , the maximum difference for each point, where the expected frequency and observed frequency are compared in four quadrants.
- ii) Calculate Z_n by $Z_n = \sqrt{n}D_n$.

iii) Calculate Z_∞ by $Z_\infty = \frac{Z_n}{1 - 0.53n^{-0.9}}$.

iv) P -value is given by: $P(> Z_\infty) = 2 \exp[-2(Z_\infty - 0.5)^2]$.

The procedure can be expressed as a function of p -value versus D_n :

$$p = 2 \exp \left[-2 \left(\frac{\sqrt{n} D_n}{1 - 0.53n^{-0.9}} - 0.5 \right)^2 \right]. \quad (3.26)$$

If $n = 1000$ is assumed, the following plot is obtained:

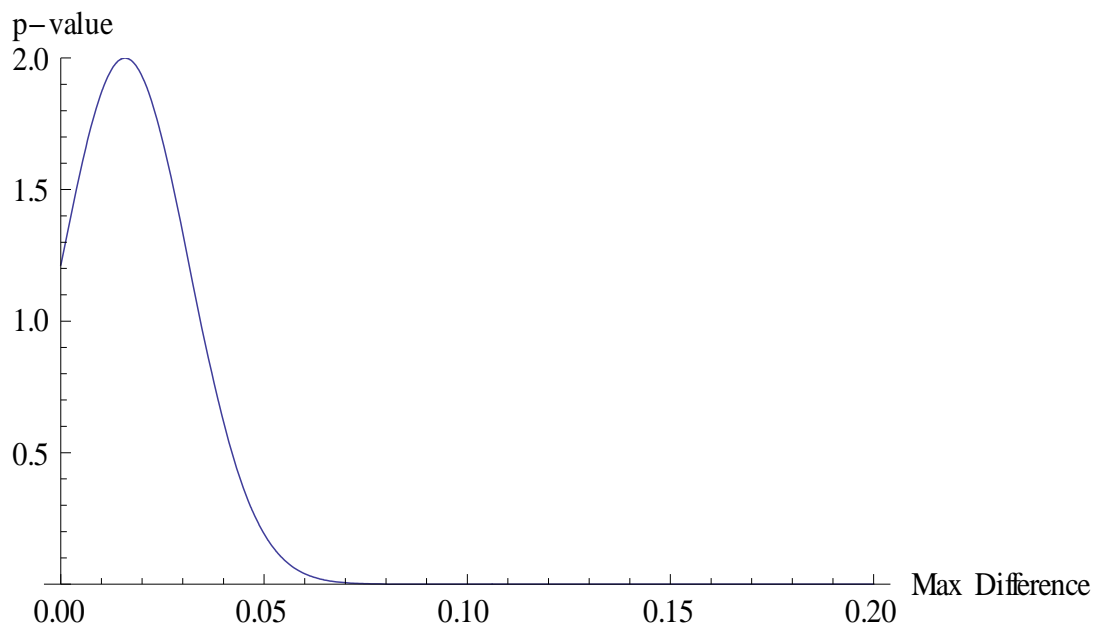


Table 3-15 Relationship between maximum difference and p -value for the 2D K-S test

It is obvious that if the maximum difference is small, the p -value given by Peacock's two-dimensional Kolmogorov-Smirnov test may give a value greater than 1, which has already been mentioned in Peacock (1983).

Based on the p -values obtained from the Chi-square goodness-of-fit test and the two-dimensional K-S test, the two tests do not agree with each other in seventeen out of one hundred and thirty five sets of estimates. A smaller p -value is found in the Chi-square goodness-of-fit test in all but two of these contradicting cases. In fact, there are only eighteen occasions such that the Chi-square goodness-of-fit test has a larger p -value than Peacock's 2D K-S test. This may be explained by the formulation of Peacock's 2D K-S test and Chi-square goodness-of-test. The formulation of Peacock's test has just been discussed previously in the preceding paragraph.

Peacock's test requires the maximum difference found between the expected and observed frequencies for all points and this maximum difference is then used to calculate the p -value. In Chi-square goodness-of-fit test, the expected and observed frequencies are compared in cells formed by the statistician. Thus, the statistician has to apply his or her own judgment in deciding how many cells are required. As discussed, since cells with equal size are preferred, the number of cells for a triangular domain would be 9, 16, 25, etc. Unfortunately, little work has been done on selecting cells for the Chi-square goodness-of-test in a bivariate distribution. 9 cells of equal size are adopted in this study to make sure merging of cells is not likely. This may affect the p -value obtained by the Chi-square goodness-of-fit test

and have a bearing on the cases which different conclusion about the proposed model interpreted by the two tests.

3.7.2 Discussion on the Three Methods of Estimation

In all the tests compiled, fourteen models proposed by the percentile matching method give rise to a p -value less than 0.1, implying the proposed model being rejected. Similarly, eight models by the method of moments and four models by the maximum likelihood estimation method are rejected with p -value being less than 0.1. According to the results of the simulation exercise, the number of rejected model indicates the models proposed by the percentile matching method is more likely to be rejected. On the other hand, there are cases where the p -value obtained by the percentile matching method is greater than the other two methods. In table 3-13, trial 2, under the Chi-square goodness-of-fit test, the model suggested by the percentile matching method overweights the other two models which are not rejected with a 0.10 level of significance. This method of percentile matching outlined in this study is not matured in terms of development. More work being done in this method may give statisticians a better estimation.

3.8 Discussion and Suggestion on Simulating Techniques and Estimating Methods

This chapter has discussed three different simulating techniques and three different estimating methods. Based on the results obtained in this chapter, the gamma factors method and maximum likelihood estimators are suggested.

The gamma factors approach is suggested over the other two methods because of its simplicity. The conditional approach requires the conditional probability density function, which is a Beta distribution in this case. A selection procedure must be carried because the second random number obtained may not agree with the domain of a Dirichlet distribution. Similar matter arises with the acceptance and rejection approach. A selection criterion must be satisfied or the random variate must be simulated again. The gamma factors approach does not have these problems. Three random numbers distributed by gamma probability density functions can easily be simulated and the random variate is then generated.

In order to estimate parameters of the conjugate function, three estimating methods have been considered. The method of maximum likelihood, although without a clear

formula for the parameters, is suggested. Unlike the other two methods, maximum likelihood estimation is unique. Moreover, the experiment demonstrated in Section 3.7 shows that the method of maximum likelihood estimation accounts for the smallest number of rejected models among the three methods of estimation. For the method of moments, although the estimates and the test results are very similar to the method of maximum likelihood estimation and the calculation is comparably much easier than the method of maximum likelihood estimation, this method gives a set of estimates which are not unique, since different choice of moments gives different estimates of parameters. Similarly, different choice of regions gives different estimates of parameters for the percentile matching method. Maximum likelihood estimation is the only method which gives a unique estimation.

CHAPTER 4 INVERTED DIRICHLET LOSS FUNCTION

Spiring (1993) introduced an innovative Reflected (or Inverted) Normal Loss Function (RNLF or INLF). This loss function penalized deviation from off-targetness and provided bounded losses at the two far ends. He further built a bivariate RNLF to depict financial loss due to deviations of two quality characteristics from their respective target values to measure monetary losses to the society as well as the manufacturer. In this section, the IbNLF of Spiring would be revisited and a loss function by inverting the Dirichlet distribution would be developed.

The Dirichlet distribution, due to its flexibility in including variables, is an ideal study target to begin with in developing a family of loss functions which are multivariate. Unlike other distributions, the Dirichlet distribution can take on as many variables as required. Its characteristic enables quality assurance professionals to conveniently build loss functions to evaluate financial loss when multiple variables are considered.

The Dirichlet distribution is additive in nature, namely, the domain is defined as long as $0 \leq x_1 + x_2 + \dots + x_k \leq 1$, where x_1, x_2, \dots, x_k are measurable process characteristics. This is a valuable characteristic when the sum of measurements of k variables must be less than or equal to a limit. This distribution may also be applied, for example, to depict the financial loss of transportation from a wholesaler to k retailers, within a specific amount of time. Applications will be demonstrated in later sections of this chapter to illustrate the feasibility of such loss function developed from the Dirichlet distribution.

4.1 Spiring's Inverted bivariate Normal Loss Function

Spiring (1993) introduced the bivariate form of his Inverted Normal Loss Function to show the possibility of projecting financial loss when more than one variable was being considered. The general form of this bivariate loss function was given as:

$$L(\mathbf{y}) = K \left\{ 1 - \exp \left[-(\mathbf{y} - \mathbf{T})^T \Gamma^{-1} (\mathbf{y} - \mathbf{T}) \right] \right\}, \quad (4.1)$$

where $\mathbf{y} = \begin{bmatrix} x \\ y \end{bmatrix}$ represents the vector of quality characteristics being studied, K is the maximum loss due to the product of a manufacturing process, $\mathbf{T} = \begin{bmatrix} T_x \\ T_y \end{bmatrix}$ is the

associated vector of target values and $\Gamma = \begin{bmatrix} \gamma_1 & \gamma_{12} \\ \gamma_{12} & \gamma_2 \end{bmatrix}$, where $\gamma_1 = \frac{\Delta_1}{4}$, $\gamma_2 = \frac{\Delta_2}{4}$ and

$\gamma_{12} = \frac{\beta\Delta_2^2}{16}$ represents the shape parameter as presented in Spiring's initial INLF. Δ_1 and Δ_2 are the Euclidean distance measured from the target value to the point where maximum loss occurs in the principle and secondary axis respectively, β is the slope of the principle axis.

4.1.1 Procedure of Building a Bivariate INLF

The product engineer specified the target value and tolerance limits of a product based on his expertise and experience. These are the main ingredients to build an IbNLF. Spiring (1993) estimated the shape parameters by approximation and came up with $\gamma_1 = \frac{\Delta_1}{4}$, $\gamma_2 = \frac{\Delta_2}{4}$ and $\gamma_{12} = \frac{\beta\Delta_2^2}{16}$, where Δ_1 and Δ_2 are the Euclidean distance measured from the target value to the point where maximum loss occurs in the principle and secondary axis respectively and γ_1 , γ_2 , and γ_{12} represent the standard deviations of the two variables and the correlation between the two variables respectively. The use of distance between target value and maximum loss occurrence was first suggested by Taguchi, Elsayed and Hsiang (1989) as an

approximation. β is the slope of the principle axis, which has to be estimated by evaluating the slope of the dominating quality characteristic. Problems may arise when both quality characteristics are equally important and thus the slope measurement may not be unique. In view of these probable problems, an alternative is considered in this thesis.

Consider a case when the financial loss of a product can only be affected by two quality characteristics. A product engineer specifies the target values of the two qualities and four tolerance limits. These tolerance limits form a rectangular region, while the Normal distribution has elliptical contours and the Dirichlet distribution has a triangular domain. In other words, an inverted probability loss function built by either a Normal distribution or a Dirichlet distribution will either over-estimate or under-estimate the financial loss. Over-estimation may cause less problems than under-estimation because high quality of products may give an unreasonably high depicted loss, which may lead to the decision of upgrading production lines by mistake. However, under-estimating the financial loss may lead to low quality products having a reasonable loss depicted and affects the reputation of the company, eventually lower consumer demand and drives the company out of business.

As the discussion extends to the bivariate (or multi-variate case) case, unfortunately, information about the correlation of the two (or more) variables is often unavailable. The lack of information about correlation creates problems, especially when an Inverted bivariate Normal loss function is desired. Data collected from samples may give some insight about the actual correlation, but nonetheless, the estimation of correlation between the two variables, requires assumptions in most occasions. Three possible scenarios: no correlation, negative correlation and positive correlation, for which values close to positive and negative one will be considered.

Recall the bivariate Normal density function:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \quad (4.2)$$

Based on the bivariate Normal distribution, the mode can be calculated as follows:

$$\begin{aligned}
f(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \\
\ln f &= \ln \left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \right) - \frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \\
\frac{\partial \ln f}{\partial x} &= -\frac{1}{2(1-\rho^2)} \left[2 \left(\frac{x-\mu_x}{\sigma_x} \right) \frac{1}{\sigma_x} - 2\rho \left(\frac{y-\mu_y}{\sigma_y} \right) \frac{1}{\sigma_x} \right] = 0 \\
\frac{\partial \ln f}{\partial y} &= -\frac{1}{2(1-\rho^2)} \left[2 \left(\frac{y-\mu_y}{\sigma_y} \right) \frac{1}{\sigma_y} - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \frac{1}{\sigma_y} \right] = 0 \\
x^* &= \mu_x = T_x \\
y^* &= \mu_y = T_y.
\end{aligned} \tag{4.3}$$

From equation (4.3), the mode of a bivariate Normal distribution is its mean, and it is selected as the target values. Hence, the maximum value of the function is:

$$\begin{aligned}
\pi(x, y, T_x, T_y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \\
&\quad \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \\
m(T_x, T_y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \\
&\quad \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{\mu_x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{\mu_y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{\mu_x-\mu_x}{\sigma_x} \right) \left(\frac{\mu_y-\mu_y}{\sigma_y} \right) \right] \right\} \\
&= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}.
\end{aligned} \tag{4.4}$$

Using the result of equation(4.4), the general form of an IbNLF was obtained:

$$\begin{aligned}
L(x, y) &= K \left\{ 1 - \frac{\pi(x, y, T_x, T_y)}{m(T_x, T_y)} \right\} \\
&= K \left[1 - \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \right].
\end{aligned}
\tag{4.5}$$

Due to the problem of lack of information provided for σ_x, σ_y and ρ , the parameters of the IbNLF, estimations of σ_x, σ_y and different scenarios of values of ρ will be considered.

4.1.1.1 No Correlation

If there is no correlation ($\rho = 0$) between the two variables, as is the case that the two quality characteristics are independent. From (4.2) the IbNLF is:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\}.$$

Assuming the target value is set

at the middle of the tolerance limits and UL_x and UL_y represent the upper limits of the two variables respectively. σ_x and σ_y can be picked from the set where

$$g(\sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[\left(\frac{x-UL_x}{\sigma_x} \right)^2 + \left(\frac{y-UL_y}{\sigma_y} \right)^2 \right] \right\} = 0$$

is satisfied.

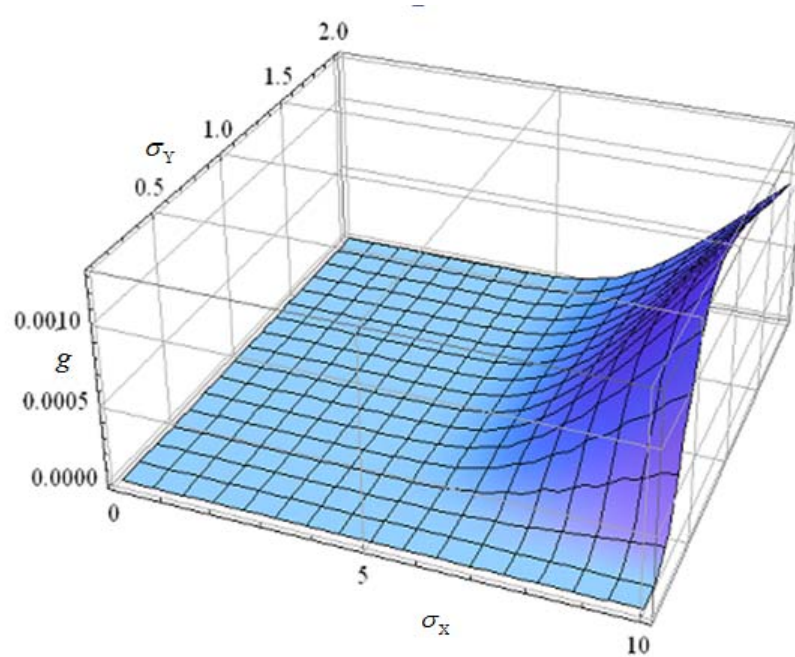


Figure 4-1 $g(\sigma_x, \sigma_y)$ with $UL_x=10$, $UL_y=2$

Although values of x and y in a normal distribution never assume the values of 0, arbitrarily small values, for example, 0.0001, can be determined in advance, and the limits will include these arbitrarily small values.

Figure 4-1 above illustrates the function $g(\sigma_x, \sigma_y)$. According to this plot for specific values, any combination of (σ_x, σ_y) , which lay on the xy -plane can be used as the scale parameters of the IbNLF for this particular example. A combination like (4, 1) may be favorable since: 1) it lies on the xy -plane and 2) it lies very close to the edge of the curly surface given by $g(\sigma_x, \sigma_y)$. A combination lying on the xy -

plane is preferable because the probability density is very close to zero, which would result in the maximum financial loss by the IPLF. Being close to the curly surface is favorable because In this sense, the elliptical opening for the IbNLF will give the maximum unit loss right on the tolerance limits given by the engineer.

4.1.1.2 Positive or Negative Correlation

Similar to the previous subsection, based on the tolerance limits given by the engineer, a function based on σ_x , σ_y and ρ can be obtained:

$$g(\sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-UL_x}{\sigma_x}\right)^2 + \left(\frac{y-UL_y}{\sigma_y}\right)^2 - \frac{2\rho(x-UL_x)(y-UL_y)}{\sigma_x\sigma_y}\right]\right\},$$

where ρ is an assumed value lying between -1 and 1. In this project, values of ± 0.95 are used to provide extreme cases.

The loss function of a k -variate Normal distribution can be established as follows:

$$f_{\mathbf{x}}(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \Sigma^{-1}(\mathbf{x} - \mathbf{m})\right\}, \quad (4.6)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_k]^T$, $\mathbf{m} = [\mu_1, \mu_2, \dots, \mu_k]^T$, and Σ is the variance-covariance

matrix, given by $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{2k} & \cdots & \sigma_k^2 \end{bmatrix}$.

It is known that the mode happens to be the mean vector of normal distribution, thus,

$[T_{X_1} \ T_{X_2} \ \dots \ T_{X_k}]^T = [\mu_{X_1} \ \mu_{X_2} \ \dots \ \mu_{X_k}]^T$, and the supremum is found at:

$$\begin{aligned}
 m &= \sup_x \pi(\mathbf{x}, \mathbf{T}) \\
 &= \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{m} - \mathbf{m})^T \Sigma^{-1}(\mathbf{m} - \mathbf{m})\right\} \\
 &= \left(\frac{1}{2\pi}\right)^{k/2} \frac{1}{|\Sigma|^{1/2}}.
 \end{aligned} \tag{4.7}$$

The loss function is then,

$$\begin{aligned}
 L(x, y) &= K \left\{1 - \frac{\pi(\mathbf{x}, \mathbf{T})}{m(\mathbf{T})}\right\} \\
 &= K \left[1 - \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \Sigma^{-1}(\mathbf{x} - \mathbf{m})\right\}\right].
 \end{aligned} \tag{4.8}$$

Figure 4-2 to Figure 4-5 are some example plots of the IbNLF with various variances

and correlation factors:

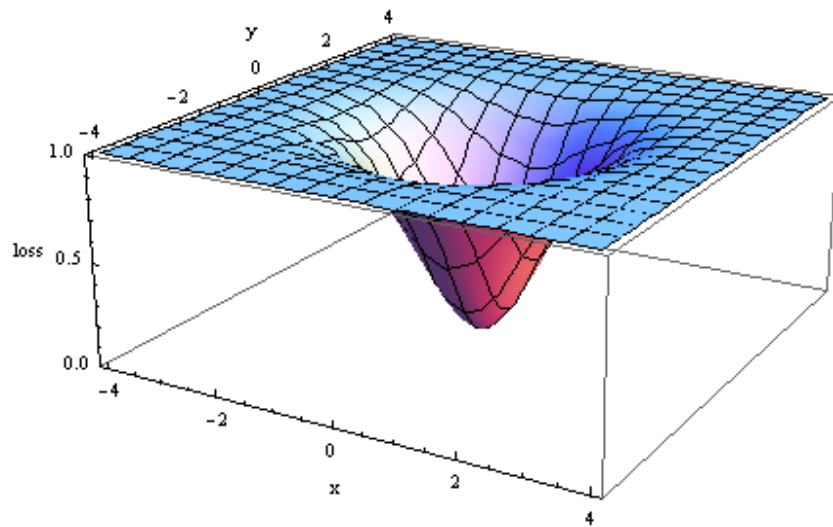


Figure 4-2 IbNLF with $T_x = T_y = 0, \sigma_x = \sigma_y = 1, \rho = 0$

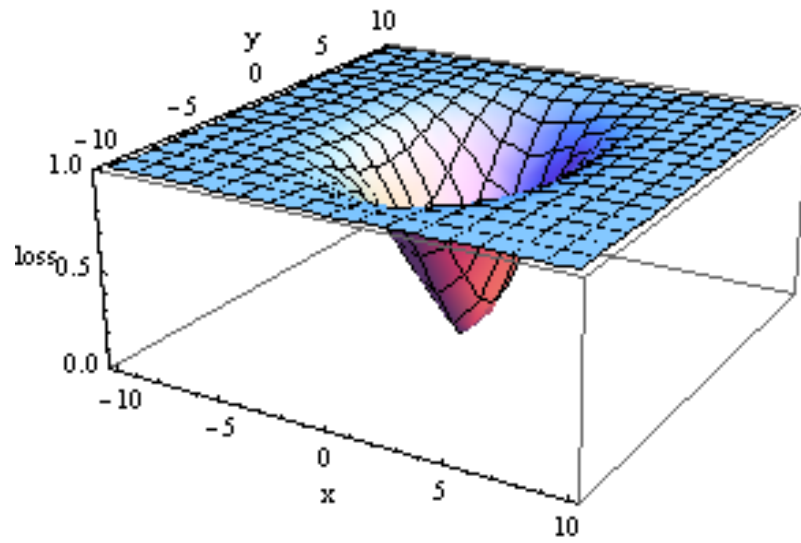


Figure 4-3 IbNLF with $T_x = T_y = 0, \sigma_x = 2, \sigma_y = 3, \rho = 0$

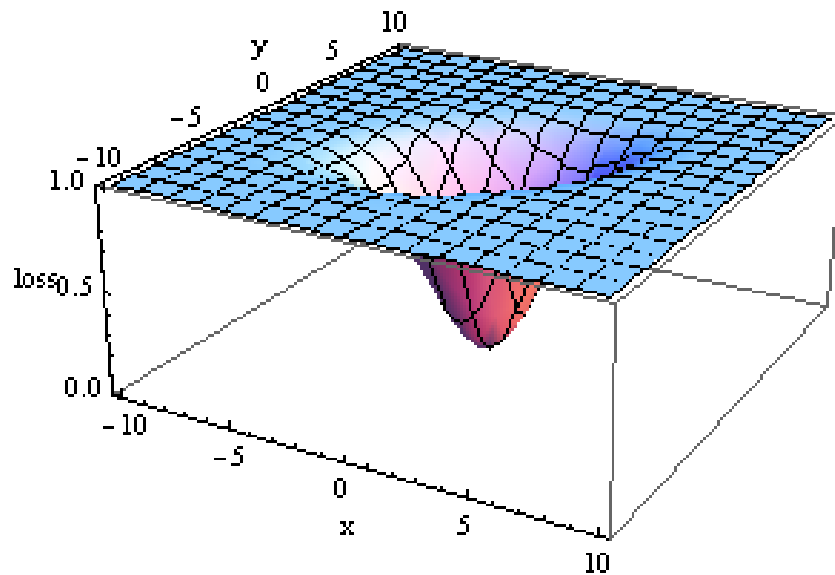


Figure 4-4 IbNLF with $T_x = T_y = 0, \sigma_x = \sigma_y = 2, \rho = 0.5$

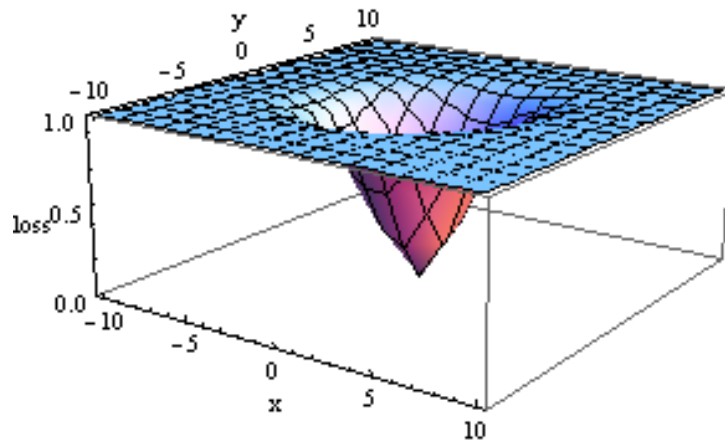


Figure 4-5 IbNLF with $T_x = T_y = 0, \sigma_x = \sigma_y = 2, \rho = 0.25$

From the above plots, one can assure that the original assumption of Spiring (1993) is indeed correct. By letting $\sigma = \frac{\Delta}{4}$, or $\Delta = 4\sigma$, given any σ , the distance between the target and where maximum loss occurs can be predicted easily.

IbNLFs and their probable correlation can be illustrated as in the contour plots (Figure 4-6):

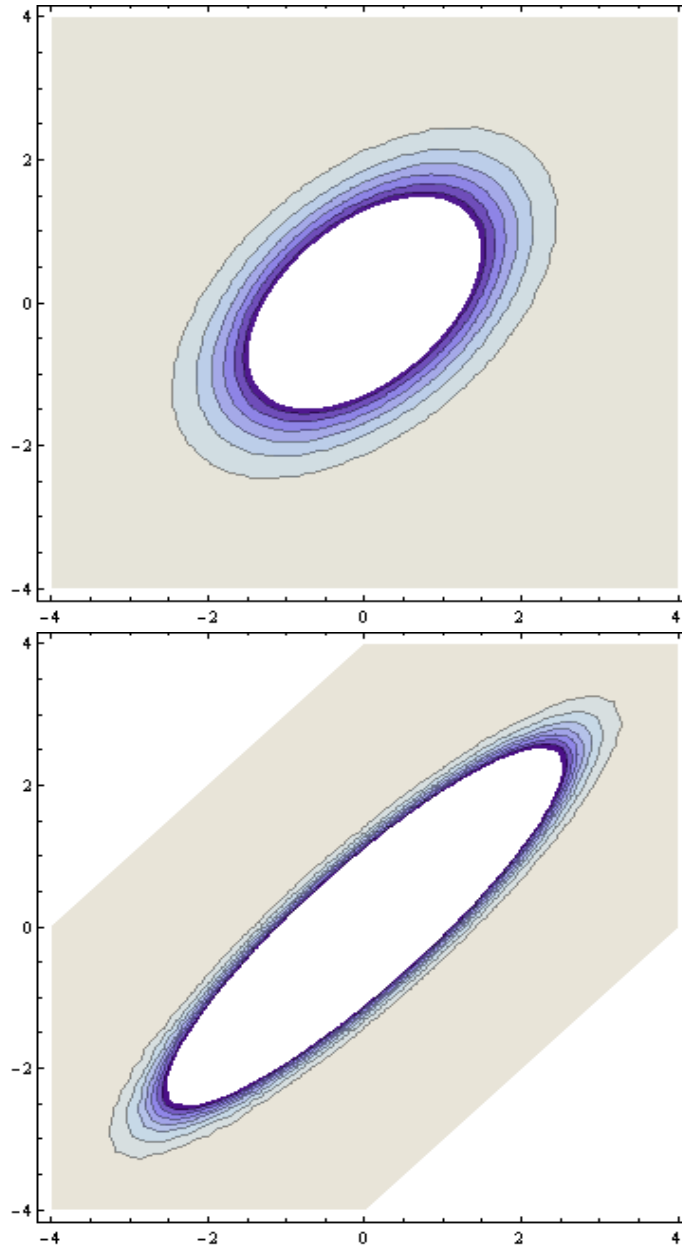


Figure 4-6 Contour plot of two different IbNLFs (above: $\rho = 0.5$, below: $\rho = 0.9$)

Both plots have the target value at the origin and variance of 2 in both axes of X and Y . The only difference is the correlation, $\rho = 0.5$ and $\rho = 0.9$ for the two contour plots respectively. When $\rho = 0$, in other words, the two variables of X and Y are independent of each other, and if variances of the two axes are the same, the contour

plot will have the same form of a circle as the IbNLF plotted as in Figure 4-2 and Figure 4-3. Increment of correlation increases the length of the oval along the line of $x = y$, where, different variances will result in the increment of length in the corresponding axis.

In Chapter 5, estimation of parameters of two IbNLF will be utilized with real data.

Detail illustration with the idea outlined in this subsection will be demonstrated.

4.1.2 Relationship of a Bivariate INLF and its Marginal Loss Functions

The relationship between two univariate probability density functions can be measured by its covariance. Consider a bivariate normal distribution with the following form:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 - 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right] \right\} \quad (4.9)$$

ρ is the correlation between X and Y , where X and Y are random variables distributed marginally by univariate normal distributions. Correlation describes the dependency of X and Y and their relationship: positively or negatively correlated. However, ρ alone cannot explain the relationship between loss functions.

Hoeffding (1940) showed that the covariance of two random variables X and Y could be written in terms of their joint distribution and respective marginal distributions in the following form:

$$\text{cov}(X, Y) = \int_{R^2} (H(x, y) - F(x)G(y)) dx dy \quad (4.10)$$

where $H(x, y)$ is the continuous joint cumulative density function, $F(x)$ and $G(y)$ are the marginal cumulative density functions, and R^2 is the two dimensional domain which the function is defined.

Expanding on Hoeffding's work, Sen (1994) found a formula of calculating the covariance between two functions and Cuadras (2002) gave a detailed proof of the following generalization:

$$\text{cov}(\alpha(X), \beta(Y)) = \int_{R^2} (H(x, y) - F(x)G(y)) d\alpha(x) d\beta(y) \quad (4.11)$$

$\alpha(x)$ and $\beta(y)$ are two functions defined on $[a, b]$, $[c, d]$ respectively. Under the assumptions of:

1. Both functions are of bounded variation, and
2. $E(|\alpha(X)\beta(Y)|)$, $E(|\alpha(X)|)$ and $E(|\beta(Y)|)$ all finite.

Then, the covariance function can be rewritten as:

$$\text{cov}(\alpha(X), \beta(Y)) = \int_a^b \int_c^d (H(x, y) - F(x)G(y)) d\alpha(x) d\beta(y) \quad (4.12)$$

Consider a bivariate Normal distribution, with marginal distributions of two univariate Normal distribution. The cumulative density function (cdf) of an

univariate normal distribution could be written as $F(x) = \frac{1}{2} \left(1 + \text{erf} \left(\frac{x - \mu}{\sqrt{2}\sigma} \right) \right)$, where

$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ or simply known as the error function encountered in

integrating the normal distribution. However, no neat expression of a bivariate normal cumulative density function can be formed. Using the integral form and substituting into Sen and Cuadras' formula gives:

$$\begin{aligned}
H(x, y) &= \int_{-\infty}^y \int_{-\infty}^x \left\{ \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \right. \\
&\quad \left. \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{t-\mu_x}{\sigma_x} \right)^2 + \left(\frac{u-\mu_y}{\sigma_y} \right)^2 - 2\rho \frac{t-\mu_x}{\sigma_x} \frac{u-\mu_y}{\sigma_y} \right\} \right] \right\} dt du \\
F(x) &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp \left[-\frac{1}{2} \left(\frac{t-\mu_x}{\sigma_x} \right)^2 \right] dt \\
G(y) &= \int_{-\infty}^y \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp \left[-\frac{1}{2} \left(\frac{u-\mu_y}{\sigma_y} \right)^2 \right] du \\
F(x)G(y) &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp \left[-\frac{1}{2} \left(\frac{t-\mu_x}{\sigma_x} \right)^2 \right] dt \int_{-\infty}^y \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp \left[-\frac{1}{2} \left(\frac{u-\mu_y}{\sigma_y} \right)^2 \right] du \\
&= \int_{-\infty}^y \int_{-\infty}^x \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left[\left(\frac{t-\mu_x}{\sigma_x} \right)^2 + \left(\frac{u-\mu_y}{\sigma_y} \right)^2 \right] \right\} dt du.
\end{aligned} \tag{4.13}$$

The associated loss functions and their derivatives are:

$$\begin{aligned}
\alpha(x) = L(x) &= K \left\{ 1 - \exp \left[-\frac{1}{2} \left(\frac{x-\mu_x}{\sigma_x} \right)^2 \right] \right\}, \\
\beta(y) = L(y) &= K \left\{ 1 - \exp \left[-\frac{1}{2} \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\}, \\
d\alpha(x) &= \frac{K}{\sigma_x} \exp \left[-\frac{1}{2} \left(\frac{x-\mu_x}{\sigma_x} \right)^2 \right] dx, \\
d\beta(y) &= \frac{K}{\sigma_y} \exp \left[-\frac{1}{2} \left(\frac{y-\mu_y}{\sigma_y} \right)^2 \right] dy.
\end{aligned} \tag{4.14}$$

Putting all together into Sen and Cuadras formula, the covariance of the two IbNLFs

can be illustrated as:

$$\begin{aligned} & \text{cov}(L(X), L(Y)) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (H(x, y) - F(x)G(y)) \frac{K^2}{\sigma_X \sigma_Y} \exp \left\{ -\frac{1}{2} \left[\left(\frac{x - \mu_X}{\sigma_X} \right)^2 + \left(\frac{y - \mu_Y}{\sigma_Y} \right)^2 \right] \right\} (4.15) \end{aligned}$$

where $H(x, y)$, $F(x)$ and $G(y)$ are defined previously.

4.2 Inverted Dirichlet Loss Function

Extending the IBLF developed by Leung and Spiring (2002), this study explores the possibility of developing a multivariate IBLF based on a standard Dirichlet distribution, with parameters of $\alpha_1, \alpha_2, \dots, \alpha_k$. A simpler case of two variables will be presented in section 4.2.1 after developing a general case of k variables here.

$$\begin{aligned}
f_{x_1, x_2, \dots, x_k} (x_1, x_2, \dots, x_k) &= \frac{\Gamma(v_1 + v_2 + \dots + v_k + v_{k+1})}{\Gamma(v_1)\Gamma(v_2)\cdots\Gamma(v_k)\Gamma(v_{k+1})} \times \\
&\quad x_1^{v_1-1} x_2^{v_2-1} \cdots x_k^{v_k-1} \left(1 - \sum_{i=1}^k x_i\right)^{v_{k+1}-1} \\
\ln f &= C + (v_1 - 1)\ln x_1 + (v_2 - 1)\ln x_2 + \cdots \\
&\quad + (v_k - 1)\ln x_k + (v_{k+1} - 1)\ln \left(1 - \sum_{i=1}^k x_i\right) \\
\frac{\partial \ln f}{\partial x_1} &= \frac{v_1 - 1}{x_1} - \frac{v_{k+1} - 1}{1 - \sum_{i=1}^k x_i} = 0 \\
\frac{\partial \ln f}{\partial x_2} &= \frac{v_2 - 1}{x_2} - \frac{v_{k+1} - 1}{1 - \sum_{i=1}^k x_i} = 0 \\
&\quad \vdots \\
\frac{\partial \ln f}{\partial x_k} &= \frac{v_k - 1}{x_k} - \frac{v_{k+1} - 1}{1 - \sum_{i=1}^k x_i} = 0 \\
T_{X_1} &= \frac{v_1 - 1}{\sum_{i=1}^{k+1} v_i - (k+1)} \\
T_{X_2} &= \frac{v_2 - 1}{\sum_{i=1}^{k+1} v_i - (k+1)} \\
&\quad \vdots \\
T_{X_k} &= \frac{v_k - 1}{\sum_{i=1}^{k+1} v_i - (k+1)}
\end{aligned} \tag{4.16}$$

The modes, with respect to each dimension, are derived as in (4.16) above. The supremum (m), loss inversion ratio (Π) and the loss function (L), can be obtained as follows:

$$\begin{aligned}
m &= \frac{\Gamma\left(\sum_{i=1}^{k+1} v_i\right)}{\prod_{i=1}^{k+1} \Gamma(v_i)} T_{X_1}^{v_1-1} T_{X_2}^{v_2-1} \dots T_{X_k}^{v_k-1} \left(1 - \sum_{i=1}^k T_{X_i}\right)^{v_{k+1}-1} \\
\Pi &= \frac{f}{m} = \left(\frac{x_1}{T_{X_1}}\right)^{v_1-1} \left(\frac{x_2}{T_{X_2}}\right)^{v_2-1} \dots \left(\frac{x_k}{T_{X_k}}\right)^{v_k-1} \left(\frac{1 - \sum_{i=1}^k x_i}{1 - \sum_{i=1}^k T_{X_i}}\right)^{v_{k+1}-1} \\
L &= K \left[1 - \left(\frac{x_1}{T_{X_1}}\right)^{v_1-1} \left(\frac{x_2}{T_{X_2}}\right)^{v_2-1} \dots \left(\frac{x_k}{T_{X_k}}\right)^{v_k-1} \left(\frac{1 - \sum_{i=1}^k x_i}{1 - \sum_{i=1}^k T_{X_i}}\right)^{v_{k+1}-1} \right] \\
&= K \left[1 - C x_1^{v_1-1} x_2^{v_2-1} \dots x_k^{v_k-1} \left(1 - \sum_{i=1}^k x_i\right)^{v_{k+1}-1} \right].
\end{aligned} \tag{4.17}$$

where $C = T_{X_1}^{1-v_1} T_{X_2}^{1-v_2} \dots T_{X_k}^{1-v_k} \left(1 - \sum_{i=1}^k T_{X_i}\right)^{1-v_{k+1}}$.

Expected value and variance can be found by the usual formula in statistics. In the derivation below, the conjugate function is assumed to be a standard Dirichlet distribution with parameters $a_1, a_2, \dots, a_k, a_{k+1}$.

$$\begin{aligned}
E[L] &= \int_0^1 \cdots \int_0^{1-\sum_{i=3}^k x_i} \int_0^{1-\sum_{i=2}^k x_i} K \left[1 - C x_1^{v_1-1} x_2^{v_2-1} \cdots x_k^{v_k-1} \left(1 - \sum_{i=1}^k x_i \right)^{v_{k+1}-1} \right] \frac{\Gamma\left(\sum_{i=1}^{k+1} a_i\right)}{\prod_{i=1}^{k+1} \Gamma(a_i)} \times \\
&\quad x_1^{a_1-1} x_2^{a_2-1} \cdots x_k^{a_k-1} \left(1 - \sum_{i=1}^k x_i \right)^{a_{k+1}-1} dx_1 dx_2 \cdots dx_k \\
&= K - KC \frac{\Gamma\left(\sum_{i=1}^{k+1} a_i\right)}{\prod_{i=1}^{k+1} \Gamma(a_i)} \times \\
&\quad \int_0^1 \cdots \int_0^{1-\sum_{i=3}^k x_i} \int_0^{1-\sum_{i=2}^k x_i} x_1^{a_1+v_1-2} x_2^{a_2+v_2-2} \cdots x_k^{a_k+v_k-2} \left(1 - \sum_{i=1}^k x_i \right)^{a_{k+1}+v_{k+1}-2} dx_1 dx_2 \cdots dx_k \\
&= K \left[1 - C \frac{\Gamma\left(\sum_{i=1}^{k+1} a_i\right) \prod_{i=1}^{k+1} \Gamma(a_i + v_i - 1)}{\prod_{i=1}^{k+1} \Gamma(a_i) \Gamma\left(\sum_{i=1}^{k+1} (a_i + v_i - 1)\right)} \right]. \tag{4.18}
\end{aligned}$$

where C is the constant as defined in (4.17).

4.2.1 Bivariate Inverted Dirichlet Loss Function

This subsection illustrates a specific case of $k = 2$. With the Dirichlet density

distribution, $f(x, y) = \frac{\Gamma(v_1 + v_2 + v_3)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)} x^{v_1-1} y^{v_2-1} (1-x-y)^{v_3-1}$, the modal point can be

obtained by (4.16) as:

$$\begin{aligned}
T_X &= \frac{v_1 - 1}{v_1 + v_2 + v_3 - 3}, \\
T_Y &= \frac{v_2 - 1}{v_1 + v_2 + v_3 - 3}.
\end{aligned} \tag{4.19}$$

Following the above procedure, its supremum, loss inversion ratio and loss function are obtained as follows:

$$m = \frac{\Gamma(v_1 + v_2 + v_3)}{\Gamma(v_1)\Gamma(v_2)\Gamma(v_3)} T_X^{v_1-1} T_Y^{v_2-1} (1 - T_X - T_Y)^{v_3-1}. \tag{4.20}$$

The parameters v_1 and v_2 can be expressed as $v_1 = \frac{T_Y - 1 + (2 - v_3)T_X}{T_X + T_Y - 1}$ and $v_2 = \frac{T_X - 1 + (2 - v_3)T_Y}{T_X + T_Y - 1}$. The loss inversion ratio becomes:

$$\begin{aligned}
\frac{\pi(x, y, T_X, T_Y)}{m} &= \left(\frac{x}{T_X}\right)^{v_1-1} \left(\frac{y}{T_Y}\right)^{v_2-1} \left(\frac{1-x-y}{1-T_X-T_Y}\right)^{v_3-1} \\
&= \left(\frac{x}{T_X}\right)^{(1-v_3)T_X/(T_X+T_Y-1)} \left(\frac{y}{T_Y}\right)^{(1-v_3)T_Y/(T_X+T_Y-1)} \left(\frac{x+y-1}{T_X+T_Y-1}\right)^{v_3-1} \\
&= C x^{(1-v_3)T_X/\varphi} y^{(1-v_3)T_Y/\varphi} (x+y-1)^{v_3-1}.
\end{aligned} \tag{4.21}$$

where $C = T_X^{(v_3-1)T_X/\varphi} T_Y^{(v_3-1)T_Y/\varphi} \varphi^{1-v_3}$ and $\varphi = T_X + T_Y - 1$, and the inverted bivariate beta loss

function (IbBLF):

$$L(x, y, T_X, T_Y) = K \left[1 - C x^{(1-v_3)T_X/\varphi} y^{(1-v_3)T_Y/\varphi} (x+y-1)^{v_3-1} \right]. \tag{4.22}$$

Based on the loss function, the expected loss is:

$$E(L) = K \left[1 - C \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} \frac{\Gamma(v_1 + a_1 - 1)\Gamma(v_2 + a_2 - 1)\Gamma(v_3 + a_3 - 1)}{\Gamma(v_1 + v_2 + v_3 + a_1 + a_2 + a_3 - 3)} \right] \quad (4.23)$$

where C is the expression in (4.21) and a_1, a_2, a_3 are the parameters of the conjugate function which is assumed to be a Dirichlet distribution.

Developing the IDLF based on the tolerance limits suggested by the product engineer would be difficult. Often a rectangular tolerance region is suggested by the product engineer, which does not correspond well to a triangular domain of a Dirichlet distribution with two variables. Linear transformation may be considered, but this may not work well in all cases. More discussion on the selection process of parameters in an IDLF will be utilized in Chapter 5 with real data.

4.3 Conjugate Functions of the Inverted Dirichlet Loss Function

In Section 4.2, the IDLF has been developed, along with its moments and relationships with its marginal functions. As mentioned previously, a loss function is used to depict the financial loss caused by deviation from the desired target over a

collection of data. However, the distribution of deviations or the behavior of measurements of the product may be distributed by another probability distribution, which is often denoted as the conjugate function. In this section, a few bivariate probability density functions are selected to demonstrate the feasibility of using a conjugate function other than a Dirichlet distribution.

In Leung and Spiring (2002), a set of conjugates were discussed, and they were the univariate gamma, exponential, beta and uniform distribution. In this thesis, the inverted Dirichlet distribution is also considered. Unlike the univariate probability density functions, statisticians developed a large number of multivariate probability distributions for common distributions like gamma and exponential. For example, in Kotz et al. (2000), thirteen bivariate exponential distributions were presented. Here, only one from each of the proposed distributions is being discussed based on the simplicity of that particular distribution.

Due to the complexity of multivariate distribution, most of the formulas are given in terms of double integrals. To avoid the mathematical complexity such that clearer insights can be obtained, only bivariate distributions and IDLF would be considered.

The most natural choice of conjugate would certainly be the Dirichlet distribution itself, which had already been discussed section 4.2.

4.3.1 Uniform Distribution

The simplest choice of conjugate is the Uniform distribution. The bivariate uniform distribution is defined as:

$$f(x, y) = \begin{cases} \frac{1}{(b-a)(d-c)}, & a \leq x \leq b, c \leq y \leq d \\ 0, & \text{otherwise} \end{cases} \quad (4.24)$$

If the behavior of data has a rectangular domain, the expected loss becomes:

$$\begin{aligned} & \int_c^d \int_a^b \frac{1}{(b-a)(d-c)} K \left[1 - \left(\frac{x}{T_X} \right)^{v_1-1} \left(\frac{y}{T_Y} \right)^{v_2-1} \left(\frac{1-x-y}{1-T_X-T_Y} \right)^{v_3-1} \right] dx dy \\ & = K - \frac{K}{(b-a)(d-c)} \int_c^d \int_a^b \left(\frac{x}{T_X} \right)^{v_1-1} \left(\frac{y}{T_Y} \right)^{v_2-1} \left(\frac{1-x-y}{1-T_X-T_Y} \right)^{v_3-1} dx dy. \end{aligned} \quad (4.25)$$

4.3.2 Inverted Dirichlet Distribution

Ghorbel (2010) and Tiao and Cuttman (1965) used two different approaches to derive the Inverted Dirichlet Distribution. Ghorbel (2010) considered individual random variables to be distributed by the gamma distributions. On the other hand, Tiao and Cuttman (1965) used an inverted transformation to derive the Inverted Dirichlet Distribution. Both approaches lead to the same probability density function and the general form is shown below if it has parameters $a_1, a_2, \dots, a_k, a_{k+1}$:

$$f(x_1, x_2, \dots, x_k) = \frac{\Gamma\left(\sum_{i=1}^{k+1} a_i\right)}{\prod_{i=1}^{k+1} \Gamma(a_i)} x_1^{a_1-1} x_2^{a_2-1} \dots x_k^{a_k-1} \left(1 + \sum_{i=1}^k x_i\right)^{-\sum_{i=1}^{k+1} a_i} \quad (4.26)$$

Expected loss of an IDLF with the Inverted Dirichlet distribution acting as the conjugate function could be illustrated as:

$$\begin{aligned}
E[L] &= \int_0^\infty \int_0^\infty \cdots \int_0^\infty K \left[1 - Cx_1^{v_1-1} x_2^{v_2-1} \cdots x_k^{v_k-1} \left(1 - \sum_{i=1}^k x_i \right)^{v_{k+1}-1} \right] \\
&\quad \times \frac{\Gamma\left(\sum_{i=1}^{k+1} a_i\right)}{\prod_{i=1}^{k+1} \Gamma(a_i)} x_1^{a_1-1} x_2^{a_2-1} \cdots x_k^{a_k-1} \left(1 + \sum_{i=1}^k x_i \right)^{-\sum_{i=1}^{k+1} a_i} dx_1 \cdots dx_{k-1} dx_k \\
&= K - KC \frac{\Gamma\left(\sum_{i=1}^{k+1} a_i\right)}{\prod_{i=1}^{k+1} \Gamma(a_i)} \int \int \cdots \int x_1^{a_1+v_1-2} x_2^{a_2+v_2-2} \cdots x_k^{a_k+v_k-2} \\
&\quad \times \frac{\left(1 - \sum_{i=1}^k x_i \right)^{v_{k+1}-1}}{\left(1 + \sum_{i=1}^k x_i \right)^{\sum_{i=1}^{k+1} a_i}} dx_1 \cdots dx_{k-1} dx_k.
\end{aligned}$$

If $k = 2$, the expected loss would become:

$$\begin{aligned}
E[L] &= \int_0^1 \int_0^1 \left[K \left(1 - \left(\frac{x}{T_x} \right)^{v_1-1} \left(\frac{y}{T_y} \right)^{v_2-1} \left(\frac{1-x-y}{1-T_x-T_y} \right)^{v_3-1} \right) \right] \\
&\quad \times \frac{\Gamma(a_1 + a_2 + a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)} x^{a_1-1} y^{a_2-1} (1+x+y)^{-(a_1+a_2+a_3)} dx dy.
\end{aligned} \tag{4.27}$$

To illustrate graphically, let $[a_1 \ a_2 \ a_3]^T = [v_1 \ v_2 \ v_3]^T = [2 \ 3 \ 4]^T$, a three-dimensional surface could be obtained as in Figure 4-7:

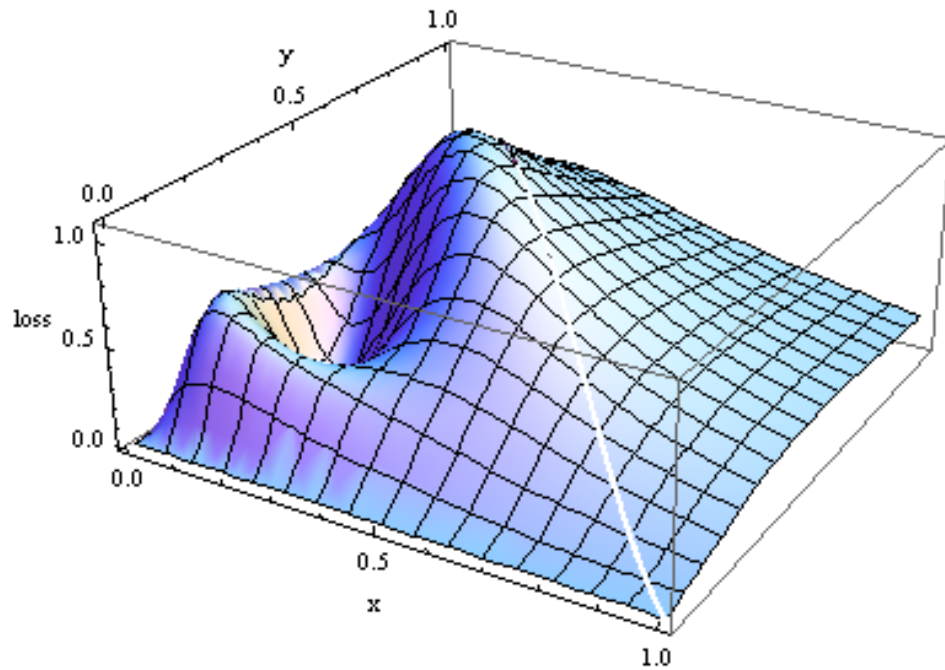


Figure 4-7 Surface of an IDDF with a conjugate distribution of Inverted Dirichlet

By built-in numerical integration method of Mathematica, an answer of 0.436552 is obtained through equation (4.27) by evaluating the volume within our desired domain.

4.3.3 Bivariate Gamma Distribution

In Kotz et al. (2000), thirteen bivariate gamma distributions were presented. In this study, the McKay's bivariate gamma is selected due to its simplicity and its common application.

McKay's bivariate gamma is defined by the following density function:

$$p(x, y) = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)} x^{a-1} (y-x)^{b-1} e^{-cy}, 0 < x < y, a, b, c > 0 \quad (4.28)$$

If $[a \ b \ c]^T = [2 \ 3 \ 4]^T$, its probability density can be illustrated as:

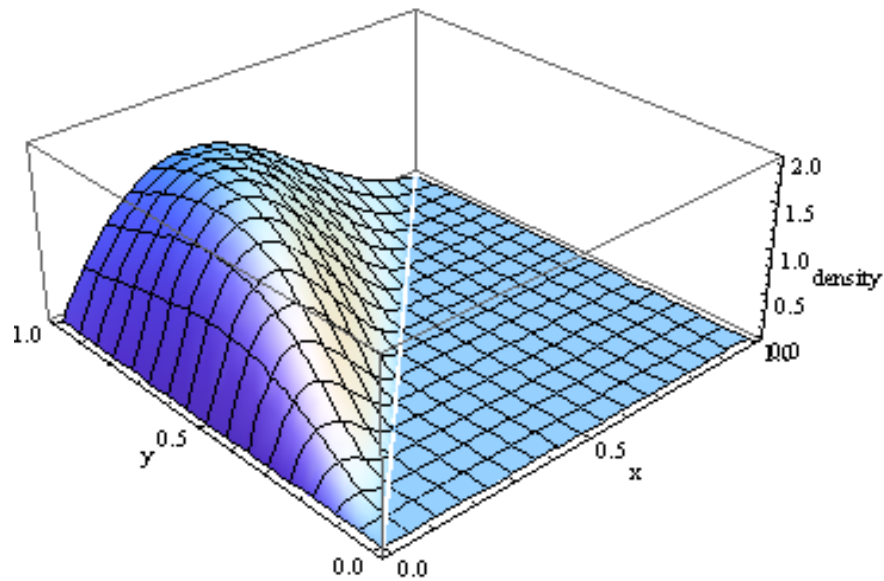


Figure 4-8 Surface of McKay's bivariate gamma[2, 3, 4]

If this probability density function is selected as the conjugate function, the general

form of the expected loss is obtained as:

$$E[L] = \frac{c^{a+b}}{\Gamma(a)\Gamma(b)} \int_0^1 \int_0^1 K \left[1 - \left(\frac{x}{T_x} \right)^{v_1-1} \left(\frac{y}{T_y} \right)^{v_2-1} \left(\frac{1-x-y}{1-T_x-T_y} \right)^{v_3-1} \right] x^{a-1} (y-x)^{b-1} e^{-cy} dx dy. \quad (4.29)$$

The above double integral, like the one shown in the Inverted Dirichlet distribution, does not have a closed analytical form. If the parameters of the IDLF are assumed as

$[v_1 \ v_2 \ v_3]^T = [2 \ 3 \ 4]^T$, the density would appear as:

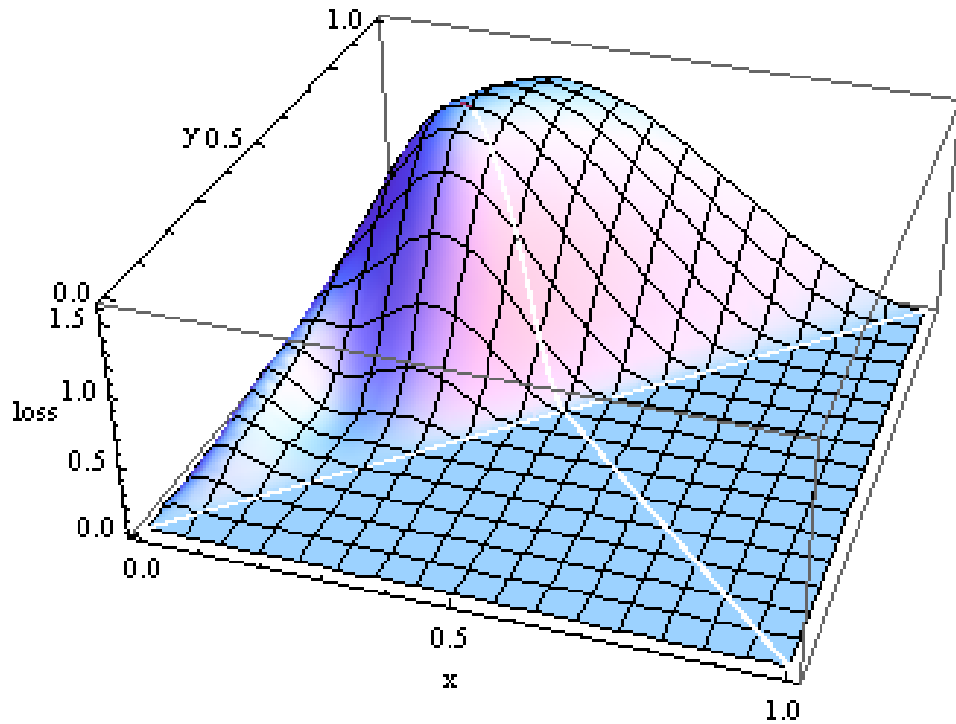


Figure 4-9 Surface of an IbDLF with a conjugate distribution of gamma[2, 3, 4]

The expected loss with the above assumptions of parameters would be 0.31439 obtained by the numerical integration function of Mathematica.

4.3.4 Bivariate Exponential Distribution

Gumbel's Bivariate Exponential distribution is being selected as one of the candidates of the conjugate of the IDLF with similar reason as McKay's bivariate gamma. The joint density function is given by:

$$p(x, y) = e^{-(x+y+\theta xy)} [(1+\theta x)(1+\theta y) - \theta], x > 0, y > 0, \theta > 0 \quad (4.30)$$

The probability density can be displayed by the following plot if $\theta = 1$:

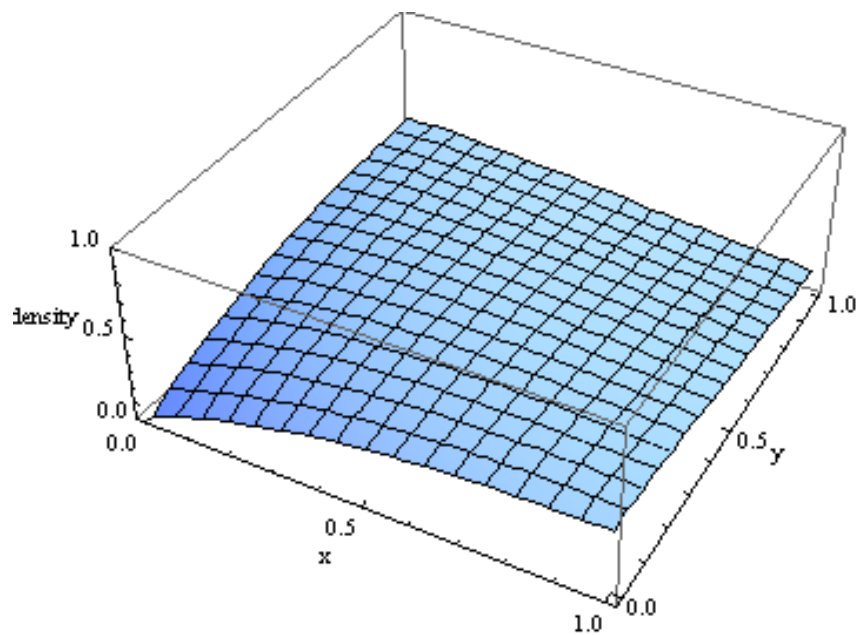


Figure 4-10 Plot of McKay's bivariate gamma distribution

The expected loss with IDLF is:

$$E[L] = \int_0^1 \int_0^1 K \left[1 - \left(\frac{x}{T_x} \right)^{v_1-1} \left(\frac{y}{T_y} \right)^{v_2-1} \left(\frac{1-x-y}{1-T_x-T_y} \right)^{v_3-1} \right] e^{-(x+y+\theta xy)} [(1+\theta x)(1+\theta y) - \theta] dx dy \quad (4.31)$$

Once again, a closed form cannot be obtained by integration. Assuming $\theta = 1, v_1 = 2, v_2 = 3, v_3 = 4$, the loss function would have the surface as followed:

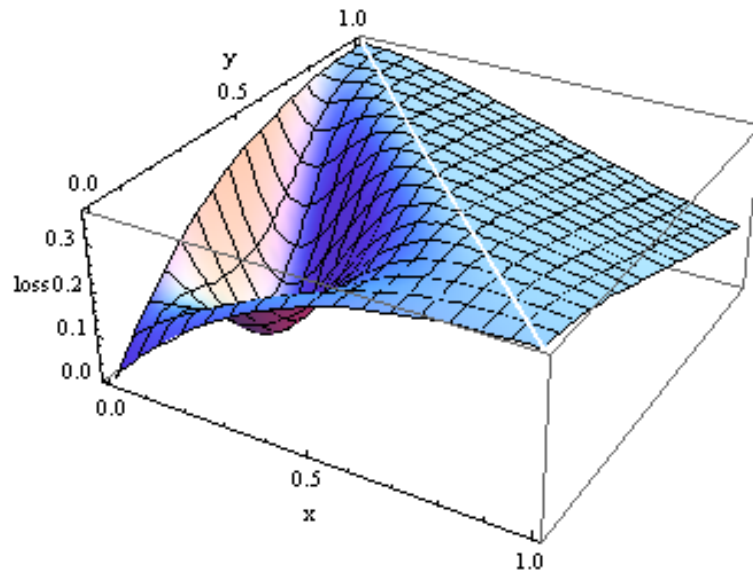


Figure 4-11 Surface of an IBDLF with a conjugate distribution of Exp[1]

Double integration based on this setting of parameters requires numerical methods to solve. For complex functions like the expected loss with bivariate inverted Dirichlet or bivariate gamma, Mathematica is able to evaluate at least one of the two integrals, while this is unlikely the case with Gumble's bivariate exponential as the conjugate

of the proposed loss function. Numerical integration gives an answer of 0.272882 for the expected loss under previously stated assumptions.

4.3.5 Bivariate Normal Distribution

The probability density function of a bivariate normal distribution has the form:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{z}{2(1-\rho^2)}\right], \quad (4.32)$$

where

$$z = \frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \quad (4.33)$$

and $\rho = \text{corr}(x, y) = \frac{V_{12}}{\sigma_1\sigma_2}$ is the correlation of x and y and V_{12} is the covariance of X

and Y .

If the covariance between x and y is assumed to be zero, the bivariate density can be

simplified to $f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_1^2}\right)\right]$. Using this

simplified form, the expected loss is:

$$E[L] = \int_0^1 \int_0^1 K \left[1 - \left(\frac{x}{T_x} \right)^{v_1-1} \left(\frac{y}{T_y} \right)^{v_2-1} \left(\frac{1-x-y}{1-T_x-T_y} \right)^{v_3-1} \right] \times \frac{1}{2\pi\sigma_1\sigma_2} \exp \left[-\frac{1}{2} \left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_1^2} \right) \right] dx dy \quad (4.34)$$

Unfortunately, similar to the case to bivariate gamma, and exponential distributions, a closed form cannot be reached. If the parameters of the above expression are assumed to be:

$$[\mu_1 \ \mu_2 \ \sigma_1^2 \ \sigma_2^2]^T = [0.5 \ 0.5 \ 1 \ 1]^T \text{ and } [v_1 \ v_2 \ v_3]^T = [2 \ 3 \ 4]^T,$$

Plots of the bivariate normal distribution and the loss function obtained were:

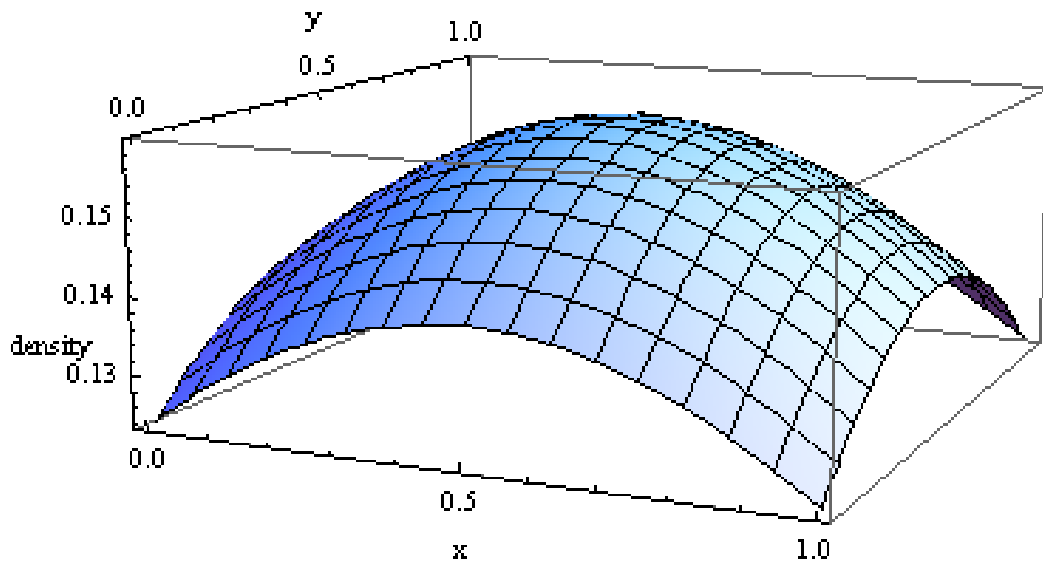


Figure 4-12 Plot of a bivariate Normal Distribution

The bivariate normal probability density function together with the assumed parameters in the IDLF has the following surface:

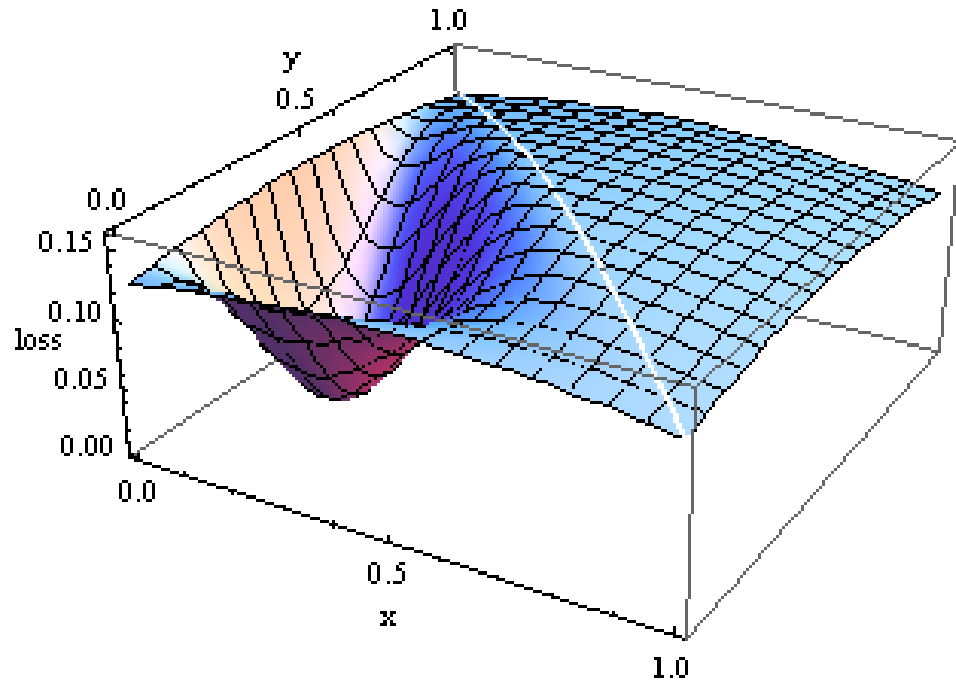


Figure 4-13 Surface of an IDLF with a conjugate distribution of Normal[.5, .5, 1, 1]

A value of 0.127583 for the expected loss with the set of assumed parameters is obtained using Mathematica.

4.3.6 Section Summary

Different bivariate probability density functions have been selected as the conjugate function in order to evaluate the expected loss when an IDLF is applied. These pdfs represent the behavior of process data. Although the financial loss may be depicted efficiently by an IDLF, the process data may be distributed by another probability density function. In other words, to accurately depict the financial loss, a statistician may need to consider other probability density function to describe the behavior of process data.

General closed form for expected loss usually cannot be found due to the complexity of the bivariate densities with different parameters. However, with the advanced computing power, as long as parameters can be numerically estimated, expected loss, variance and covariance can be calculated by numerical integration with computing software.

4.4 Properties of Multivariate Inverted Probability Loss Functions

Leung and Spiring (2004) developed a list of associated properties of the loss inverted ratio and inverted probability loss functions, which were covered in Section 2.7 of this thesis. In this section, investigation will be done to determine whether these properties are also legitimate for higher dimension loss functions.

4.4.1 Properties of the Loss Inversion Ratio (LIR)

The Loss Inversion Ratio (LIR) is defined as follow. Let X be the quality of interest of a process. Let $\pi(x, T)$ be the probability density function of X in terms of the target value, T and m is the maximum of $f_x(x)$. Then the ratio $\frac{\pi(x, T)}{m}$ is the percentage of the x values that are missed with respect to the target, and it is called the Loss Inversion Ratio. It is bounded between 0 and 1. The percentage loss $1 - \frac{\pi(x, T)}{m}$ is the penalty it pays subject to the maximum loss incurred in the process.

Properties of the LIR under the univariate case stated by Leung and Spiring (2004)

are listed as follows:

1. The LIR represents the relative proportional loss associated with any x value and hence $1 - \pi(x, T) / m$ represents the proportional penalty paid at any value of x .
2. The r -th moment of the LIR is bounded by 0 and 1. The expectation equals to 1 if $r = 0$.
3. The variance of the LIR is:

$$V[\pi(X, T) / m] = E\{[\pi(X, T) / m]^2\} - \{E[\pi(X, T) / m]\}^2,$$

and the variance shown above is bounded by 0 and 1.

For multivariate inverted probability loss functions, due to the fact that more variables are included, the loss inversion ratio becomes slightly more compact. If \underline{x} represent values within the domain of some qualities of a process which the manufacturer desires to depict loss from and \underline{T} represent the target values of those qualities. The multivariate LIR (mLIR) becomes $\pi(\underline{x}, \underline{T}) / m$ and the properties showed by Leung and Spiring (2004) can be extended as:

1. The LIR represents the relative proportional loss associated with any set of values \underline{x} and hence $1 - \pi(\underline{x}, \underline{T})/m$ represents the proportional penalty paid at this specific set of values \underline{x} .
2. The r -th moment of the LIR is bounded by 0 and 1. The expectation equals to 1 if $r=0$.
3. The variance of the loss inversion ratio is $V[\pi(\underline{X}, \underline{T})/m] = E\{[\pi(\underline{X}, \underline{T})/m]^2\} - \{E[\pi(\underline{X}, \underline{T})/m]\}^2$, and it is bounded between 0 and 1.

Proof:

1. Since $\pi(\underline{x}, \underline{T}) \leq m$, hence $0 \leq 1 - \frac{\pi(\underline{x}, \underline{T})}{m} \leq 1$.

2. With the r -th moment:

$$0 \leq \pi(\underline{x}, \underline{T})/m \leq 1$$

$$0 \leq [\pi(\underline{x}, \underline{T})/m]^r \leq 1$$

$$0 \leq \int_R [\pi(\underline{x}, \underline{T})/m]^r f_{\underline{X}}(\underline{x}) d\underline{x} \leq \int_R f_{\underline{X}}(\underline{x}) d\underline{x}$$

$$0 \leq E\{[\pi(\underline{x}, \underline{T})/m]^r\} \leq 1$$

$$\text{When } r=0, E\{[\pi(\underline{x}, \underline{T})/m]^0\} = E\{1\} = \int_R f_{\underline{X}}(\underline{x}) d\underline{x} = 1.$$

3. Due to the properties of variance, maximum value of second moment is 1 and the minimum value of the first moment is 0, thus, it is proved that the variance of mLIR is always bounded between 0 and 1.

4.4.2 Properties of the Inverted Probability Loss Functions

Leung and Spiring (2004) also showed various properties of the Inverted Probability Loss functions and have been listed in Section 2.7. Since the inverted probability loss function introduced in this thesis is based on multivariate probability density function, these properties of the inverted probability loss functions are thus modified as follows:

1. The general form of the risk function for multivariate IPLFs (ImPLFs) is

$$E[L(\underline{X}, \underline{T})] = E\{K[1 - \pi(\underline{X}, \underline{T})/m]\} = K\{1 - E[\pi(\underline{X}, \underline{T})/m]\}$$

2. The ImPLF is bounded between 0 and K, where K is the maximum loss incurred when the target vector is missed.

3. The variance of $L(\underline{X}, \underline{T})$ arises from using $f_R(\underline{x})$ as the process characteristics distribution given by

$$V[L(\underline{X}, \underline{T})] = K^2 \{E\{[\pi(\underline{X}, \underline{T})/m]^2\} - \{E[\pi(\underline{X}, \underline{T})/m]\}^2\}.$$

4. The variance of an ImPLF is K^2 times the variance of mLIR. Hence it is bounded between 0 and K^2 .

5. The loss inversion ratio is scale invariant under linear transformation.

6. The shape of IPLF is scale invariant under linear transformation.

7. The loss function is scale invariant under linear transformation.

8. The risk function is scale invariant under linear transformation.

Property 2 can be proved easily with a similar technique as illustrated in Section

4.4.1. Since $0 \leq \pi(\underline{x}, \underline{T})/m \leq 1$, by the following:

$$\begin{aligned}
0 &\leq \pi(\underline{x}, \underline{T})/m \leq 1 \\
0 &\leq 1 - \pi(\underline{x}, \underline{T})/m \leq 1 \\
0 &\leq K[1 - \pi(\underline{x}, \underline{T})] \leq K \\
0 &\leq \int_{\Omega} K[1 - \pi(\underline{x}, \underline{T})] f_R(\underline{x}) d\underline{x} \leq \int_{\Omega} K f_R(\underline{x}) d\underline{x} \\
0 &\leq E[L(\underline{X}, \underline{T})] \leq 1
\end{aligned}$$

Another interesting property of ImPLF is that the variance is always bounded between 0 and K^2 (Property 4). According to the definition of variance, the following is obtained:

$$\begin{aligned}
V[L(\underline{X}, \underline{T})] &= E\{[L(\underline{X}, \underline{T})]^2\} - E\{[L(\underline{X}, \underline{T})]\}^2 \\
&= K^2 \left[E\{[1 - \pi(\underline{X}, \underline{T})/m]^2\} - \{1 - E[\pi(\underline{X}, \underline{T})/m]\}^2 \right] \\
&= K^2 \left[E\{[\pi(\underline{X}, \underline{T})/m]^2\} - \{E[\pi(\underline{X}, \underline{T})/m]\}^2 \right].
\end{aligned}$$

With Property 2 of Section 4.4.1, it can be concluded that the variance of an inverted multivariate probability loss function will be bounded between 0 and K^2 .

Scale invariant (Property 7) is an important feature in the inverted probability loss functions introduced by Leung and Spiring (2004). If $\pi(\underline{x}, \underline{T})$ is a continuous multivariate pdf, having unique maximum at \underline{T} , then under any linear transformation, the loss inversion ratio, the IPLF and its shape and the risk function are all scale invariant. Under the linear transformation in multiple variants,

$$\begin{aligned} y_1 &= a_1 x_1 + b_1 \\ y_2 &= a_2 x_2 + b_2 \\ &\vdots \\ y_k &= a_k x_k + b_k \end{aligned}, \text{ where } a_1, a_2, \dots, a_k \text{ and } b_1, b_2, \dots, b_k \text{ are scalars and } \underline{X}, \underline{Y} \text{ are distributed}$$

by $f(\underline{x})$ and $g(\underline{y})$ respectively, thus:

$$\begin{aligned} g(y_1, y_2, \dots, y_k) &= f(x_1, x_2, \dots, x_k) |J| \\ &= f(x_1, x_2, \dots, x_k) \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \dots & \frac{\partial x_1}{\partial y_k} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_2}{\partial y_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_k}{\partial y_1} & \frac{\partial x_k}{\partial y_2} & \dots & \frac{\partial x_k}{\partial y_k} \end{vmatrix} \\ &= f(x_1, x_2, \dots, x_k) \begin{vmatrix} \frac{1}{a_1} & 0 & \dots & 0 \\ 0 & \frac{1}{a_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{a_k} \end{vmatrix} \\ &= \frac{1}{a_1 a_2 \dots a_k} f\left(\frac{y_1 - b_1}{a_1}, \frac{y_2 - b_2}{a_2}, \dots, \frac{y_k - b_k}{a_k}\right). \end{aligned} \quad (4.35)$$

Thus, $f(x_1, x_2, \dots, x_k) = a_1 a_2 \dots a_k g(y_1, y_2, \dots, y_k)$, and

$$\sup f(x_1, x_2, \dots, x_k) = \sup [a_1 a_2 \dots a_k g(y_1, y_2, \dots, y_k)] = a_1 a_2 \dots a_k \sup g(y_1, y_2, \dots, y_k).$$

Since, $m = \sup f(x_1, x_2, \dots, x_k)$, and if $m' = \sup g(y_1, y_2, \dots, y_k)$, it follows that

$m = a_1 a_2 \dots a_k m'$ or $m = \frac{1}{|J|} m'$. As $\frac{1}{|J|}$ is a scalar, with the following calculation:

$$\frac{g(y_1, y_2, \dots, y_k)}{m'} = \frac{g(y_1, y_2, \dots, y_k)}{|J| m} = \frac{f(x_1, x_2, \dots, x_k) |J|}{|J| m} = \frac{f(x_1, x_2, \dots, x_k)}{m} \quad (4.36)$$

It is proved that the loss inversion ratio before and after the linear transformation are indeed equivalent.

To prove the multivariate IPLF (Property 6) is also scale-invariant, Property 5 may be applied:

$$L[\underset{\sim}{y}, \underset{\sim}{T}'] = K[1 - g(y_1, y_2, \dots, y_k) / m'] = K[1 - f(x_1, x_2, \dots, x_k) / m] = L[\underset{\sim}{x}, \underset{\sim}{T}] \quad (4.37)$$

The risk function, $E[L(\underset{\sim}{y}, \underset{\sim}{T}')]$ can be shown to be equivalent to $E[L(\underset{\sim}{x}, \underset{\sim}{T})]$ by the followings:

$$\begin{aligned} E[L(\underset{\sim}{y}, \underset{\sim}{T}')] &= E\left\{K\left[1 - \frac{g(y_1, y_2, \dots, y_k)}{m'}\right]\right\} \\ &= E\left\{K\left[1 - \frac{f(y_1, y_2, \dots, y_k)}{m}\right]\right\} \\ &= E[L(\underset{\sim}{y}, \underset{\sim}{T})]. \end{aligned} \quad (4.38)$$

The above proofs have shown the properties of IPLF also work for loss functions built on multivariate probability density functions.

4.5 Fitting an Inverted Probability Loss Function

Spiring (1993) gave an approximation for fitting his Inverted Normal Loss Function. Fitting the INLF requires estimators for mean and standard deviation. Spiring suggested to use the target value as the estimator for mean. For the standard deviation, one quarter of the Euclidean distance from the target value to the point where maximum loss occurred was suggested. Similar procedure was introduced for his IbNLF. For a bivariate Normal distribution, covariance would also be required and $\gamma_{12} = \beta \frac{\Delta_2^2}{16}$. β was the slope of the principal axis and Δ_2 was the Euclidean distance from the target value to the point where maximum loss occurred on the secondary axis. Estimation of the covariance factor might be difficult since the slope of the principal axis might be hard to identify.

Sun et al. (1996) commented on the inverted normal loss function and suggested to fit parameters of an INLF by a method similar to the least squared method. This

method would require a set of measured quality characteristic and the actual loss corresponded to those measured points.

Leung and Spiring (2002) demonstrated an easy way to obtain the estimators of an IPLF by using the point where half of the maximum loss was depicted. This method required less information, since usually there is insufficient information obtained from the product engineer.

In this research project, it is assumed that the only information given by the product engineer is the target values and tolerance limits. Investigation will be done, with this little amount of information, on how to come up with reliable estimations for the parameters of IbPLFs. Detailed work will be illustrated through the examples in Chapter 5.

CHAPTER 5 APPLICATIONS

The applications of loss functions in quality assurance settings have grown remarkably among practitioners and quality assurance researchers as well as the important recognition of off-targetness by both customers and manufacturers. Kulkarni and Prybutok (2004) applied the exact form of loss function in Sun et al. (1996) in optimizing process investment and variance reduction decision for process improvement. Pan (2007) used a modified inverted normal loss function to assess manufacturing and environmental risks. Kulkarni (2008a, 2008b) examined a joint lot-sizing and processing investment problem with random yield and back orders using a modified form of inverted normal loss function. In this chapter, applications of inverted bivariate probability loss functions (IbPLF) in a wave soldering process and a manufacturing process of fasteners will be examined. The distribution of the process data will be used to build the conjugate distribution for the application of IbPLF. Parameters of the conjugate distribution will be estimated using method of maximum likelihood, method of moments and percentile matching method. Hence, the assumed distribution will be verified using bivariate Chi-square goodness-of-fit test and bivariate Kolmogorov-Smirnov test.

5.1 Wave Soldering Process

Wave soldering is a process by which electronic components are soldered to a printed circuit board (PCB) to form an electronic assembly. The PCB first enters a fluxing zone, a fluxer applies flux to the underside of the board. The process uses a tank to hold a quantity of molten solder which is monitored by the viscosity level. The components are inserted into or placed on the PCB and the loaded PCB is passed across a pumped wave or waterfall of solder. The solder wets the exposed metallic areas of the board, creating a reliable mechanical and electronic connection. Some defects may occur during the soldering process if the amount of solder and the viscosity level deviate from the pre-determined level. Cracks occur due to mechanical stress and can cause loss of conductivity. Cavities are caused from either contaminated surfaces, lack of flux or insufficient heating which will decrease in strength and conductivity. If the wrong solder thickness is produced, the product may be susceptible or too thin for the load placed upon it. The manufacturer believes the amount of solder used and the level of viscosity of the molten solder are the two quality characteristics which have the strongest impact to the quality of solder paste.

A sample of fifty specimens is collected and the weight of solder powder and the viscosity level are recorded. The tolerance limits of weight of solder powder and percentage of viscosity are (180g, 220g) and (10.5%, 12.5%) respectively. The target values are given by the engineers at the midpoint of the tolerance limits.

Data collected is shown in the following table:

Table 5-1 Collected data of solder paste

Solder Powder	Viscosity	Solder Powder	Viscosity	Solder Powder	Viscosity
202.3	11.5	200.2	11.45	204.1	11.49
200.4	11.55	184.2	11.82	211.6	11.37
195.3	11.61	196.4	11.65	217.4	11.26
206.2	11.43	192.7	11.55	186.2	11.74
206.3	11.44	189.8	11.63	180.9	11.83
210.7	11.37	202.4	11.46	218.4	11.27
220.8	11.21	205.4	11.47	209.4	11.34
207.6	11.34	196.4	11.53	206.5	11.45
201.6	11.56	219.6	11.28	210.1	11.43
181.6	11.82	214.5	11.31	208.2	11.42
185.4	11.77	212.4	11.39	206.3	11.41
218.5	11.26	201.8	11.46	205.5	11.4
186.2	11.73	218.2	11.26	210.8	11.37
199.5	11.55	205.6	11.47	188.4	11.61
198.1	11.56	201.4	11.48	203.4	11.5
197.2	11.56	196.1	11.61	200.2	11.49
202.4	11.51	193.7	11.51		

Figure 5-1 showed the scatter plot of the collected data:

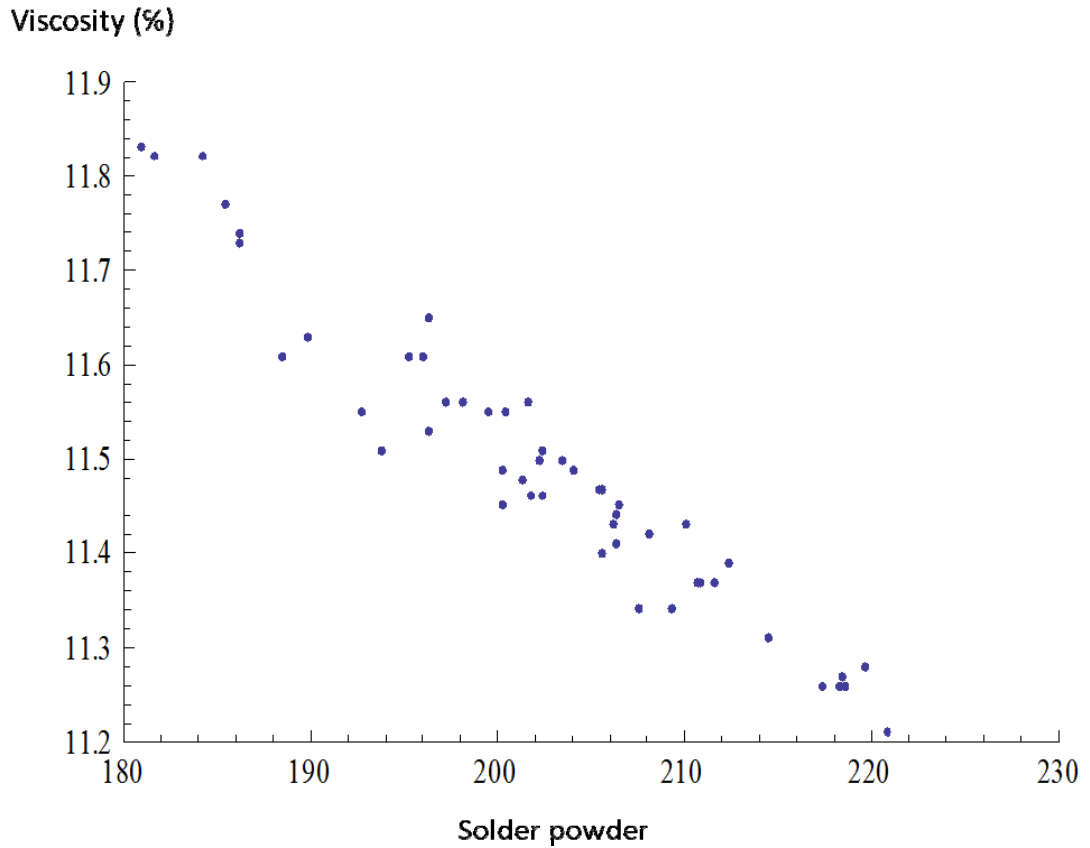


Figure 5-1 Scatter plot of solder paste data

Although a triangular shape is shown in the scatter plot, the data can also fit to a bivariate Normal distribution in order to select parameters of an IbNLF. For the conjugate function, parameters will be estimated by the method of maximum likelihood and the percentile matching method.

5.1.1 Maximum Likelihood Estimators

The maximum likelihood estimators for the parameters of a bivariate normal distribution are:

$$\begin{aligned}\hat{\mu}_X &= \bar{x} \\ \hat{\mu}_Y &= \bar{y} \\ \hat{\sigma}_X = s_X &= \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \\ \hat{\sigma}_Y = s_Y &= \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} \\ \hat{\rho} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\hat{\sigma}_X \hat{\sigma}_Y}\end{aligned}\tag{5.1}$$

Using fifty data points, the maximum likelihood estimates are

$$\left[\bar{x} \quad \bar{y} \quad s_X \quad s_Y \quad \hat{\rho} \right] = [202.366 \quad 11.4896 \quad 10.1532 \quad 0.151234 \quad -0.96559].$$

5.1.2 Percentile Matching Method

For the univariate case, the percentile matching method requires to equate, for example, the 5th and 95th percentiles of the empirical distribution and population, in

order to yield estimates on the parameters of a certain probability density function. For the bivariate case, extending the idea of univariate case, the percentile matching method requires to build a number of regions, which the observed frequency and the expected frequency in each region have to be matched. A potential list of estimates will be considered. For each set of potential estimates, the expected frequency of each pre-determined region will be compared to the observed frequency and the set of potential estimates which yield the smallest difference will be confirmed as the estimates obtained by the percentile matching method.

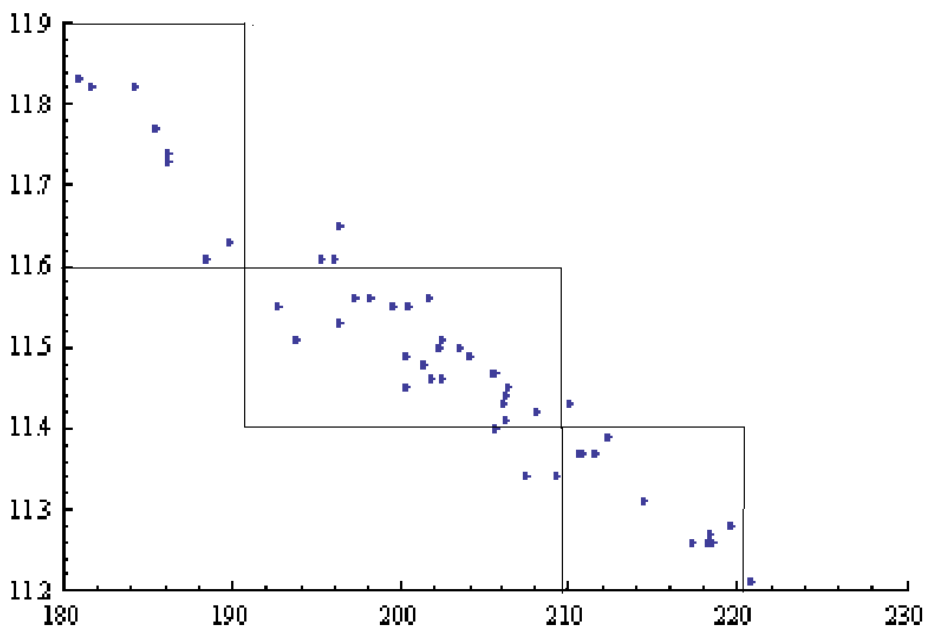


Figure 5-2 Illustration of regions used in the modified percentile matching method

Three regions are built encompassing most of the samples. (See Figure 5-2.)

The maximum likelihood estimates are used as a starting point and 20840 sets of values near to this starting point are considered as the potential set of estimates. For each set of the estimates estimated by mle, plus and minus ten percent of each estimates are used as the upper and lower limit. For each variable, each interval between plus and minus ten percent is then dissected into 8 equal sub-intervals, except for the estimate of correlation, since correlation cannot be less than -1. After lengthy calculation, the set of percentile matching estimates for the parameters of a bivariate normal distribution obtained is:

$$\left[\bar{x} \quad \bar{y} \quad s_x \quad s_y \quad \hat{\rho} \right] = [202.366 \quad 11.4896 \quad 10.407 \quad 0.136111 \quad -0.869031].$$

5.1.3 Chi-Square Goodness-of-Fit Test

As discussed previously in Chapter 3, the Chi-Square goodness-of-fit test requires the region being tested to be dissected into sub-regions. The problem with our data is that equal-sized sub-regions cannot be formed due to the fact that the data points are packed closely together. With the equal-probability principle, the 50 data points are dissected into ten sub-regions, each with approximately the same expected frequency. Ten sub-regions are chosen because there are five parameters to be

estimated. In other words, more than six sub-regions must be formed because the degrees of freedom for a Chi-square goodness-of-fit Test equal to $n-k-1$, where n denotes the number of sub-regions and k denotes the number of parameters to be estimated. Bearing in mind those sub-regions with expected frequency less than five must be merged with a nearby sub-region, ten seems to be a reasonable pace to start with.

The detail structure of the ten sub-regions is shown in Figure 5-3. Numerical integration is performed to find out the expected values of each sub-region, based on the two set of estimates, namely maximum likelihood estimators and percentile matching estimators.

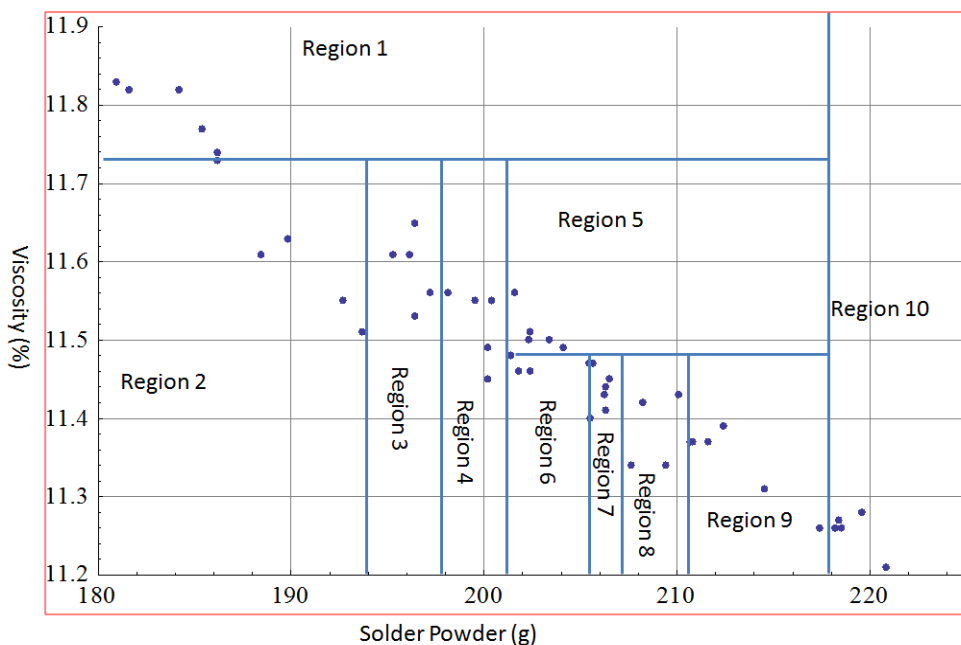


Figure 5-3 Illustration of cells dissection in the Chi-square goodness-of-fit test

Chi-square goodness-of-fit test statistics are then obtained in the following table:

Table 5-2 Observed versus expected frequency for the solder paste example with maximum likelihood estimates

Maximum Likelihood Estimates	$\bar{x} = 202.366, \bar{y} = 11.4896,$ $s_x = 10.1532, s_y = 0.151234, \hat{\rho} = -0.96559$				
Before combining cells		After combining cells			
Region	Expected Frequency	Region	Expected Frequency	Observed Frequency	Test Statistic
1	1.9304	1	9.5595	10	0.020298
2	7.6291	2	6.4301	5	0.318064
3	6.4301	3	5.6447	5	0.073633
4	5.6447	4	4.1271	5	0.184622
5	4.1271	5	5.0732	5	0.001056
6	5.0732	6	4.7594	5	0.012163
7	2.3851	7	5.8681	5	0.128423
8	5.8681	8	8.5379	10	0.250382
9	7.092			Sum	0.988642
10	1.446				
Remaining	2.3742				

Table 5-3 Observed versus expected frequency for the solder paste example with estimators obtained by the modified percentile matching method

Percentile Matching Method	$\bar{x} = 202.366, \bar{y} = 11.4896,$ $s_x = 10.407, s_y = 0.136111, \hat{\rho} = -0.869031$				
Before combining cells		After combining cells			
Region	Expected Frequency	Region	Expected Frequency	Observed Frequency	Test Statistic
1	1.1981	1	9.7873	10	0.004622
2	8.5892	2	6.2986	5	0.267736
3	6.2986	3	5.5137	5	0.04786
4	5.5137	4	5.9529	5	0.152534
5	5.9529	5	4.5868	5	0.037223
6	4.5868	6	4.4068	5	0.079851
7	1.8684	7	4.9547	5	0.000414
8	4.9547	8	8.4993	10	0.264975
9	6.8214			Sum	0.855215
10	1.6778				
Remaining	2.5384				

After combining cells, the test statistics obtained are 0.9887 and 0.8552 respectively for the method of maximum likelihood estimation and the method of percentile matching estimation. Each test is based on a Chi-square goodness-of-fit test with 2 degrees of freedom and the critical value of 4.60517 when $\alpha = 0.1$. Hence, the null hypothesis of data coming from a bivariate normal distribution is not rejected.

The shape of the conjugate function built from using maximum likelihood estimates is shown in (Figure 5-4) while the one using percentile matching estimates can be seen in (Figure 5-5).

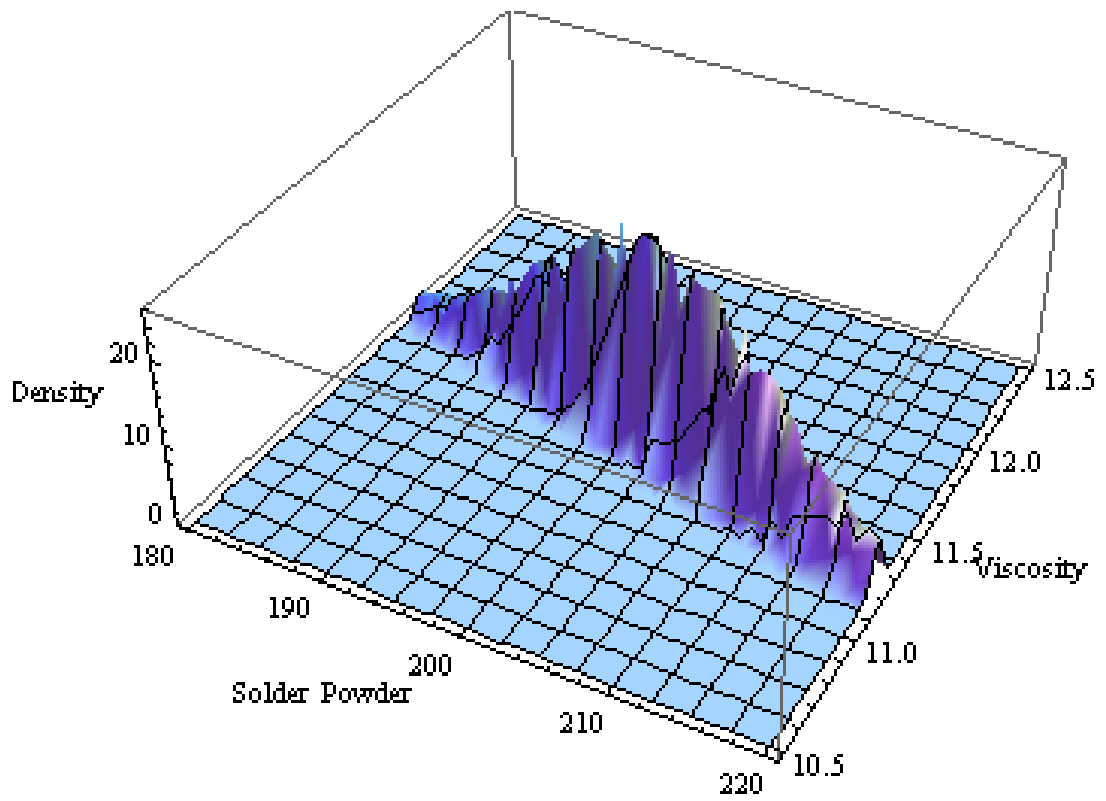


Figure 5-4 Conjugate function based on the maximum likelihood method

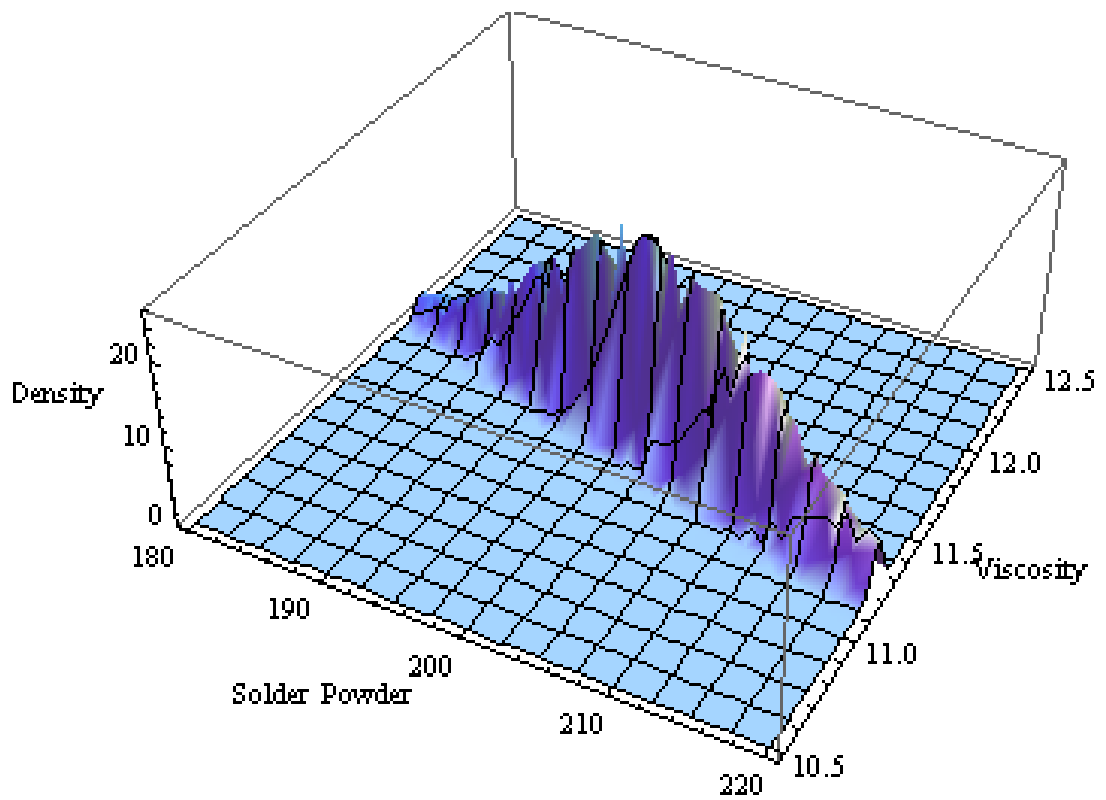


Figure 5-5 Conjugate function based on the modified percentile matching method

Both plots show that the density lie within the upper and lower limits of the weight of solder powder and of the level of viscosity.

5.1.4 Expected Loss

According to the information provided, target values are designed to be 200 grams and 11.5% for the weight of solder powder and viscosity respectively, with tolerance limits of (180, 220) and (10.5, 12.5) for the two quality characteristics. Assumed no further information is available, the choices of parameters for the IbNLF, such that non-negative losses can be collected over the appropriate region of the process data, will be discussed in this chapter.

Often, correlation is a measurement with little information provided. For this specific application, three scenarios are considered: no correlation, positive and negative correlation.

Consider the case with no correlation. Maximum financial loss of K will occur as the measurement reaches either end of the tolerance limits. Substituting the upper

limits into the bivariate normal distribution with zero correlation, and the following is obtained:

$$g(s_x, s_y) = \frac{1}{2\pi s_x s_y} \exp \left[-\left(\frac{220 - 200}{s_x} \right)^2 - \left(\frac{12.5 - 11.5}{s_y} \right)^2 \right] \quad (5.2)$$

Assuming $g(s_x, s_y) \geq 0$, all possible sets of estimates can be identified by plotting the function and its contour plot. The function $g(s_x, s_y)$ and its contour plot are plotted as follows:

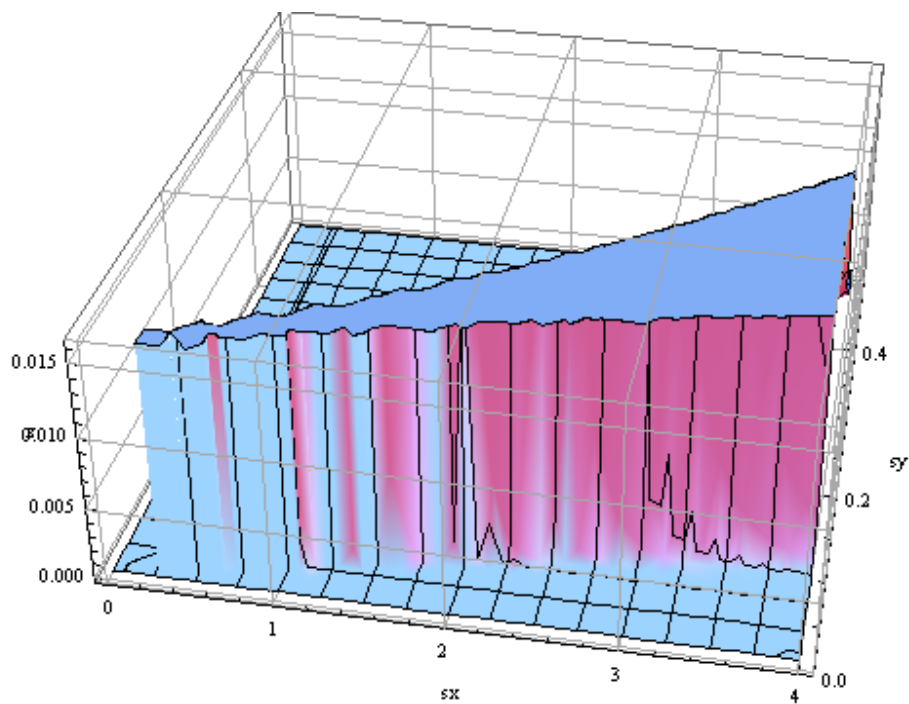


Figure 5-6 Three-dimensional plot of $g(s_x, s_y)$ in solder paste example

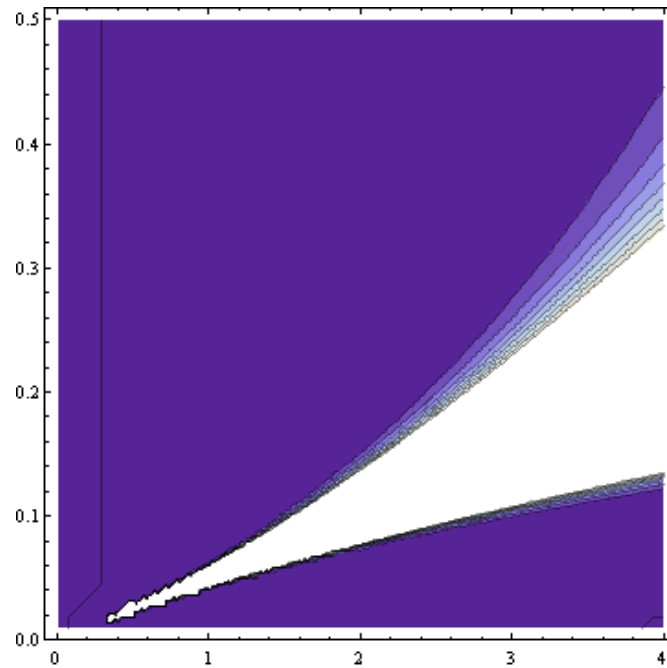


Figure 5-7 Contour plot of $g(s_x, s_y)$ in solder paste example

As illustrated in the contour plot (Figure 5-7), it can be seen that any combination of s_x and s_y in the region colored in deep purple would be an appropriate selection. A combination close to the boundary is favorable, or right at the edge of the surface which strikes out from the xy -plane. Hence, a combination of relatively large standard deviations and yet obtaining maximum financial loss will be a good choice.

Similar procedure can be done to the scenarios when the correlation of the conjugate function is assumed to be 0.95 and -0.95. Estimates of the conjugate function and the IbNLF under different scenarios are recorded in Table 5-4. Given the estimates

of both conjugate function and IbNLF, expected losses are depicted and are located in the last row of Table 5-4.

The proposed IbNLF when correlation is assumed to be zero are illustrated as follows:

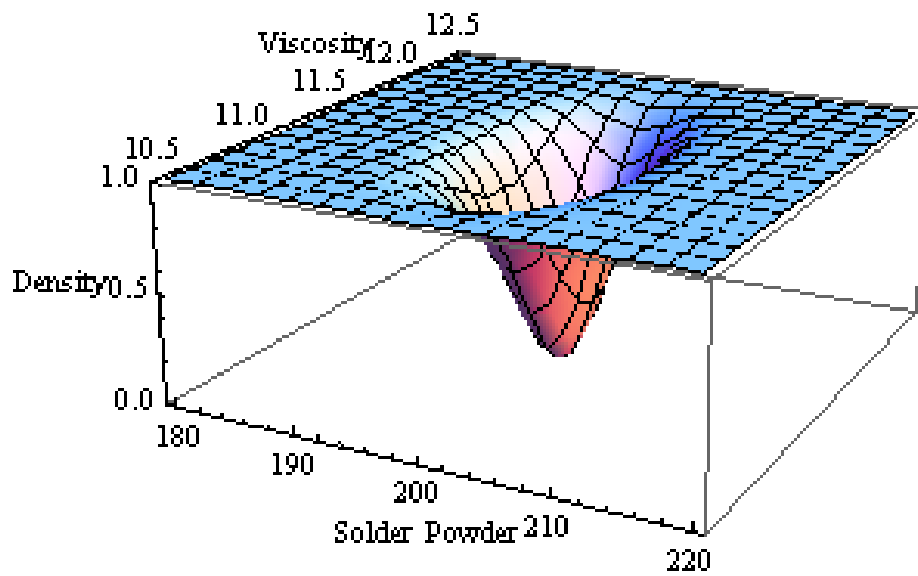


Figure 5-8 Plot of IbNLF when correlation is assumed to be zero

When correlation is assumed be -0.95, the IbNLF becomes:

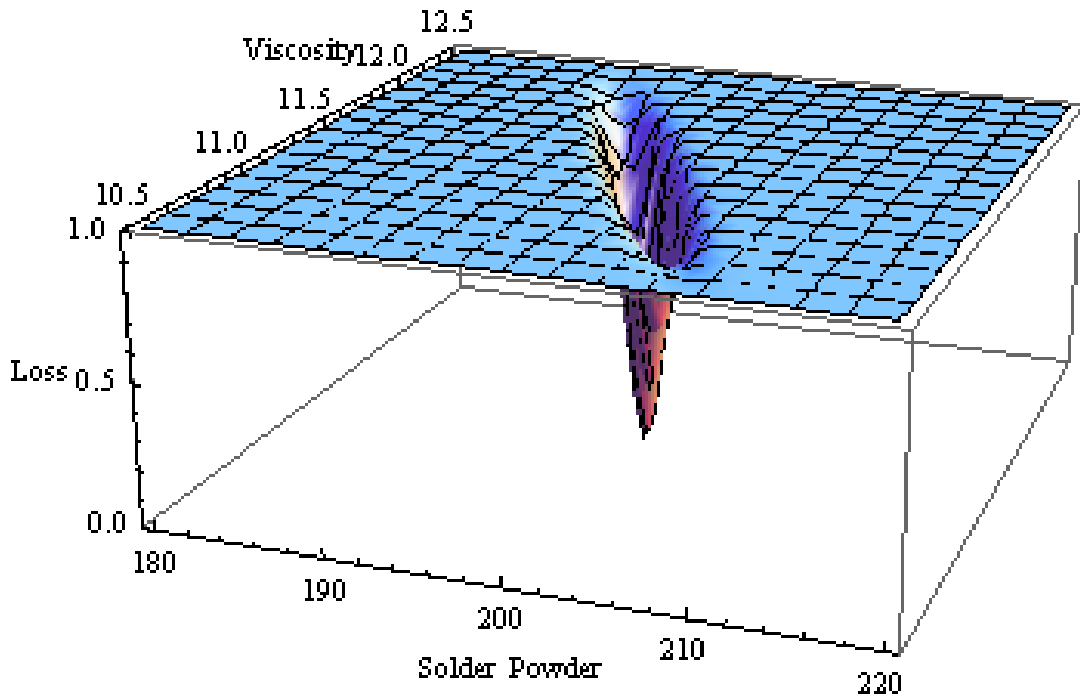


Figure 5-9 Plot of IbNLF when correlation is assumed to be -0.95

Table 5-4 Table of estimators of the conjugate function and the IbNLF

		MLE	Percentile Matching Method	MLE	Percentile Matching Method	MLE	Percentile Matching Method
Conjugate	\bar{X}	202.366	202.366	202.366	202.366	202.366	202.366
	\bar{Y}	11.4896	11.4896	11.4896	11.4896	11.4896	11.4896
	S_x	10.1532	10.407	10.1532	10.407	10.1532	10.407
	S_y	0.151234	0.136111	0.151234	0.136111	0.151234	0.136111
	r	-0.96559	-0.869031	-0.96559	-0.869031	-0.96559	-0.869031
IbNLF	μ_x	200	200	200	200	200	200
	μ_y	11.5	11.5	11.5	11.5	11.5	11.5
	S_x	3	3	3	3	3	3
	S_y	0.3	0.3	0.24	0.24	0.24	0.24
	r	0	0	0.95	0.95	-0.95	-0.95
Expected Loss		0.674118	0.676977	0.869268	0.864375	0.841244	0.844254

From the above table, the largest depicted financial loss for any one unit of solder paste is obtained when a positive correlation is assumed and its value is 0.869268.

The smallest financial loss depicted is 0.674118 when the parameters of the conjugate function are estimated by the maximum likelihood estimators and zero correlation is assumed for the IbNLF. An interesting finding is that the financial loss depicted with assumptions of positive and negative correlations are not different too much.

Similar expected losses obtained with positive and negative correlation can be explained by plotting the relationship between expected loss and correlation, and the relationship between the IbNLF and correlation if all other variables are held constant.

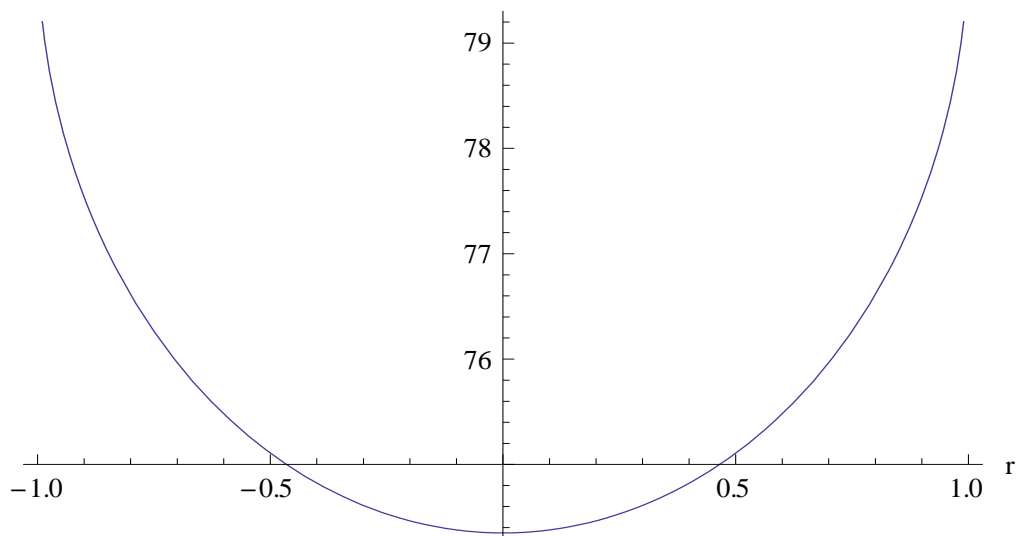


Figure 5-10 Plot of volume of IbNLF versus correlation

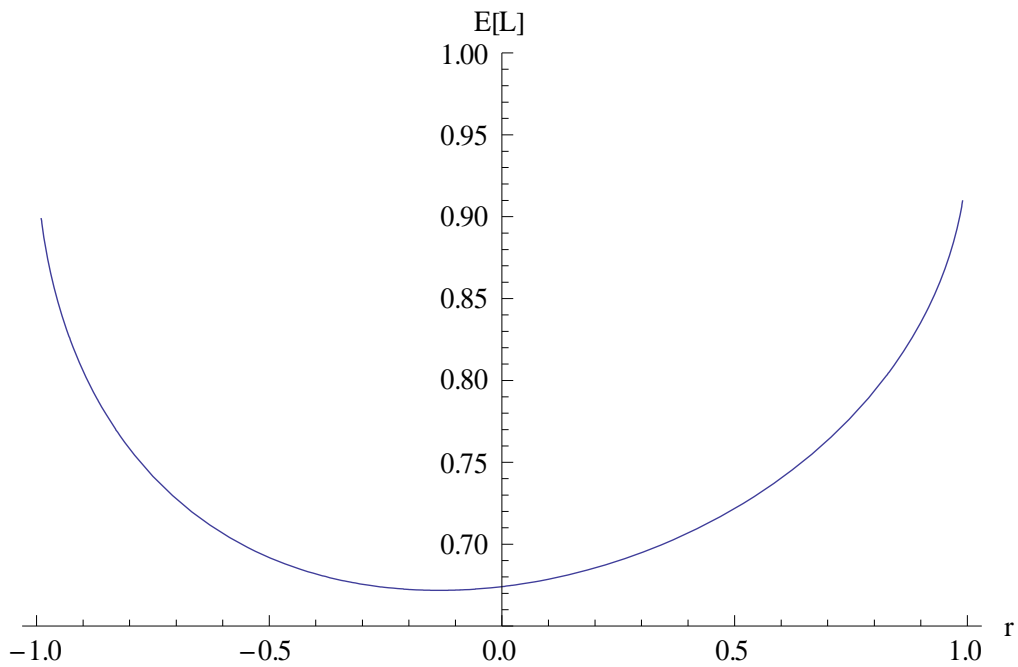


Figure 5-11 Plot of E[L] versus correlation

The volume under the IbNLF within the tolerance limits is a parabolic function. This can be explained analytically. When the correlation is positive, the two variables have a positive slope. There are two possible cases under this scenario: i) x is greater than \bar{x} and y is greater than \bar{y} and ii) x is smaller than \bar{x} and y is smaller than \bar{y} . Both cases end up with a negative sign inside the exponential function of the IbNLF. When the correlation is negative, the two possible cases would be: i) x is less than \bar{x} and y is greater than \bar{y} and ii) x is greater than \bar{x} and y is smaller than \bar{y} . Both cases once again end up with a negative sign inside the exponential function of the IbNLF. Based on this analysis, the IbNLF will always be

positive. As the magnitude of correlation increases, the quotient of $(1 - r^2)$ becomes very small and lead to a larger expected loss. Thus, the two expected losses depicted with correlation of positive and negative 0.95 are similar to each other.

5.2 Fasteners

A local merchandising company, who serves as the agent of a metal parts company in Germany, provided samples of two types of fasteners for this study. This German company specializes in fastening technology with state-of-the-art production facilities. The two types of fasteners provided are: i) Self-tapping Threaded Insert and ii) Pressed-in Threaded Insert.

5.2.1 Self-tapping Threaded Insert

Threaded insert is a self-tapping fastener for the creation of wear-free, vibration resistant screw joints with high loading capacity in materials with low shearing

strength. It is suitable for installation in duroplastics, thermoset plastics, aluminum, aluminum and magnesium alloys. Self-tapping inserts provide chip-free installation, but still provide strong holding power. The design of the self-tapping threaded insert is shown in the following diagram:

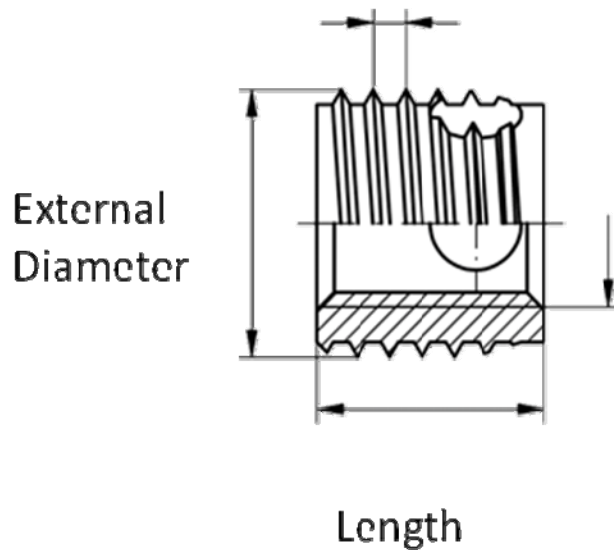


Figure 5-12 Design of the Self-tapping Threaded Insert

Measurements are collected by a Mitutoyo caliper, model number CD-6”C. Error of the caliper is about ± 0.001 inch, according to similar but later product. The two qualities under monitor are the length and diameter. Data collected from 51 samples are recorded in Table 5-5:

Table 5-5 Collected data of the self-tapping threaded inserts in mm

Length	Diameter	Length	Diameter	Length	Diameter
13.94	11.94	14	11.89	13.92	11.9
13.92	11.88	13.88	11.94	13.95	11.94
13.95	11.84	13.94	11.95	13.9	11.88
13.93	11.93	14.01	11.89	13.94	11.94
13.94	11.95	13.95	11.96	13.95	11.94
13.94	11.92	13.94	11.96	13.93	11.95
14.02	11.96	13.92	11.92	13.93	11.93
13.91	11.93	13.96	11.9	13.94	11.95
13.94	11.93	13.95	11.95	13.93	11.92
13.96	11.95	13.95	11.96	13.95	11.94
13.93	11.92	13.95	11.95	13.94	11.93
13.94	11.95	13.93	11.91	13.95	11.93
13.95	11.94	13.97	11.96	14.01	11.87
13.95	11.94	14.02	11.93	14.02	11.93
13.94	11.96	13.98	11.97	13.95	11.93
13.95	11.93	13.94	11.92	13.94	11.94
13.94	11.92	13.96	11.94	13.93	11.94

Based on the data collected, a scatter plot was done as the following:

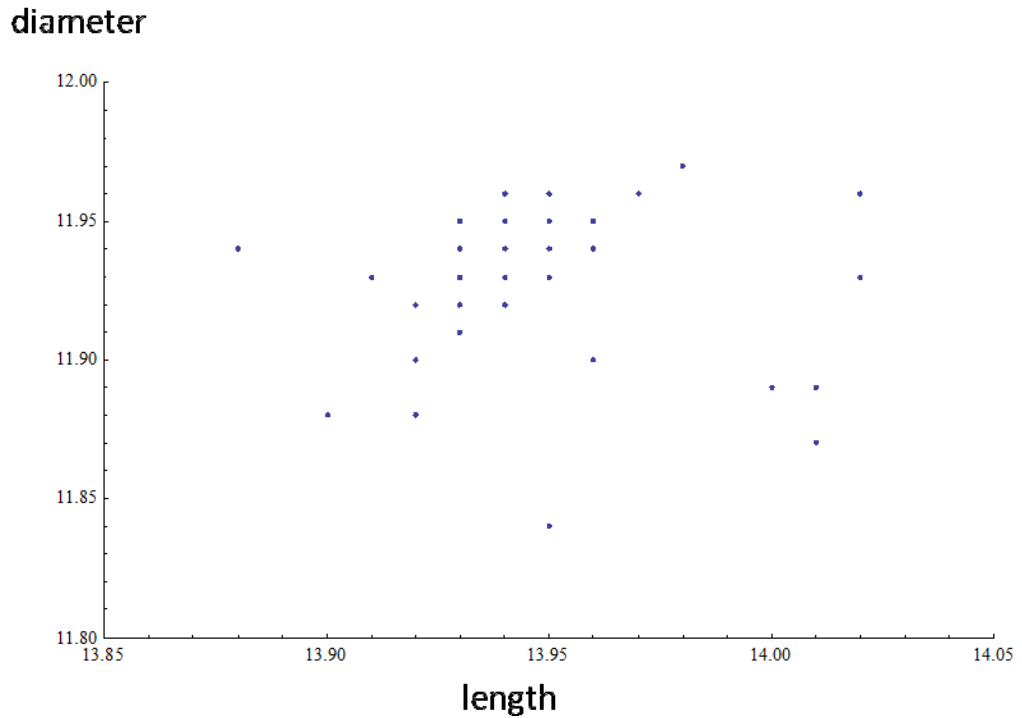


Figure 5-13 Scatter plot of Self-tapping Threaded Inserts data in mm

Data points are spreading out near the centre of the tolerance, which are (13.85mm, 14.15mm) and (11.8mm, 12.0mm) for the length and diameter respectively. As the measurements spread out nicely, a bivariate inverted normal loss function is suggested to depict the financial loss with this self-tapping threaded inserts.

Using the approach similar to the solder paste example, maximum likelihood estimates are:

$$[\hat{\mu}_x \quad \hat{\mu}_y \quad s_x \quad s_y \quad \hat{\rho}] = [13.9486 \quad 11.9304 \quad 0.0286301 \quad 0.0258169 \quad -0.0204941].$$

Using the maximum likelihood estimates as the starting values for the percentile matching method, a set of estimates is obtained:

$$[\hat{\mu}_x \quad \hat{\mu}_y \quad s_x \quad s_y \quad \hat{\rho}] = [13.9486 \quad 11.9304 \quad 0.0257671 \quad 0.0232352 \quad -0.0184447].$$

Based on the two sets of estimates, two conjugate functions are obtained. Their plots are displayed below:

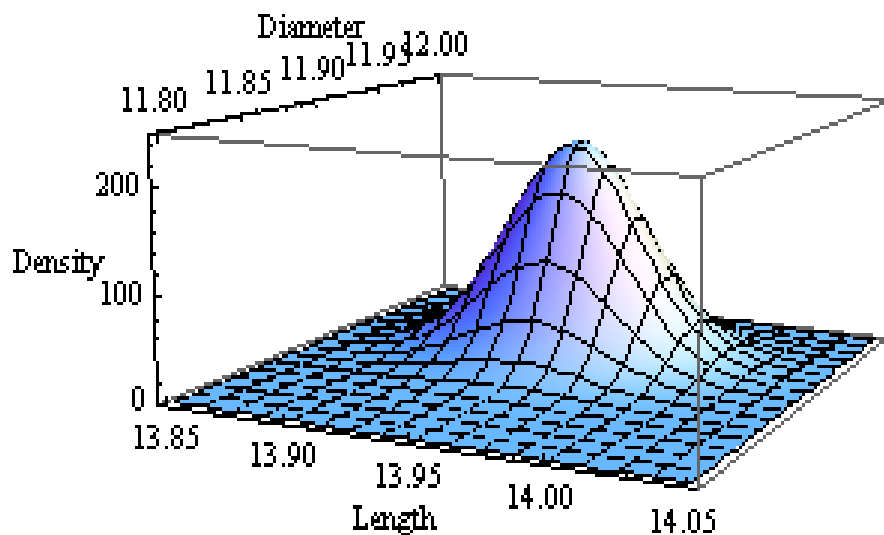


Figure 5-14 Conjugate function of self-tapping threaded insert by mle

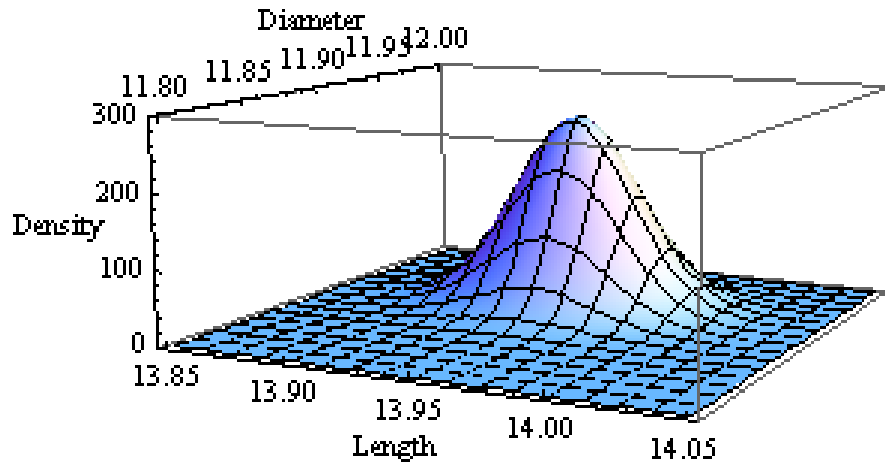


Figure 5-15 Conjugate function of self-tapping threaded insert by percentile matching method

Because of the smaller standard deviations and a correlation of smaller magnitude, the conjugate function in (Figure 5-15) has a larger supremum than the conjugate function plotted in (Figure 5-14).

From the scatter plot (Figure 5-13), duplications are observed at various points. The setup of sub-regions for the Chi-Square goodness-of-fit test becomes very complicated as the number of sub-regions must be greater than the degrees of freedom with five parameters to be estimated. Since the sample indicates that the products are very precise and sticks to the arithmetic mean closely, 10 sub-regions are a reasonable setup. The detail setup can be seen in Figure 5-16:

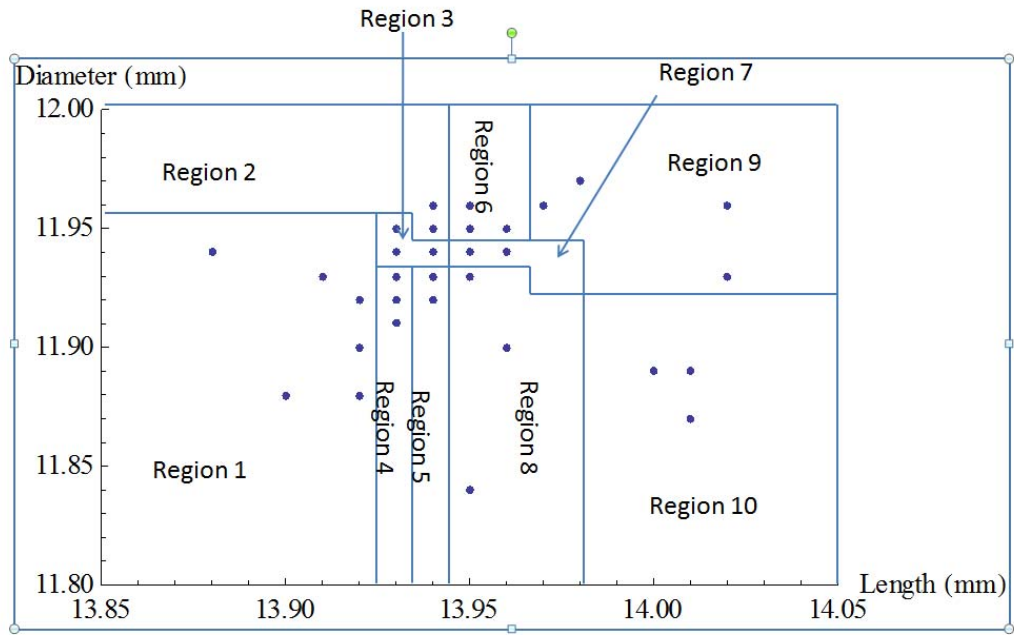


Figure 5-16 Illustration of cells dissection in the Self-tapping Threaded Inserts example

The expected and observed number of observations after cells being combined is:

Table 5-6 Observed versus expected frequency with the two estimation methods

Maximum Likelihood Method				Percentile Matching Method			
Region	Expected	Observed	$(E_i - O_i)^2 / E_i$	Region	Expected	Observed	$(E_i - O_i)^2 / E_i$
1	8.48834	6	0.729452	1	7.74511	6	0.393204
2	4.71092	6	0.352739	2	4.23094	6	0.739687
3	2.46866	5	2.595611	3	2.85807	5	1.605232
4	3.54175	5	0.600407	4	3.5178	5	0.624514
5	3.8429	5	0.348404	5	4.27866	5	0.121611
6	3.81958	5	0.364802	6	3.92054	5	0.297213
7	4.1681	6	0.805129	7	4.84566	6	0.274989
8	10.9613	5	3.242051	8	11.8663	5	3.973107
9	6.01218	5	0.170405	9	5.27239	5	0.014073
10	2.98627	3	6.31E-05	10	2.38926	3	0.156117
Chi-square Test Stat.			9.209065	Chi-square Test Stat.			8.199745

The test statistics obtained by our two models are 9.209065 and 8.199745 for the maximum likelihood method and percentile matching method respectively. The p -values for the two models are 0.0561 and 0.0845. With $\alpha = 0.10$, the critical value for a Chi-square distribution with 4 degrees of freedom is 7.77944, and 9.48773 if $\alpha = 0.05$. In other words, the two models obtained by the maximum likelihood method and percentile matching method are not rejected with 95% confidence.

In the test above, cells with expected frequency less than five were not merged with nearby cells. In order to perform the Chi-square goodness-of-fit test without bending this rule of thumb, a test with 8 sub-regions is then performed and the sub-regions are distributed as below:

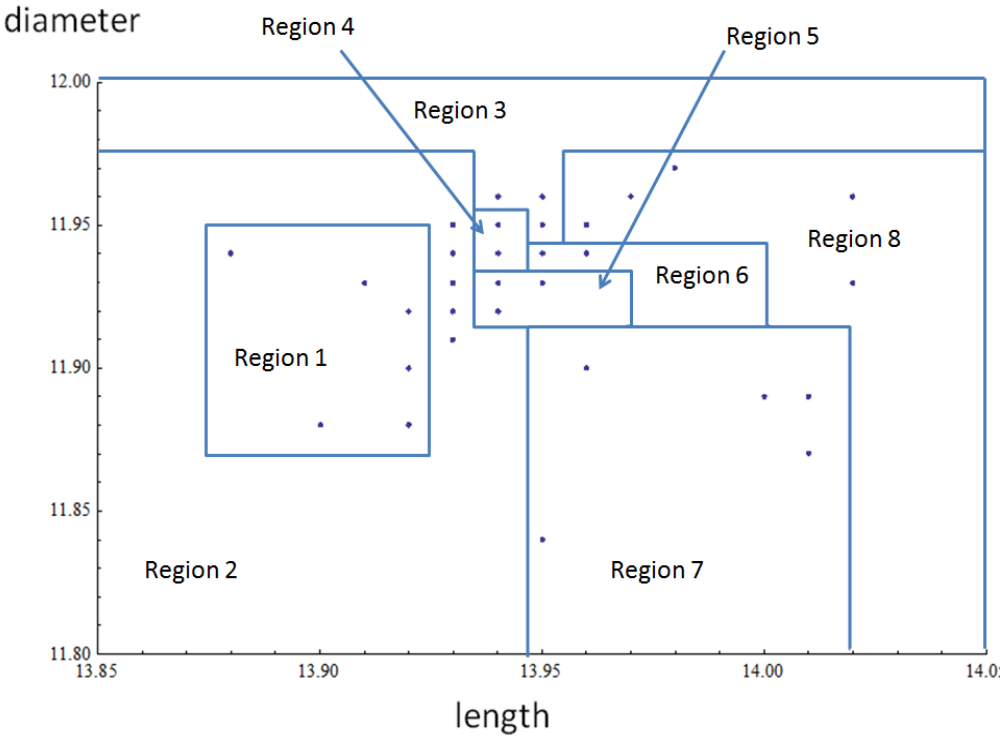


Figure 5-17 Illustration of cells distribution when number of cells are assumed to be 8

Calculation of the Chi-square goodness-of-fit test statistic can be found in the following table:

Table 5-7 Calculation of Chi-square goodness-of-fit test statistic

Before Merging							
Maximum Likelihood Method				Percentile Matching Method			
Region	Expected	Observed		Region	Expected	Observed	
1	7.73304	6	0.729452	1	7.14903	6	0.393204
2	8.76959	7	0.352739	2	8.87431	7	0.739687
3	5.36232	6	2.595611	3	4.71142	6	1.605232
4	1.757	7	0.600407	4	2.06239	7	0.624514
5	6.86008	8	0.348404	5	8.25394	8	0.121611
6	6.63196	6	0.364802	6	7.23156	6	0.297213
7	7.7689	5	0.805129	7	7.26926	5	0.274989
8	5.91342	6	3.242051	8	5.37279	6	3.973107

After Merging							
Maximum Likelihood Method				Percentile Matching Method			
Region	Expected	Observed		Region	Expected	Observed	
1	7.73304	6	0.388389	1	7.14903	6	0.18467819
2 and 4	10.73027	14	0.99635277	2	8.87431	7	0.39586604
3	5.36232	6	0.07583206	3 and 4	6.84911	13	5.52384876
5	6.86008	8	0.18941727	5	8.25394	8	0.0078127
6	6.63196	6	0.06021952	6	7.23156	6	0.20973898
7	7.7689	5	0.98685878	7	7.26926	5	0.70839961
8	5.91342	6	0.00126764	8	5.37279	6	0.07321939
		Test Stat.	2.69833704			Test Stat.	7.10356367

After merging the sub-regions with expected frequency less than 5, there is one degree of freedom for the Chi-square goodness-of-fit test. According to the obtained test statistics, the model built by maximum likelihood method has a p -value of 0.10 and the model built by percentile matching method has a p -value of 0.0077.

Based on the results of the two Chi-square goodness-of-fit tests with different number of sub-regions, one can conclude that the proposed models are not very good. The results also show that: i) different way of setting up the sub-regions will lead to a different test statistic and thus, the result is not unique among all statisticians who perform the test; ii) the rule of merging of sub-regions with expected frequency less than five may not be as important as in the univariate case. The p -values obtained from the models built by maximum likelihood method are very similar, which may indicate loosening of this rule may be compensated by building of sub-regions.

Arguments have been made against the Chi-Square goodness-of-fit test in assigning cells and being less stringent. A bivariate Kolmogorov-Smirnov test is performed to verify the conclusions given by the Chi-square goodness-of-fit test.

Peacock (1983) extended the well known Kolmogorov-Smirnov test to a bivariate case. His procedure requires calculation to be done at each sample point. For each sample point, Peacock suggested to use the sample point as the origin and considered the four quadrants around it. For all four quadrants, calculation of the difference between the observed density and expected density has to be done, based on the distribution with proposed estimates. The maximum absolute difference becomes the test statistics to test whether the hypothesized distribution is acceptable or not.

Calculations have been done by Mathematica. A maximum absolute difference is found to be 0.264707 in the model with maximum likelihood estimates. This maximum difference, also known as D_n , is then handled with the following procedures:

$$\begin{aligned}
 Z_n &= \sqrt{n}D_n = \sqrt{51}(0.264707) = 1.89039 \\
 Z_\infty &= \frac{Z_n}{1 - 0.53n^{-0.9}} = 1.91995 \\
 P(> Z_\infty) &= 2 \exp\left[-2(Z_\infty - 0.5)^2\right] = 0.035493
 \end{aligned} \tag{5.3}$$

A p -value of 0.0354593 is then obtained. Indicating, if $\alpha = 0.05$, our proposed model will be rejected, while, if $\alpha = 0.01$, our model will not be rejected. Similarly, for the modified percentile matching method, a p -value of 0.0236954 is obtained. A conclusion of not rejecting when $\alpha = 0.01$ is thus being made.

If the result from Chi-square goodness-of-fit test is to be recalled, although the conclusion does not exactly agrees with the bivariate Kolmogorov-Smirnov test (rejection for $\alpha = 0.05$), it is well known that the Kolmogorov-Smirnov test is meant to be more conservative and more powerful than the Chi-square goodness-of-fit test.

If the two set of estimates obtained above are not rejected with significance level $\alpha = 0.05$ from the Chi-square goodness-of-fit test, they can be used in obtaining the conjugate function and deriving the expected loss as in the solder paste example. In order to obtain the estimates of the standard deviations for the IbNLF, similar procedures from the previous application of solder paste can be applied. A function with the target values given and expressed in terms of the standard deviations and correlation can be set up as:

$$f = \frac{1}{2\pi S_x S_y \sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[\left(\frac{14.05-13.95}{S_x} \right)^2 + \left(\frac{12-11.9}{S_y} \right)^2 - \frac{2r(14.05-13.95)(12-11.9)}{S_x S_y} \right] \right\} \quad (5.4).$$

Due to the fact that there is almost zero correlation obtained in the conjugate function, a correlation coefficient of zero is assumed and the result is obtained and showed in the following table:

Table 5-8 Estimates of the conjugate function and IbNLF

		MLE	Percentile Matching Method	MLE	Percentile Matching Method
Conjugate	\bar{X}	13.9486	13.9486	13.95	13.95
	\bar{Y}	11.9304	11.9304	11.9	11.9
	S_x	0.0286301	0.0257671	0.0286301	0.0257671
	S_y	0.0258169	0.0232352	0.0258169	0.0232352
	r	-0.0204941	-0.0184447	-0.0204941	-0.0184447
IbNLF	μ_x	13.95	13.95	13.95	13.95
	μ_y	11.9	11.9	11.9	11.9
	S_x	0.0275	0.0275	0.0275	0.0275
	S_y	0.0275	0.0275	0.0275	0.0275
	r	0	0	0	0
Expected Loss		5.18134	4.8671	3.95475	3.53967

The last two columns assume that if the manufacturer can fix up the machine and shift the mean dimension of X and Y back to the target. Results show, for a piece of fastener which costs approximately HK\$8, although the sample standard deviations are found to be very small already, the expected financial loss seems to be a little high. This may be caused by over-estimating the standard deviations for the IbNLF. Once again, neither the standard deviations nor the correlation coefficient can be found precisely. Statisticians can only come up with close estimation based on their own judgment and experience.

5.2.2 Pressed-in Threaded Inserts

A pressed-in threaded insert can be installed into the receiving hole by either a hand lever or a small press. Ultrasonic or heat equipment are not required to install this particular insert into material like hard plastic. Detail design can be seen in the following diagram:

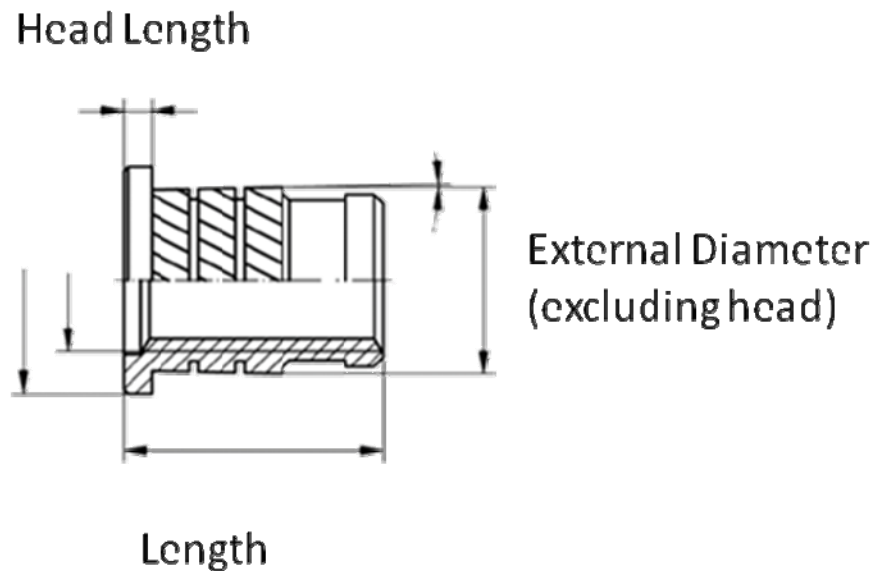


Figure 5-18 Design diagram of the pressed-in threaded insert

The length of an insert is an important quality characteristic of a fastener as it contributes to the amount of pullout strength that it resists. The total length of such a fastener can be separated into the length of head and length of body. If x denotes the length of body and y denotes the length of head, $x + y \leq z$, where z stands for the

total length. This relationship, if linear transformations are performed, will end up as $x' + y' \leq 1$, which is a triangular region which is suitable to the domain of a Dirichlet distribution. Thus, an IDLF will be used to depict the financial loss due to the deviation of these two quality characteristics away from their target values.

Measurements of 60 samples are recorded as:

Table 5-9 Collected data on the pressed-in threaded inserts in mm

Length of Body	Length of Head	Length of Body	Length of Head	Length of Body	Length of Head
12.72	1	12.75	1	12.68	1.06
12.8	0.94	12.73	0.97	12.73	1.02
12.7	1	12.72	1	12.78	0.93
12.73	0.99	12.66	1.04	12.71	1.04
12.72	0.99	12.68	1.04	12.73	0.98
12.65	1.01	12.75	0.95	12.66	1.02
12.67	1.03	12.7	1.03	12.71	1.01
12.69	1.02	12.78	0.98	12.69	1.03
12.68	1.03	12.69	0.99	12.66	1.06
12.78	0.94	12.72	1	12.69	1.01
12.71	0.98	12.75	0.97	12.71	0.97
12.72	1.01	12.71	1	12.68	1.05
12.65	1.05	12.71	1.02	12.68	1.02
12.7	1.01	12.71	1.01	12.71	1.02
12.73	0.96	12.72	1	12.73	1.03
12.77	0.94	12.75	0.98	12.66	1.06
12.74	1.02	12.7	0.99	12.74	0.95
12.77	0.99	12.79	0.96	12.72	0.99
12.72	1.01	12.73	0.98	12.75	0.99
12.77	0.97	12.63	1.11	12.78	0.97

5.2.2.1 Transformation

The product engineer provided the tolerance limits for the total length and the body length of the pressed-in self inserts. The limits given for the total length are (13.6, 13.8) and the limits for the length of head are (0.9, 1.1) in unit mm. In order to demonstrate the feasibility of the IDLF, the financial loss depicted by the body length and the length of head are being studied. Because the domain of a Dirichlet distribution is defined as $x \geq 0, y \geq 0, x + y \leq 1$, a linear transformation is performed to show that the length of body and the length of head can be well described by a Dirichlet distribution.

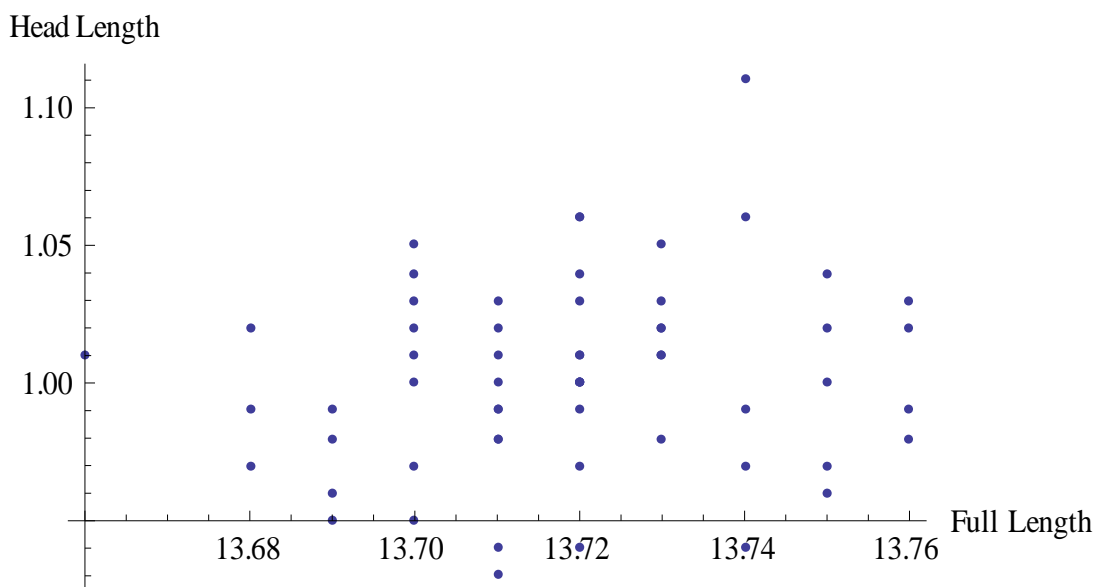


Figure 5-19 Scatter plot of the original measurements

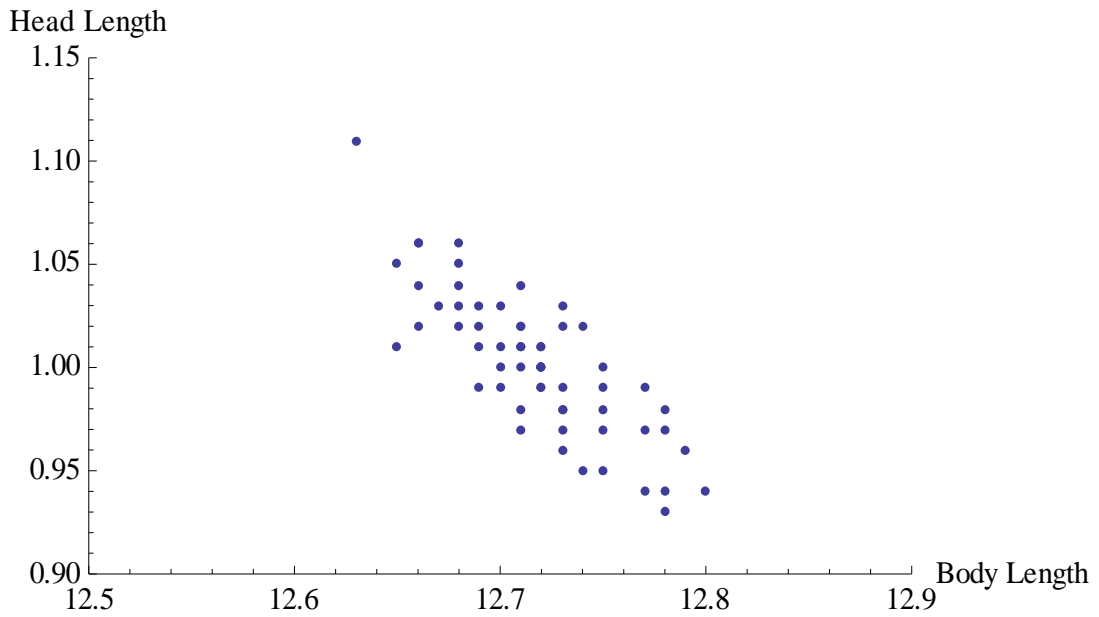


Figure 5-20 Scatter plot of length of head vs length of body

According to Leung and Spiring (2002), if Y has a generalized beta distribution with parameters $\alpha > 0, \beta > 0$ and y is bounded between p and q , where p and q are the lower and upper tolerance limits of the measurement of Y , then Y could be transformed to a standard beta distribution, which is the same distribution used to develop the inverted beta loss function. The probability density function of Y was given as:

$$g(y) = \frac{1}{B(\alpha, \beta)(q-p)} \left(\frac{y-p}{q-p} \right)^{\alpha-1} \left(\frac{q-y}{q-p} \right)^{\beta-1}, \quad p < y < q, \quad (5.5)$$

based on the linear transformation of $X = \frac{Y-p}{q-p}, |J| = \left| \frac{dy}{dx} \right| = q-p$.

For the bivariate case, similar procedure can be followed. If X_1, X_2 are distributed by a Dirichlet distribution with parameters of $\alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0$, Y_1, Y_2 can be transformed into X_1, X_2 by $X_1 = \frac{Y_1 - p_1}{q_1 - p_1}, X_2 = \frac{Y_2 - p_2}{q_2 - p_2}, |J| = (q_1 - p_1)(q_2 - p_2)$, where p_1, q_1 are the lower and upper tolerance limits of X_1 and p_2, q_2 are the lower and upper tolerance limits of X_2 . Thus, the bivariate probability density function of Y_1, Y_2 is:

$$g(y_1, y_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \left(\frac{y_1 - p_1}{q_1 - p_1}\right)^{\alpha_1 - 1} \left(\frac{y_2 - p_2}{q_2 - p_2}\right)^{\alpha_2 - 1} \left(1 - \frac{y_1 - p_1}{q_1 - p_1} - \frac{y_2 - p_2}{q_2 - p_2}\right)^{\alpha_3 - 1} \times \frac{1}{(q_1 - p_1)(q_2 - p_2)}, p_1 < y_1 < q_1, p_2 < y_2 < q_2 \quad (5.6)$$

It is easy to show that:

$$f(x_1, x_2) = g(y_1, y_2)|J| = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} (1 - x_1 - x_2)^{\alpha_3 - 1} \quad (5.7)$$

According to the tolerance limits given by the product engineer, $p_1 = 12.6, q_1 = 12.9, p_2 = 0.9, q_2 = 1.2$, and based on the above setting, the transform data are demonstrated in Figure 5-21.

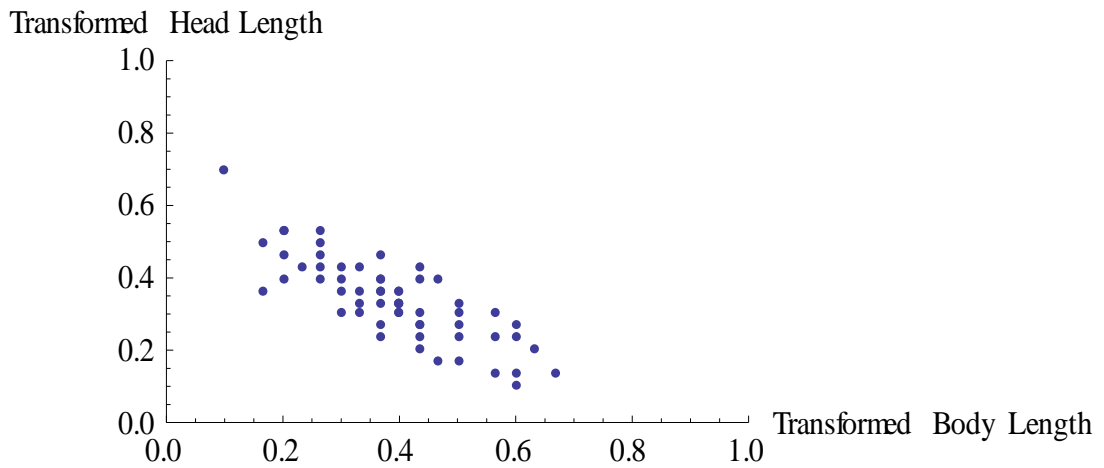


Figure 5-21 Scatter plot of transformed data collected from pressed-in threaded inserts

5.2.2.2 Construction of conjugates

After performing the linear transformation, the following three sets of estimates are obtained:

	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$
Method of Moments	5.3257	4.65618	3.71277
Maximum Likelihood	6.86421	6.04221	5.02687
Percentile Matching	5.5057	4.83618	3.51277

A Chi-square goodness-of-fit test is performed with 9 sub-regions. The detail setup of sub-regions can be seen in Figure 5-17.

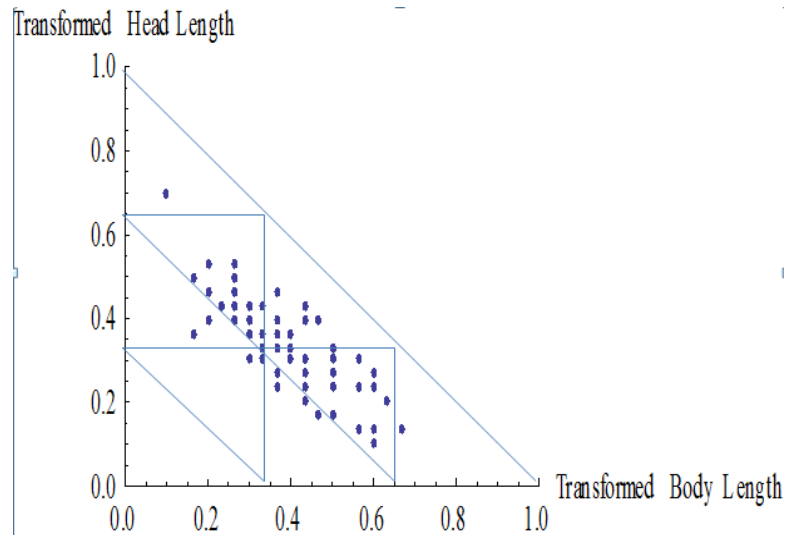


Figure 5-22 Illustration of cell dissections for the Chi-square goodness-of-fit test

Table 5-9 illustrates the test statistics obtained from the three proposed models.

Table 5-10 Observed versus expected frequency for the three estimation methods

Method of Moments			Maximum Likelihood Estimators			Modified Percentile Matching		
Expected	Observed	$(E_i - O_i)^2 / E_i$	Expected	Observed	$(E_i - O_i)^2 / E_i$	Expected	Observed	$(E_i - O_i)^2 / E_i$
8.068046	10	0.462620899	10.06256	10	0.000388942	9.924255	10	0.000578109
11.47539	9	0.533975169	10.73085	9	0.279180281	11.16416	9	0.419520009
5.58461	6	0.030897207	7.03511	6	0.152300776	6.66182	6	0.065748656
18.41349	23	1.142427317	17.3751	23	1.820967937	16.8348	23	2.257804728
16.4584	12	1.207731648	14.79632	12	0.528469616	15.41494	12	0.756526798
		3.377652241			2.781307553			3.500178301

There is one degree of freedom for each of the three models. As indicated in Table 5-9, test statistics of 3.38, 2.78, and 3.5 are obtained for the three models respectively. The critical value of one degree of freedom is 3.84146 when $\alpha = 0.05$,

and 2.7 when $\alpha = 0.10$. In other words, all three models are not rejected by the Chi-square goodness-of-fit test with 0.05 level of significance.

The followings are three plots of conjugate functions built by the fore-mentioned sets of estimates:

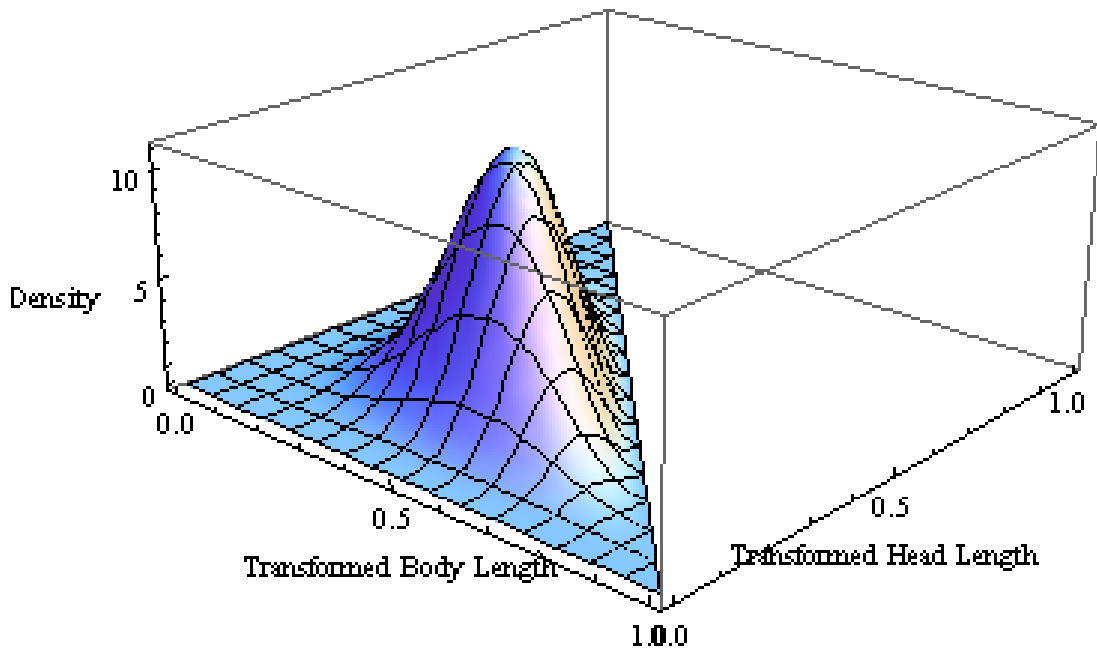


Figure 5-23 Conjugate function built by method of moments estimates

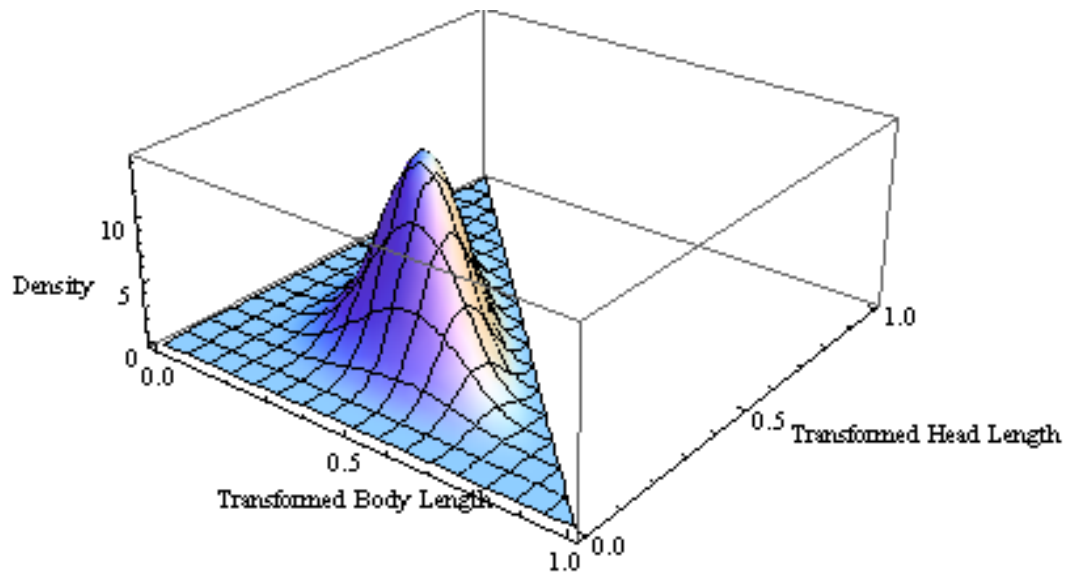


Figure 5-24 Conjugate function built by mle estimates

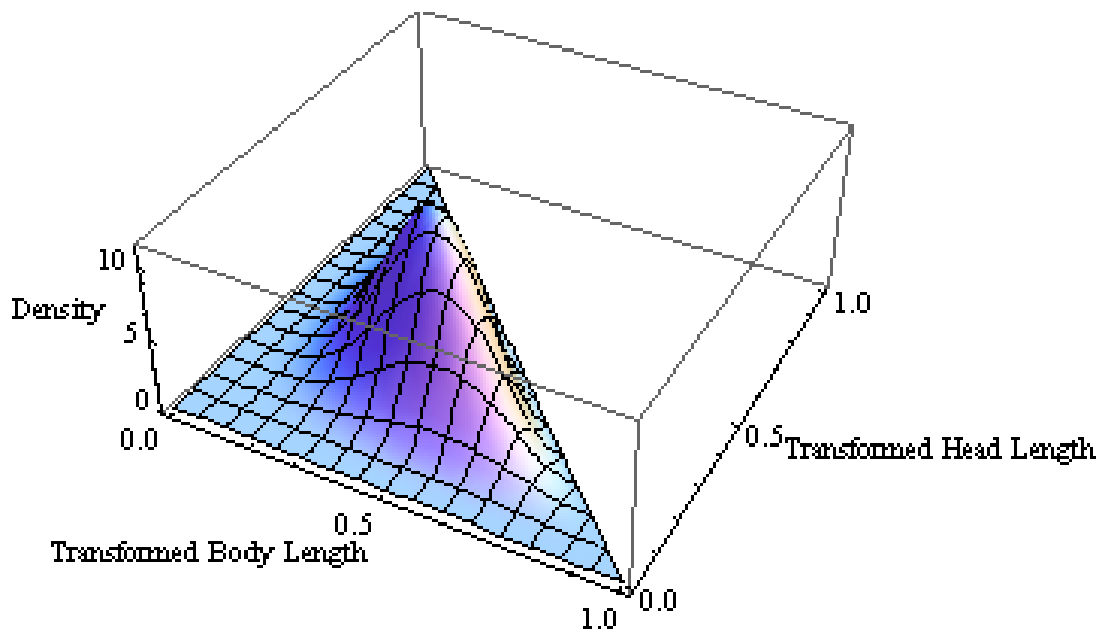


Figure 5-25 Conjugate function built by percentile matching estimates

Peacock's 2-Dimensional Kolmogorov-Smirnov test is applied to verify whether the dissection into nine sub-regions is appropriate. Based on the calculation similar to

the previous section, a maximum distance calculated inside the four quadrants around each point is equal to 0.147138. With the test statistic developed by Peacock,

$$\begin{aligned}
 Z_n &= \sqrt{n}D_n = \sqrt{60}(0.151807) = 1.17589 \\
 Z_\infty &= \frac{Z_n}{1 - 0.53n^{-0.9}} = 1.19174 \\
 P(> Z_\infty) &= 2 \exp\left[-2(Z_\infty - 0.5)^2\right] = 0.76808
 \end{aligned}
 \tag{5.8}$$

The significance level being calculated equals to 0.76808, which results in a failure to reject the hypothesized model with the method of moments. A summary of the test statistics and their p -values can be found in the following table:

Table 5-11 Calculation for Peacock's 2D K-S test

	Method of Moments	Maximum Likelihood Method	Percentile Matching Method
D_n	0.151807	0.170832	0.151426
Z_n	1.17589	1.32326	1.17294
Z_∞	1.19174	1.3411	1.18875
p -value	0.76808	0.485901	0.774447

According to the table above, all three sets of estimates are not rejected based on Peacock's 2D Kolomogorov-Smirnov tests. The conclusions of not rejecting models with the three sets of estimates agree with the Chi-square goodness-of-fit test with nine triangular sub-regions. This result confirms the assumption of dissecting the triangular domain into equal-sized triangles is indeed reasonable and feasible.

5.2.2.3 Expected Loss

Based on the assumed target values and tolerances in this example, and making use of the results yielded from Chapter 3, the followings are obtained after transformation:

$$T_x = \frac{12.7-12.6}{0.3} = \frac{1}{3} = \frac{a_1-1}{a_1+a_2+a_3-3},$$

$$T_y = \frac{1-0.9}{0.3} = \frac{1}{3} = \frac{a_2-1}{a_1+a_2+a_3-3}.$$

Setting the two equations equivalent, a relationship of $a_1 = a_2 = a_3$ is obtained.

Given the relationship of $a_1 = a_2 = a_3$, there are infinite sets of parameters satisfying the relationship while giving the desired target values. Assume $a_1 = a_2 = a_3 = 3$ and $K=1$, the expected losses depicted by the three estimation methods listed as follows:

Table 5-12 Expected loss depicted by the three estimation methods

Estimation Method	Method of Moments	MLE	Percentile Matching
Expected Loss	0.351154	0.28814	0.368964

As mentioned before, there are many possible sets of parameters which may be adopted as the parameters for our IDLF. According to the calculation in the preceding paragraph in this subsection, the three parameters are found to be equivalent. Expected loss can be viewed as a function of the parameters and the following plot is therefore obtained:

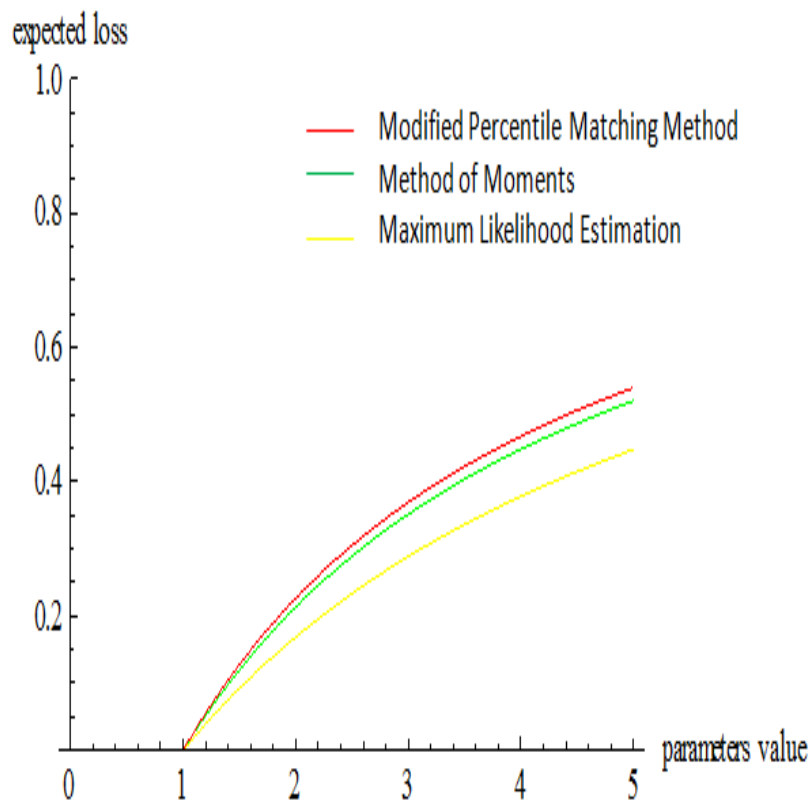


Figure 5-26 Expected loss vs. values of parameters

As shown in Figure 5-26, maximum likelihood estimation gives the smallest expected loss among the three methods of estimation. In other words, the most

conservative depiction of financial loss, in this case, is given by the maximum likelihood estimates.

An assumption of all parameters of the IDLF be equal to three is made previously in this subsection. Leung and Spiring (2002) demonstrated how to select the parameters of an IBLF when the exact measurement where one half of the maximum loss would result is known. In the remaining of this sub-section, a similar approach is used.

Assume the product designer agrees that half of the maximum loss would be obtained when $x = \frac{1}{6}$ and $y = \frac{1}{6}$. In other words, aside from the given tolerances, an IDLF will be formed, which will depict half of maximum loss when the measurements obtained are approximately $x = \frac{1}{6}$ and $y = \frac{1}{6}$. The Dirichlet probability density function, if $a_1 = a_2 = a_3 = a$, can then be expressed as:

$$\begin{aligned}
 L &= 1 - \left(\frac{x}{T_x}\right)^{a-1} \left(\frac{y}{T_y}\right)^{a-1} \left(\frac{1-x-y}{1-T_x-T_y}\right)^{a-1} \\
 &= 1 - \left(\frac{1/6}{1/3}\right)^{a-1} \left(\frac{1/6}{1/3}\right)^{a-1} \left(\frac{2/3}{1/3}\right)^{a-1} \\
 &= 1 - 2^{1-a} = \frac{1}{2}
 \end{aligned} \tag{5.9}$$

which can be solved easily and $a = 2$ is obtained. A plot showing the relationship between the value of parameter and the expected loss under the specified setting can be seen below:

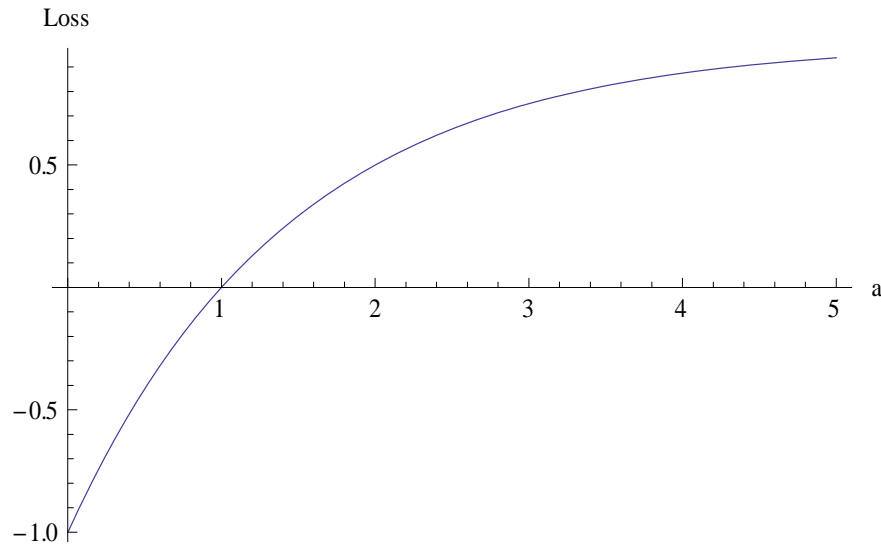


Figure 5-27 Relationship between the estimator and loss

Based on this finding, the parameters of the proposed IbDLF are formed as

Dirichlet $[a_1 = a_2 = a_3 = 2]$ and the loss function becomes:

$$L = 2 \left[1 - \left(\frac{x}{T_x} \right)^{2-1} \left(\frac{y}{T_y} \right)^{2-1} \left(\frac{1-x-y}{1-T_x-T_y} \right)^{2-1} \right], \quad (5.10)$$

where the unit cost of one such fastener is US\$2.

The loss function can be illustrated in the following plot:

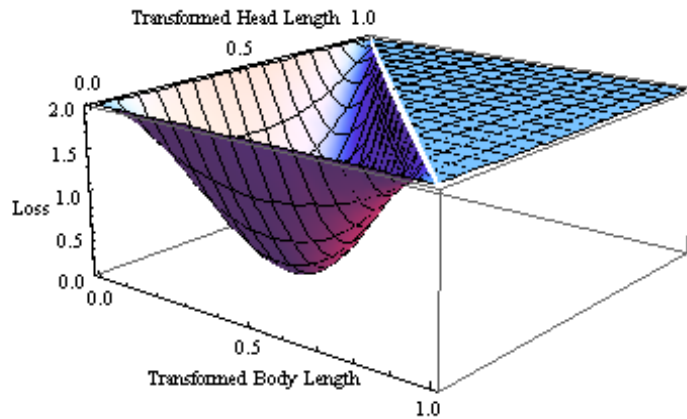


Figure 5-28 Plot of IDLF with respect to the two transformed measurements

Therefore, the expected loss can be calculated by integrating the multiple of the loss function and the conjugate function obtained above, the following results are obtained:

Table 5-13 Estimators for the conjugate function and the IDLF

		Method of Moments	MLE	Percentile Matching Method
Conjugate	α_1	5.3257	6.86421	5.5057
	α_2	4.65618	6.04221	4.83618
	α_3	3.71277	5.02687	3.51277
IDLF	a_1	2	2	2
	a_2	2	2	2
	a_3	2	2	2
Expected Loss		0.425886	0.336538	0.452096

The three methods give approximately the same expected loss at between US\$0.33 and US\$0.45 when the maximum loss of one pressed-in threat insert. This relatively low expected loss may result from the small sample standard deviations found in the

two variants. In the following section, financial loss is once again being depicted using an IbNLF.

5.2.2.4 Expected Loss Calculated by IbNLF

In order to show the difference in maximum loss obtained from an IbNLF and an IDLF, the same sample of pressed-in threaded inserts is being used. Since the three methods of computing estimates of the conjugate function give relatively close results, only the estimates of MLE are used in this sub-section. With calculation similar to previous sub-sections, following estimates are obtained:

Table 5-14 Table for the estimators obtained by MLE

\bar{x}	\bar{y}	S_x	S_y	r
12.7167	1.002	0.0381517	0.034533	-0.81716

Peacock's two dimensional KS test gives a maximum difference of 0.119006, which results in a p -value of almost 1, indicating no evidence is found to reject the proposed distribution.

Since X and Y seem to have a strong positive relationship, three different correlation coefficients of -0.95, 0 and 0.95 are considered to build the IbNLF. The expected losses due to different scenarios are listed in the following table:

Table 5-15 Expected loss depicted with three different scenarios

T_x	T_y	S_x	S_y	r	Expected Loss
12.7	1	0.04	0.04	0	0.000562065
12.7	1	0.06	0.06	-0.95	0.000693885
12.7	1	0.03	0.03	0.95	0.0011366

Based on the result found in this subsection and the previous subsection, it can be seen that the expected losses depicted by IbNLF are much lower than those depicted by IbDLF.

A few points are worth-mentioning regarding the inspection method suggested here. By fixing the target values and upper limits being fed into the bivariate normal function, the standard deviations may be varied by the range of potential standard deviations used in plotting. It can be illustrated by the following plot where a bivariate normal function with standard deviations between 0 and 2 is plotted.

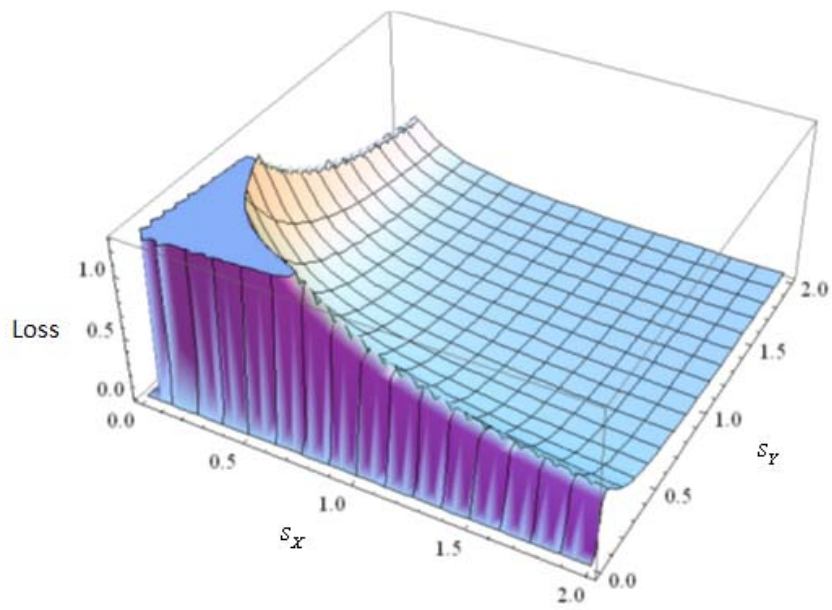


Figure 5-29 Plot of loss versus values of estimates when correlation is assumed to be 0

By inspecting Figure 5-29, it may be suggested that having standard deviations of (1, 1) would end up with a bivariate normal function similar to the following:

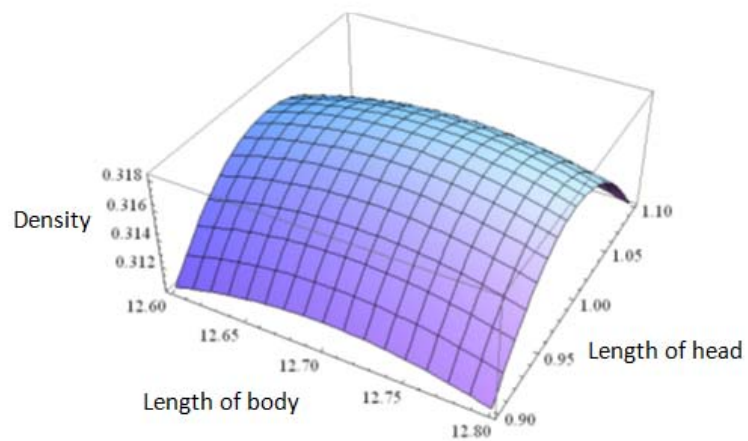


Figure 5-30 The pdf with the estimates obtained

It can be observed, although it is not exactly a loss function, when the tolerance limits are met, the probability density function does not approach to zero. Thus, the above probability density function does not satisfy the requirement of inverted probability loss functions.

5.2.2.5 Expected loss calculated by the two univariate IBLF

Aside from comparing the expected loss depicted by the IDLF and the IbNLF, comparison between the IDLF and the two IBLFs give an interesting result.

Based on the transformed data obtained in the previous sections, the method of moments estimates and maximum likelihood estimates are obtained. The values of estimators can be found in the following table:

Table 5-16 Estimators of the transformed lengths obtained by method of moments and MLE

	Transformed Body		Transformed Head	
	MME	MLE	MME	MLE
a	5.3257	5.25193	5.41738	5.26459
b	8.36896	8.27399	10.5161	10.229

Since there is no other constraint in building the Inverted Beta Loss Function, the same value of $1/6$ is chosen as the point where half of the maximum loss will be depicted as assumed in constructing the IDLF. Consider a standard Beta Distribution with parameters a and b , in other words, $Beta(a, b)$. Based on the fact that the target value is $1/3$, it can be shown that the relationship between the two estimators is $b = 2a - 1$. Adopting this assumption will grant the following function and graph:

$$L = 1 - 2^{3-a-2(2a-1)} 5^{2a-2} = 0.5 \quad (5.11)$$

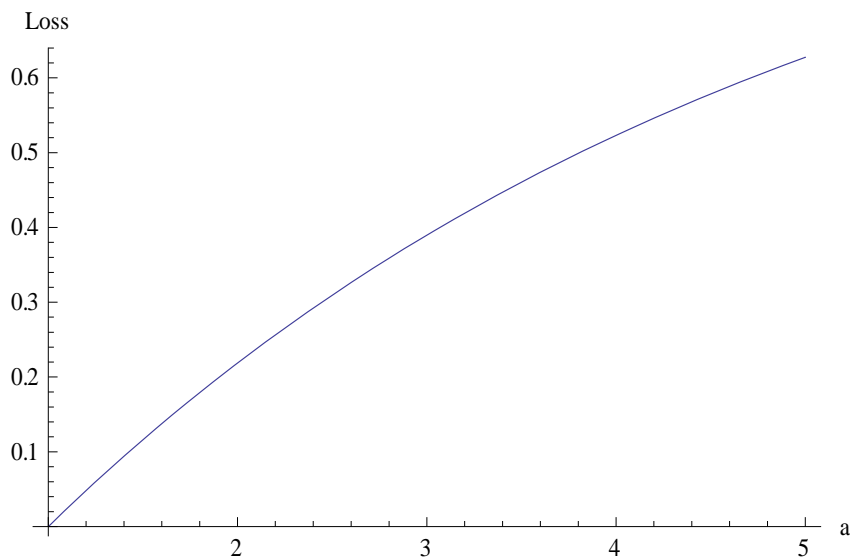


Figure 5-31 Relationship between loss and values of estimators

Letting the equation to be equivalent to 0.5, parameter a is estimated with a value of 3.80785, which gives the set of estimators to be [3.80785, 6.6157].

The expected losses calculated by the estimates obtain previously are shown in the following table.

Table 5-17 Expected loss depicted by MME and MLE for the two individual quality characteristic

	Transformed Body		Transformed Head	
	MME	MLE	MME	MLE
Expected Loss	0.237103	0.238325	0.195292	0.199436

In this section, expected losses of deviations of two individual quality characteristics were shown. Adding the expected losses of transformed body length and transformed head length together, would yield similar results as depicted by the IDLFs. Slight differences are observed, which can be explained by correlation between the two quality characteristics.

5.2.2.6 Brief comments about the IDLF

In this chapter, demonstration on how to apply an IbNLF and IDLF are done. A key point worth mentioning is certainly the domain of the quality characteristics. The tolerance limits given by the product engineer usually forms a rectangular region with finite values. This region causes complication for applying either an IbNLF or an IDLF. An IbNLF, based on a normal distribution, has a domain of \mathfrak{R}^2 and the IDLF has a domain of a triangular region. Both loss functions have a domain which

are not rectangular. This problem increases the complexity of obtaining parameters for either IPLF.

In terms of domain, the IDLF is more restrictive than the IbNLF. A triangular tolerance region is not common. As demonstrated in the application of the pressed-in threaded inserts, an IDLF was chosen because the addition of two quality characteristics must be less than or equal to a certain limit.

The setup of Chi-square goodness-of-fit test for a Dirichlet distribution with two variables seems to be easier than a bivariate Normal distribution. This is because: i) the bivariate Normal distribution has more parameters to be estimated than the Dirichlet distribution, and ii) the setup of sub-regions for a bivariate Normal distribution is arbitrary as it is hard to compose sub-regions with equal probability or equal region.

5.3 Choosing of Inverted Probability Loss Functions and Conjugate Functions

Solder paste and fasteners have been studied to demonstrate the feasibility of depicting financial losses by either an IbNLF or an IDLF. Discussion about how to select the appropriate IbPLF and the conjugate function will be discussed in this subsection.

To the author's best knowledge, up to now there are only two IbPLs, including the IDLF developed in this project. The Dirichlet distribution, having a triangular domain, in the 2-variable case can only be used when the two quality characteristics are also having a triangular domain. In other words, an IDLF is only useful when the sum of two measurements is less than a certain limit.

A bivariate Normal distribution has \mathcal{R}^2 as its domain. This feature gives flexibility to an IbNLF as any tolerance limits with two quality characteristics must be a subset of its domain. As showed in the application of pressed-in threaded inserts, financial loss can also be depicted by an IbNLF. Although an IbNLF is very flexible, it causes

a few problems: i) difficulty of obtaining parameters for the IbNLF, and ii) the rectangular tolerance limits is only a subset of the domain of the IbNLF.

Choosing a conjugate function is another important aspect in depicting financial loss using an IbPLF. For univariate IPLFs, closed forms may be easily obtained, even when the IPLF and the conjugate function do not come from the same probability density function. When the number of variables increases, closed forms are often impossible to obtain when the choice of IbPLF and the conjugate function come from two different bivariate probability density functions and numerical integration is often the only solution. It is desirable that the IbPLF built by a bivariate probability density function and the conjugate function belong to the same family of distribution. Selecting the same probability density function is easier, in terms of calculating the expected loss as both functions have the same domain.

CHAPTER 6 CONCLUSION

Globalization creates enormous economic opportunities to the world. On the other hand, fierce global competition has arisen for manufacturers. For a manufacturer to survive, and to make profit, its product must be precise and durable. Loss functions are used to evaluate and depict financial losses due to variation from the target values specified by the product engineer. Manufacturers will be better off if they can make good use of loss functions to assess their risk properly and appropriately.

Goalpost theorem has been the dominating guideline in the field of quality control until Dr. Taguchi introduced his quadratic loss function (Taguchi, 1986). The quadratic loss function satisfies the need of depicting financial loss due to manufacturing deviation within the tolerance limits. This is a step forward, since, according to the goalpost theorem, financial loss only occurs when the measured quality of a product is beyond the tolerance limits.

In this study, a multivariate inverted Dirichlet loss function has been developed. Reviews of various inverted probability loss functions (IPLFs), including the

multivariate and bivariate inverted normal distribution, properties of inverted probability loss functions, have been elaborated. Process data have been used to show the feasibility of bivariate IPLFs.

Works have also been done on simulation, modeling and justification of the proposed models. Three popular methods of simulating Dirichlet distributions have been studied, and they are the acceptance and rejection approach, conditional approach and the gamma factors approach. There may be situations which the variables are additive, but data may be difficult or impossible to be collected, generating random Dirichlet variates may be necessary.

Estimates regarding the simulated and real data, have been obtained with the method of moments, maximum likelihood estimation, alongside with the modified percentile matching method proposed in this thesis. Bivariate Chi-square goodness-of-fit test is the main tool used in this study to verify the adequacy of models with estimates obtained by the three methods mentioned above. Tables have been set up to illustrate, based on the same set of parameters for random variates generation, the three methods of simulation are often similar to each another. One thousand random variates have been simulated by each simulation method, with these generated

sample points, different estimation methods are performed and verified by the bivariate Chi-square goodness-of-fit test and Peacock's two-dimensional Kolmogorov-Smirnov test for each underlying conjugate distribution. Five trials are repeated for each simulation method.

Model verification is critical in developing an IPLF. This task is tackled by the Chi-square goodness-of-fit test. In this project, it is suggested to dissect the domain of a Dirichlet distribution into triangles of equal area. To verify this idea, a two-dimensional Kolmogorov-Smirnov test has been performed. Since 1983, various scholars have worked on developing two-dimensional K-S tests including Peacock (1983), Fasano and Franceschini (1987), Cooke, and Chan (2008). In this study, the bivariate Kolmogorov-Smirnov test developed by Peacock (1983) has been adopted because it has the longest history and it is almost distribution-free. In the two demonstrated examples of fasteners, Peacock's 2-dimensional Kolmogorov-Smirnov test resulted in a slightly smaller p -value for the self-tapping threaded inserts and a similar p -value for the pressed-in threaded inserts. Indicating the proposal of forming cells based on the contour plot is reasonable.

One of the contributions of this project lies in the proposal of the modified percentile matching method. Percentile matching method is often an afterthought in modeling. Insurance companies used it to fit models because it takes care of extreme values; however, occasionally it does not provide enough information for data in the middle part of the distribution. The approach suggested in this study started with a slightly different rationale: matching up the number of observed points to the number of expected points using integration within a rectangle with vertices containing the origin and these specific selected points. The selection criterion is based on rectangles being drawn contained the most number of sample points. The set of estimates obtained by method of moments is used as the starting point. According to tables 3-6 to 3-14, the three sets of estimates are usually similar to each other. With the starting point fixed, a cube is formed, for the case of bivariate data, with vertices containing plus or minus ten percent of the starting point. This cube, is then dissected into smaller cubes, depends on how precise the investigation is to be done. The closest integrated value to the target value done by counting is then selected to be the set of estimates. The beauty of this method is that, as long as integration or numerical integration can be done, a set of estimates can be obtained. The downside of this method is obviously a much longer computing time required. Results based on simulated data have shown that, the set of estimates obtained by the percentile

matching method usually give the same conclusion of whether to reject a proposed model as the other two methods of estimates. Based on the current set up of regions where expected and observed frequencies have to be counted, the result seems acceptable. According to the real data, p -values obtained for the three methods of estimations are indeed competitive, which is encouraging for the modified percentile matching method. More work is definitely required to study this modified percentile matching method as it remains one of the future research interest of the author.

Extension in IPLFs to multi-dimensional has also been explored. Spiring (1993) introduced the first INLF and IbNLF. In this project, the bivariate INLF is used to depict the financial loss of solder paste and self-tapping threaded inserts. According to the sample data, both examples suggest satisfactory result in terms of modeling and an expected loss is depicted based on the IPLF formed.

Upon the completion of this thesis, the author's research interest remains in developing ImPLFs based on other probability distributions including gamma and exponential. The author is keen to dig deeper in two areas discussed in this thesis: i) percentile matching method and ii) Chi-square goodness-of-fit test in higher

dimension. In this thesis, the feasibility of these two approaches is shown; however, whether they are one-time wonder or universal true remains a further investigation.

REFERENCES

Chan, I. Parallelizing a 2D Kolmogorov-Smirnov Statistic. [online] available at:

<<http://web.mit.edu/ianchan/KS2D/ks2dpres.ppt>>

Cooke, A. The muac algorithm for solving 2d KS test. [online] available at:

<<http://www.acooke.org>>

Crosby, P.B. (1979). *Quality is Free: the Art of Making Quality Certain*, McGraw-Hill, New York.

Crosby, P.B. (1996). *Quality is Still Free: Making Quality Certain in Uncertain Times*, McGraw-Hill, New York.

Cuadras, C.M. (2002). On the covariance between functions. *Journal of Multivariate Analysis*, 81, 19 – 27.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

Fasano, G. and Franceschini, A. (1987). A multidimensional version of the Kolmogorov-Smirnov test. *Mon. Not. R. astr. Soc.*, **225**, 155 – 170.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*, MacMillan.

Ghorbel, M. (2010). On the inverted Dirichlet distribution. *Communications in Statistics: Theory and Methods*, Vol. 39, 1, 21 – 37.

- Hazewinkel, M. (1988). *Encyclopaedia of Mathematics: an Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia"*
Dordrecht, Boston.
- Hoeffding, W. (1940). *Masstabinvariante Korrelationstheorie*. Schriften Math. Inst. Univ. Berlin, 5, 181 – 233.
- Johnson and Kotz (1970) *Continuous Univariate Distributions – 2*, John Wiley, New York.
- Johnson, Kotz and Balakrishnan (2000). *Continuous Multivariate Distributions, Volume 1: Models and Applications*, John Wiley, New York.
- Johnson, M.E. (1987). *Multivariate Statistical Simulation*, Wiley, New York.
- Klugman, Panjer and Willmot (1998). *Loss Models: From Data to Decisions*, Wiley, New York.
- Kulkarni S. and Prybutok V. (2004). Process investment and loss functions: Model and analysis. *European Journal of Operational Research*, Vol. 157, 1, 120 – 129.
- Kulkarni S. (2008a). Loss-based quality costs and inventory planning: General models and insights. *European Journal of Operational Research*, Vol. 188, 2, 428 – 449.

Kulkarni S. (2008b). On a multi-product model of lot-sizing with quality costs.

International Journal of Production Economics, Vol. 112, 1002 – 1110.

Leon, R.V., and Wu, C.F.J. (1992) A theory of performance measures in parameter design. *Statist. Sinica*, 2(2), 335 - 337.

Leung, B.P.K. (1999). PhD Thesis, *Contributions to Industrial Statistics*, University of Manitoba, Canada.

Leung, B.P.K. and Spiring, F.A.(2002). The inverted beta loss function: properties and applications. *IIE Transactions*, 34, 1101 – 1109.

Leung, B.P.K. and Spiring, F.A. (2004). Some properties of the family of inverted probability loss functions. *Quality Technology and Quantitative Management*, 1, 125 – 147.

Leung, B.P.K. and Tsoi, F. (2009). Application of inverted probability loss functions on a product delivery problem. *Proceedings of IWSSSCM 2009*, 335 – 344.

Narayanan, A. (1991). Maximum likelihood estimators of the parameters of the Dirichlet distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 40(2), 365 – 374.

- Pan, J.N. and Pan, J. (2006). A comparative study of various loss functions in the economic tolerance design. *2006 IEEE International Conference on Management of Innovation and Technology*, 783 – 787.
- Pan, J.N. and Wang, J.H. (2000). A study of loss functions for product interference analysis. *Industrial Engineering Research*, **2(1)**, 80 – 100.
- Peacock, J.A. (1983). Two-dimensional goodness-of-fit testing in astronomy. *Mon. Not. R. astr. Soc.*, 202, 615 – 627.
- Pearson (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine*, Series 5, 50 (302), 157 – 175.
- Ronning, G. (1989). Maximum likelihood estimation of Dirichlet distributions. *Journal of Statistical Computation and Simulation*, 32(4), 215 – 221.
- Ross, P.J. (1996). *Taguchi Techniques for Quality Engineering: Loss Function, Orthogonal Experiments, Parameter and Tolerance Design*, McGraw-Hill, New York.
- Sen, P.K. (1994). The impact of Wassily Hoeffding's research on nonparametrics. *The collected works of Wassily Hoeffding*, Springer-Verlag, New York, 29 – 55.

- Spiring, F.A. (1993) The reflected normal loss function. *Canadian Journal of Statistics*, 3, 321 – 330.
- Spiring, F.A. and Yeung, A.S. (1998) A general class of loss functions with industrial applications. *Journal of Quality Technology*, 30, 152 – 162.
- Sun, F., Laramee, J. and Ramberg, J. (1996) On Spiring's inverted normal loss function. *Canadian Journal of Statistics*, 2, 241 – 249.
- Taguchi, G. (1986) *Introduction to Quality Engineering: Designing Quality into Products and Processes*, Kraus, White Plains, NY.
- Taguchi, G., Elsayed, E.A., and Hsiang, T. (1989) *Quality Engineering in Production Systems*, McGraw-Hill, New York.
- Tiao, G. and Cuttman, I. (1965). The inverted Dirichlet distribution with applications. *Journal of the American Statistical Association*, 60, 793 – 805.
- Tukey, J.W. (1977) *Exploratory data analysis*. Addison-Wesley, Reading, M.A.
- Wikipedia (2011). Dirichlet distribution. [online] Available at: http://en.wikipedia.org/wiki/Dirichlet_distribution.

APPENDIX

Mathematica files used to build an IDLF

Calculating moments of the raw data:

```
Moments[a_]:=Block[{i}, i=1;x1=0;x2=0;y1=0;y2=0; EX=0;EX2=0;EY=0;EY2=0;
```

```
While[i<=60,
```

```
    x1=x1+a[[i,1]];
```

```
    x2=x2+a[[i,1]]^2;
```

```
    y1=y1+a[[i,2]];
```

```
    y2=y2+a[[i,2]]^2;
```

```
    i++;];
```

```
EX=x1/60;
```

```
EX2=x2/60;
```

```
EY=y1/60;
```

```
EY2=y2/60;
```

```
];
```

Calculating the parameters by method of moments:

```
MME[a_,b_,c_,d_]:=Block[{i},
```

```

MMEa1=((a-b)*a)/(b-a^2);

MMEa2=(a-b)*c/(b-a^2);

MMEa3=(a-b)*(1-a-c)/(b-a^2);

Print["MMEa1: ", MMEa1];

Print["MMEa2: ", MMEa2];

Print["MMEa3: ", MMEa3];

];

```

Calculating the parameters by maximum likelihood estimators

```

mle[a_,b_,c_,d_]:=Block[{i}, (* function to calculate the maximum likelihood
estimators *)

```

```

n=60;

```

```

Iij={n*PolyGamma[1,a],n*PolyGamma[1,b], n*PolyGamma[1,c]};

```

```

Dia=DiagonalMatrix[Iij];

```

```

G={{-n*PolyGamma[1,a+b+c],-n*PolyGamma[1,a+b+c],-

```

```

n*PolyGamma[1,a+b+c]},{-n*PolyGamma[1,a+b+c],-

```

```

n*PolyGamma[1,a+b+c],-n*PolyGamma[1,a+b+c]},{-

```

```

n*PolyGamma[1,a+b+c],-n*PolyGamma[1,a+b+c],-

```

```

n*PolyGamma[1,a+b+c]}}};

```

```

V=Inverse[Diag+G];

Print["V: ", V];

i=1;

templogx=0;

templogy=0;

templogz=0;

data=d;

logdata=N[Log[data]];

While[i<=n,

    templogx=templogx+logdata[[i,1]];

    templogy=templogy+logdata[[i,2]];

    templogz=templogz+Log[1-d[[i,1]]-d[[i,2]]];

    i++;

];

u={a,b,c};

g={ {n*PolyGamma[0,a+b+c]-

n*PolyGamma[a]+templogx},{n*PolyGamma[0,a+b+c]-

n*PolyGamma[0,b]+templogy},{n*PolyGamma[0,a+b+c]-

n*PolyGamma[0,c]+templogz} };

```

```

Print["g: ", g];

u1=u+V.g;

Print["first trial: ", u1];

mlea1=u1[[1,1]];

mlea2=u1[[2,1]];

mlea3=u1[[3,1]];

];

```

Estimators calculated by the percentile matching method:

```
pm[l_,m_,n_]:=Block[{i,j,k}, (* function of percentile matching *)
```

```
list={ }; list2={ };
```

```
e=l-.2; a=1/3; b=1/3; c=2/3; d=2/3; f=m-.2; g=n-.2; i=0; j=0; k=0;
```

```
While[i<20,
```

```
    j=0;
```

```
    While[j<20,
```

```
        k=0;
```

```
        While[k<20,
```



```

p=NIntegrate[Gamma[e+i/50+f+j/50+g+k/50]/(Gamma[e+i/5
0]*Gamma[f+j/50]*Gamma[g+k/50])*x^(e+i/50-1)*y^(f+j/50-1)*(1-
x-y)^(g+k/50-1),{x,0,a},{y,0,b}];

```

```

q=NIntegrate[Gamma[e+i/50+f+j/50+g+k/50]/(Gamma[e+i/5
0]*Gamma[f+j/50]*Gamma[g+k/50])*x^(e+i/50-1)*y^(f+j/50-1)*(1-
x-y)^(g+k/50-1),{x,0,a},{y,0,d}];

```

```

r=NIntegrate[Gamma[e+i/50+f+j/50+g+k/50]/(Gamma[e+i/50
]*Gamma[f+j/50]*Gamma[g+k/50])*x^(e+i/50-1)*y^(f+j/50-1)*(1-x-
y)^(g+k/50-1),{x,0,c},{y,0,b}];

```

```

AppendTo[list,{p,q,r,e+i/50,f+j/50,g+k/50}];

```

```

AppendTo[list2,{p,q,r}];

```

```

k++;];

```

```

j++;];

```

```

i++;];

```

```

distance[list];

```

```

];

```

Calculating the distance between a point and the desired value:

```
distance[temp_]:=Block[{i}, i=1;

dtemp=0; dlist={ }; loc=0; Dim=8000;

While[i<=Dim,

dtemp=(temp[[i,1]]-counter1/60)^2+(temp[[i,2]]-
counter2/60)^2+(temp[[i,3]]-counter3/60)^2;

AppendTo[dlist,{ dtemp,i }];

i++;];

i=1;

bench=dlist[[i,1]];

While[i<=Dim,

If[dlist[[i,1]]<bench,{ bench=dlist[[i,1]],loc=i }];

i++;];

Print[list[[loc]],dlist[[loc]];

pma1=list[[loc,4]];

pma2=list[[loc,5]];

pma3=list[[loc,6]];

];
```

Counting number of observations inside each sub-region for the percentile matching

method:

```
Counter[a_]:=Block[{i},
```

```
  i=1;aa=1/3;bb=1/3;cc=2/3;dd=2/3;counter1=0;counter2=0;counter3=0;
```

```
  While[i<=60,
```

```
    If[a[[i,1]]<aa && a[[i,2]]<bb,counter1++];
```

```
    If[a[[i,1]]<aa && a[[i,2]]<dd, counter2++];
```

```
    If[a[[i,1]]<cc && a[[i,2]]<bb, counter3++];
```

```
    i++;
```

```
  ];
```

```
  Print[counter1];
```

```
  Print[counter2];
```

```
  Print[counter3];];
```

Counting number of observations for the Chi-square goodness-of-fit test:

```
count[a_]:=Block[{i}, (*function to count number of points in specific regions for
```

```
chi-square goodness-of-fit test *)
```

```
  i=1; counter1=0; c11=0;c12=0;c13=0;c14=0;c15=0;c16=0;c17=0;
```

```
  c21=0;c22=0;c23=0;c24=0;c25=0;c31=0;c32=0;c33=0;c4=0;
```

counter2=0;counter3=0;counter4=0;

While[i<=60,

If[a[[i,1]]<=1/4, counter1++;];

If[a[[i,1]]<=1/4 && a[[i,1]]+a[[i,2]]<=1/4, c11++;];

If[a[[i,1]]<=1/4 && a[[i,1]]+a[[i,2]]>1/4 && a[[i,2]]<=1/4, c12++;];

If[a[[i,1]]<=1/4&& a[[i,1]]+a[[i,2]]<2/4 && 1/4<a[[i,2]]<=2/4, c13++;];

If[a[[i,1]]<=1/4 && a[[i,1]]+a[[i,2]]>2/4 && 1/4<a[[i,2]]<=2/4, c14++;];

If[a[[i,1]]<=1/4&& a[[i,1]]+a[[i,2]]<=3/4 && 2/4<a[[i,2]]<=3/4, c15++;];

If[a[[i,1]]<=1/4 && a[[i,1]]+a[[i,2]]>3/4 && 2/4<a[[i,2]]<=3/4, c16++;];

If[a[[i,1]]<=1/4 && a[[i,2]]>3/4, c17++;];

If[1/4<a[[i,1]]<=2/4,counter2++;];

If[1/4<a[[i,1]]<=2/4&& a[[i,1]]+a[[i,2]]<=2/4 && a[[i,2]]<=1/4, c21++;];

If[1/4<a[[i,1]]<=2/4 && a[[i,1]]+a[[i,2]]>2/4 && a[[i,2]]<=1/4, c22++;];

If[1/4<a[[i,1]]<=2/4&& a[[i,1]]+a[[i,2]]<=3/4 && 1/4<a[[i,2]]<=2/4, c23++;];

If[1/4<a[[i,1]]<=2/4 && a[[i,1]]+a[[i,2]]>3/4 && 1/4<a[[i,2]]<=2/4, c24++;];

If[1/4<a[[i,1]]<=2/4 && a[[i,2]]>2/4, c25++;];

If[2/4<a[[i,1]]<=3/4,counter3++;];

If[2/4<a[[i,1]]<=3/4&& a[[i,1]]+a[[i,2]]<=3/4 && a[[i,2]]<=1/4, c31++;];

If[2/4<a[[i,1]]<=3/4 && a[[i,1]]+a[[i,2]]>3/4 && a[[i,2]]<=1/4, c32++;];

```

If[2/4<a[[i,1]]<=3/4 && a[[i,2]]>1/4, c33++];

If[a[[i,1]]>3/4,counter4++];

If[a[[i,1]]>3/4, c4++];

i++;

];

Print["x<1/4: ", counter1];

Print[c11, " ", c12, " ", c13, " ", c14, " ", c15, " ", c16, " ", c17];

Print[c11+c12+c13+c14+c15+c16+c17];

Print["1/4<x<2/4: ",counter2];

Print[c21, " ", c22, " ", c23, " ", c24, " ", c25];

Print[c21+c22+c23+c24+c25];

Print["2/4<x<3/4: ", counter3];

Print[c31, " ", c32, " ", c33];

Print[c31+c32+c33];

Print[c4];

Print["y<1/3: ", c11+c12+c21+c22+c31+c32+c4];

Print["1/3<y<2/3: ", c13+c14+c23];

Print["y>2/3: ", c15];

Print["Total: ", counter1+counter2+counter3+counter4];];

```