INVERTED PROBABILITY LOSS FUNCTIONS AND ITS APPLICATIONS

YAU, KA HO

M. Phil.

The Hong Kong Polytechnic University

2015

HONG KONG POLYTECHNIC UNIVERSITY

DEPARTMENT OF APPLIED MATHEMATICS

RESEARCH THESIS

Inverted Probability Loss Functions and Its Applications

Ka-Ho YAU

A thesis submitted in partial fulfilment of

the requirements for the degree of Master of Philosophy

June 2013

Certificate of Originality

I hereby declare that this 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 the text.

____________________ (Signed)

YAU, Ka Ho (Name)

Abstract

In most statistical and decision problems, nearly no attention is paid to the precise mathematical form of the loss function. However, the choice of a particular loss function seriously affects the resulting inferences and estimations. This dissertation investigates a general class of loss functions based on the reflection or inversion of a probability density function, Inverted Probability loss function, which was proposed by Spiring and Yeung (1998). We modified the Inverted Probability loss function to be a more generalisation of the original one. To the best of my knowledge and belief, it is the first time to establish such results in the literature.

We firmly advocate that there are some novelties in the Inverted Probability Loss Functions and there are even more applications when applying them. In this report, we show the broad coverage and the flexibility of the Loss Functions to make a more robust expected loss.

Acknowledgements

First of all, I would like to express my deepest gratitude to my chief supervisor, Dr. Bartholomew P.K. Leung, for giving me an absolute freedom to have the opportunity to have this independent study. He also permitted me to take as many courses as we plan, including the courses in other universities in Hong Kong. Without this golden opportunity, I may not be able to stimulate the ideas and have a fruitful discussions with others.

Secondly, I would also like to give my sincere appraisal to those friends and the professors in different universities to let me open my eyes to the frontiers of different aspects of statistics, especially Professor Mourad Ismail in the City University of Hong Kong, Professor Bing-yi Jing and Prof Shiqing Ling in the Hong Kong University of Science and Technology. Their answers to my question during class are very inestimable to provide me some understandings and insights in the research level. I also need to send my special thanks to Professor Heung Wong in our Department of Applied Mathematics for strengthening my doubts to the original thoughts on this topic.

This dissertation is in the fifth version and catastrophic in the research journey. Nevertheless, I need to thank my colleagues to throw away all my hardcopy raw materials and the original draft. I also need to show my courtesy to the technician formatting the whole hard disk during my final stage instead of just the system partition while upgrading, but he told me the opposite. As Edison said, ‘there is a great value in disaster’. I have to restart the writing from scratch and a lot of mistakes are eliminated.

Last but not least, I also need to say thanks to all staff members in the Department of Applied Mathematics and all participants in the course ‘Writing in the Sciences’ by Dr Kristin L Sainani on the platform Coursera for their friendship, feedback and assistance in all ways. Under no circumstances would I survive in the dynamic circle without having their advice from our beloved staff members. I also have learnt many varied views and knowledge from the platform provided by Coursera of Stanford University, which revolutionises the limited scope of the tertiary education. For instance, on this platform, I can access to the views and experiences by Professor Bradley Efron on writing a scientific manuscript. These invaluable experiences sum up to contribute an excellent learning experience being in lack before.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This first chapter provides a little background and motivation in the first section, then exhibits a little information about my main contributions and finally manifests the organisation of this whole report.

1.1 Background and Motivation

Decision makers in manufacturing industries pay high attention to quality assurance. As a result, the use of statistics for accessing quality receives even larger attention when Taguchi (Taguchi, 1986) proposed his quality management philosophy and strategy. In decision theory and quality assurance field, loss functions are used to reflect the monetary loss or economic loss caused by the deterioration of the product characteristics from the target quality.

However, Berger (Berger, 1985) even emphasised that the loss function should be bounded and concave, because the loss function also mimics the negative of the utility, whereas the squared-error loss, Taguchi quadratic loss in quality control, or absolute error loss is unbounded and even disturb the convexity. Spiring and Yeung

(1998) proposed a framework to cover all the unbounded losses from some observations with some applications in quality control. They also called the losses made in this framework as The Inverted Probability Loss Functions. All such loss functions enjoy the boundedness and preserves the convexity due to the requirements of the unimodality. Therefore, it should totally supersede the squared-error one; but most do not have the interests in studying it and some even choose a conservative view to prefer the traditional quadratic Loss Functions to the parametric Inverted Probability Loss Function.

Therefore, this reason already gives a sounded motivation to have a research about this Inverted Probability Loss Functions to understand how this concept is unique and interesting.

1.2 Data set for illustrations

Since it is required to have some applications for realising the results in this report, a data set is chosen for this purpose. To prevent self-plagiarism and need to re-use this data many times, we now discuss about the data set here and refer it back when necessary.

The following data set is from Leung and Spiring (2002, 2004). It is a realisation of the random variable for the perforation pull strength. The data set is as follows:

Some further background information was also provided in Leung and Spiring (2004). The aim for a lottery ticket seller was to sell the lottery tickets as many as possible, so it wanted to sell the tickets via vending machines as well such that any buyers could buy the tickets in a convenient way. For putting the tickets in

40.6	47.6	49.5	52.8	45.0	51.6	48.5	58.3	46.5	53.5
48.9	58.3	42.0	41.0	47.3	47.5	47.0	54.5	47.7	44.8
47.5	47.0	41.7	54.8	42.6	56.5	52.4	55.9	42.2	52.6
50.5	49.7	48.6	58.6	43.7	53.7	47.6	55.0	45.0	54.6
55.0	46.6	51.8	51.0	46.2	53.8	56.9	48.6	47.6	44.0
49.4	53.7	44.2	52.0	44.5	48.1				

Table 1.1: The pull strength data set

a vending machine, the volume should be as large as possible so the cost, for instance, resupply cost, manual cost and transportation cost, was minimised. Inside a vending machine, the tickets were packed, folded and stacked in columns. The vending machine had to recognise that the buyer had inserted sufficient funds and dispensed the tickets with the same face value as the funds inserted through the dispensing slot. It also needed to identify certain characteristics such that the tickets were dispensed in full and sustained the force of tearing a ticket. To reduce the force required to tear a ticket, the tickets had to be designed as perforated along the margins and tough enough not to be torn in half.

Therefore, the tickets had to pass through the pull strength test such that the design and the process was optimal. If the pull strength was higher than 60 pounds per square inch (psi), the tickets will not be torn in the perforated margin. However, if the pull strength was lower than 40, another tickets will be pulled as well. In either situations, the vending machine is jammed as the next operation or mechanism is distorted.

From the Table 1.1 and the information provided, the pull strength is the random variable under interest. Although Leung and Spiring (2002, 2004) claimed that this data set follows a Beta distribution, but we have another thought on this data set.

For dealing with a data set, we propose to have an empirical study beforehand and give a summary about the ingredients of this pull strength.

Number of Data	56
Mean	49.401786
Median	48.6
Mode	47.6
Interquartile Range	7.3
Variance	22.519104
Skewness	0.10518813
Excess Kurtosis	0.87615755

Table 1.2: Empirical study of the pull strength

Thus from Table 1.2, the random variable for pull strength is also close to symmetric Gaussian. Further, there is no any constraint that the pull strength must be in the interval of $[40, 60]$, even though the maximum loss is attained. That is, there is still a possibility that the pull strength is beyond the interval but maybe the probability is very small. We also need a visualisation tool for understanding this data set to estimate the density. As a result, the empirical distribution is plotted in the following:

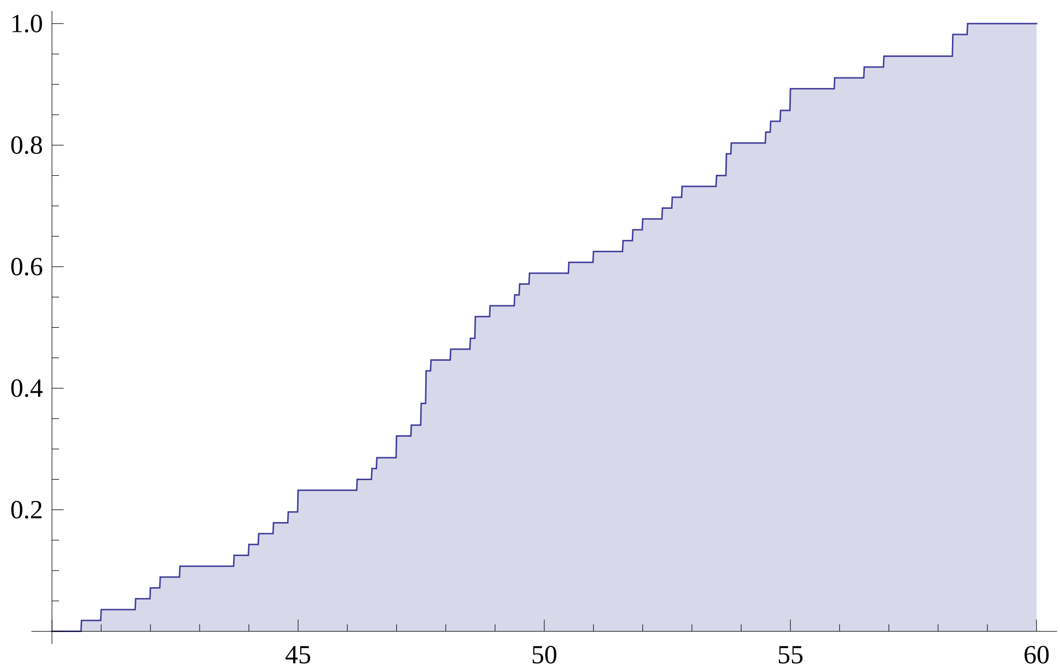
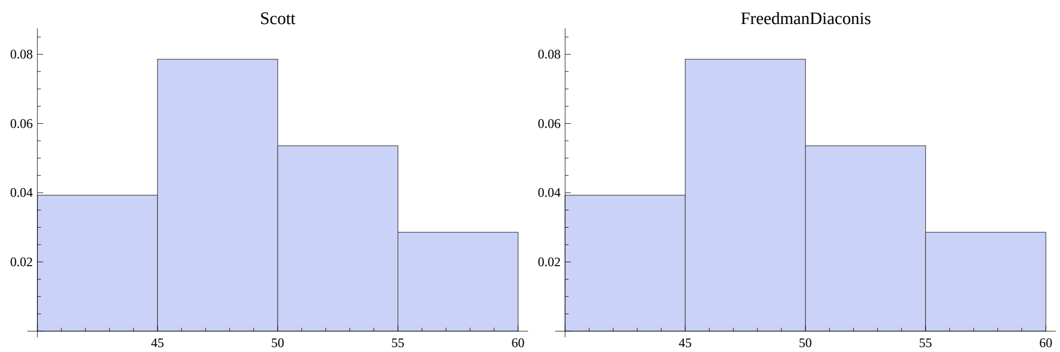


Figure 1.1: Empirical CDF of the pull length

For the density of this data set, we also have to use the worst tool for visualisation, histogram, for which the selection of number of bins is generally problematic. There are two opposing uncertainties in estimating a density from a histogram. One is the coarseness of the histogram and then more number of bins generates a better result. However, the other is the inaccuracy of the height of a bin, so the situation of larger bins are better. These two are very hard to get a balance.

We decide to fit the L_2 theory of univariate histogram in Scott (1992) such that the density estimator is consistent and the mean square error is minimised. We choose the criteria of Scott (1979) and Freedman and Diaconis (1981) as two references to select the bin widths. For Scott method, the bin width is chosen to have a $3.5\hat{\sigma}n^{-1/3}$; while for Freedman–Diaconis method, the method is more robust and the bin width is equal to $2(IQ)n^{-1/3}$, where IQ is the interquartile range. Both histograms from the two methods are illustrated in Figure 1.2.



(a) Scott method (b) Freedman–Diaconis method

Figure 1.2: Histograms from each method

From all the information we obtained, the pull strength is seemed to have a Gaussian distribution. However, the number of data is not very large and we assume that the data set all follow a particular distribution without any exceptional changes. To estimate the parameters of Gaussian distribution, we choose to use a bootstrapped estimate. We sample all the data set with replacement and the resampled values should behave like a particular sample from the original population. Statistics of a sample from the original data set should thus simulate sample statistics for the population. For each resampled data, we use the maximum log-likelihood method to estimate the parameters in Gaussian distribution. The whole processes are repeated 10,000 times and we choose the average value of the parameters.

The result is that the data set may follow $N(49.40, 4.69^2)$. Moreover, we need a diagnostic checking to know whether the result is significant. Similarly, we also choose the bootstrapped test to understand whether the data set is Gaussian distributed and the whole processes are also repeated 10,000 times. Table 1.3 summarises the results of the hypothesis testing.

	Statistic	p-Value
Anderson-Darling	0.473091	0.773567
Cramér-von Mises	0.0820403	0.68014
Jarque-Bera ALM	1.92141	0.290532
Pearson χ^2	12.75	0.237987
Shapiro-Wilk	0.972673	0.232307

Table 1.3: Test statistics and p-value of each test

Therefore, the H_0 that the data set follows Gaussian distribution with mean as 49.40 and variance as 4.69^2 is not rejected at even the significance level of 10%. From now, when we refer to this data, we believe that this data is Gaussian distributed with $N(49.40, 4.69^2)$. However, the true distribution from Leung and Spiring (2002) is a transformed $Beta(2.0994, 2.3184, 40, 60)$. With these two distributions, we will conduct some comparisons in the following chapters.

For the loss, there is no standard at all and as in Spiring and Yeung (1998), ‘after lengthy discussion, a loss ... was agreed upon.’ Therefore, it is on a case-by-case basis. In other words, it is meaningless to simply compare different losses by choosing one with the lower expected loss, in which the (almost everywhere) zero function must be the best candidate if so. As a result, the best and the most suitable loss function has to be decided by different stack holders with their consent.

1.3 Main Contributions

The research of the Inverted Probability Loss Functions is quite limited to the best of our knowledge. Hence, our major contributions are to find some evidences to make a good foundation for it and to extend its usages in a wider range of applications. Since the research is so limited, many new results are found and presented in the following chapters.

In Chapter 2, we study briefly about the differences between Taguchi loss and Spiring–Yeung framework of losses to give a general picture of the loss functions in quality assurance. A short comparisons on different distributions of process characteristics are provided as well and it is the first similar study in the literature to the best of my knowledge.

In Chapter 3, we have introduced a certain new losses and finally modify the Spiring–Yeung of losses. Moreover, one of the most beautiful result is the discovery of Inverted Student-t loss, which is also a generalisation of Inverted Normal loss and even unknown Inverted Cauchy loss.

In Chapter 4, we study the common loss function in truncated situation to test the limit of the Spiring–Yeung framework how to pretend a bounded loss to be an unbounded loss. In general, the shortcomings of IPLF occur when dealing some functionals cannot guarantee the existence of a density. We also study the whole exponential family for constructing the Inverted Probability loss function so far in the research literature. Definitely, we also find the reasons why some are more interesting and discuss what conditions can make a member of exponential family being applied in Spiring –Yeung framework.

In Chapter 5, we know that the Spiring–Yeung framework will meet its limit when facing bounded support. Therefore, we study the distribution based on quantile function, the 5-parameter Generalised Lambda distribution and find some properties useful for constructing IPLF. Since this distribution has a variable support, it can easily mimic other distributions with a suitable choice of parameters. This is also a generalisation of the study in Spiring and Yeung (1998) that they only use the special case of this distribution to form Inverted Tukey loss.

In Chapter 6, we present some branded-new applications that will use IPLFs as a tool. Since the loss function under IPLF in a truncated situation does not work well, other method of transformation needs to be considered. One of the methods is to (negatively) exponentiate the loss, because exponential function is an absolutely continuous. The exponentiated loss is where the original loss is exponentiated and to be fitted in Spiring–Yeung framework. Some common losses such as quadratic loss, absolute loss, LINEX loss can be exponentiated to make it bounded by using the framework while preserving all properties of the original loss. Since Spiring–Yeung framework cannot cover any unbounded losses, but this approach complements to provide some losses with the similar properties as those original losses by exponentiating. Further, it also shows that the optimal estimator from a particular loss is always inadmissible with respect to another loss.

Finally, in Chapter 7, we try to study a general class of conjugate loss, which wants to explain why the observation of Spiring that the exponentiated loss is a reflected Gaussian density. It also introduces the Bayes risk compared with the frequentist risk. Further, we can have an answer why we choose the Spiring loss is the most secure loss in conjugate sense.

1.4 Organisation of Thesis

This thesis consists of 8 chapters and the following is a summary. Chapter 1 provides an overview, background and motivation of the whole study. Chapter 2 presents a short review of Loss Functions, the main concept of the whole study. Some new losses are introduced to have similar results so far in literature in Chapter 3. Chapter 4 discusses introduces the scope and the questions relevant to the study. This chapter mainly exhibits a lot of concerns in different aspects. Then some preliminary results are demonstrated in Chapter 5. Chapter 6 discusses some applications. Chapter 7 discusses the loss functions from Inverted Probability framework from a more rigorous view. Chapter 8 proposes some limitations, what we have done and concludes the whole thesis.

1.5 Notations

- $\mathbb{R}$ Real number field
- $\mathbb{R}^k$ k-Cartesian product of $\mathbb{R}$
- $\mathcal{B}_X$ Borel algebra of X
- IPLF Inverted Probability Loss Function
- CDF cumulative distribution function
- $^\intercal$ Transpose
- L Loss function
- $\Gamma(\cdot)$ Gamma function
- $\mathbf{B}(\cdot)$ Beta function
- $\mathbb{E}$ Expectation operator
- $\mathbb{P}$ Probability operator
- $Q(\cdot)$ Quantile function
- $\mathbb{1}_X$ Indicator function such that X is true
- $\langle\cdot,\cdot\rangle$ Scalar product

Chapter 2

A Short Survey of Loss Functions in Quality Assurance

This chapter surveys the loss functions in the field of quality assurance and discusses each major breakthrough in each section. Inverted Probability Loss Functions (IPLFs), as the main theme of this study, will be discussed more deeply and some later modifications are also well discussed.

To motivate later results, we will study the traditional quadratic loss functions as a starting point and then follow different arguments by some researchers step-by-step to bring out the ideas of creating the IPLFs.

2.1 Taguchi-type Loss Function

2.1.1 Taguchi Loss

Taguchi (Taguchi, 1986; Taguchi et al., 1989) suggested a quadratic or squared-error Loss Function to motivate and illustrate losses together with the variance of the product quality from a process target. The form of such quadratic Loss Function proposed by Taguchi is

$$L\left(x,T\right) = B\left(x-T\right)^2 \tag{2.1}$$

where $B > 0$ is a constant and T is the target value. Figure 2.1 shows one of the particular example of quadratic loss functions in quality control. Indeed, the quadratic loss function has its advantages:

- it can be seen as a function approximated by the Taylor series expansion about the target and up to the quadratic term.

- it fits the widely-used variance and squared-error loss functions and under a certain conditions of Gauss–Markov theorem, the estimator is minimum variance unbiased.

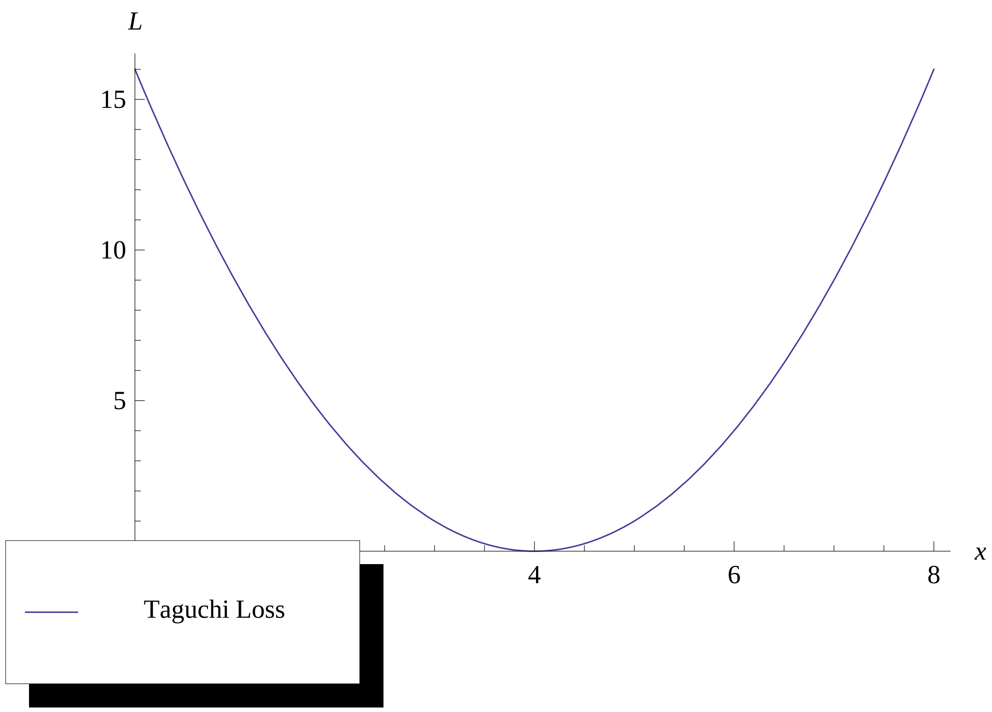


Figure 2.1: A Taguchi quadratic loss with $B = 1$ and $T = 4$

However, there are some defects associated with this quadratic loss function, being criticised by Spiring (Spiring, 1993) and Sun, Laramée and Ramberg (Sun et al., 1996). For instance, the Taguchi loss function increases without bounds; and this finally led more remedies proposed to overcome some difficulties. Another problem is that the quadratic loss function is symmetric around the target, which is not suitable in some situations.

The major solutions are the followings: Ryan loss and Barker loss, while keeping the shape and some properties of Taguchi loss.

2.1.2 Ryan Loss

Ryan (Ryan, 2012) proposed in 1989 a more general form of the quadratic Loss Function as below to overcome the unbounded loss:

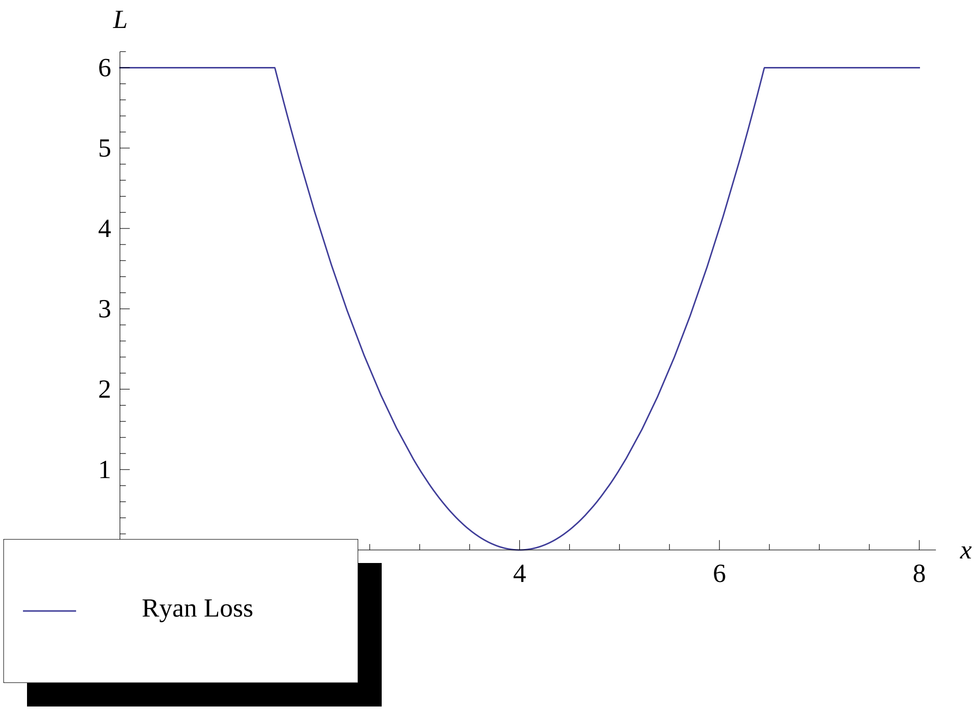


Figure 2.2: A Ryan loss with $(K, B, T) = (6, 1, 4)$

$$L(x,T) = \begin{cases} B(x-T)^2 & |x-T| < \sqrt{K/B} \\ K & |x-T| \geq \sqrt{K/B} \end{cases} \tag{2.2}$$

where $K > 0$ and $B > 0$ may not be equal, but both are constants. The Figure 2.2 shows the modifications such that the Taguchi loss becomes bounded.

In comparison with Taguchi loss, the expected loss under Ryan loss is more controlled and minor because of the boundedness of the loss function.

2.1.3 Barker Loss

It is sometimes impossible to preset a same value of loss realistically and assumes that the amount of loss is symmetric and then Barker (Barker, 1990) also introduced the following quadratic Loss Function:

$$L(x,T) = \begin{cases} B_1(x-T)^2 & x < T \\ B_2(x-T)^2 & x \geq T \end{cases} \tag{2.3}$$

where $B_1 > 0$ and $B_2 > 0$ are both constants and similar to Ryan one, B_1 does not necessarily equal to B_2. Clearly if $B_1 = B_2$, Barker loss becomes Taguchi loss and so Barker loss is a generalisation of Taguchi loss. The following graph, Figure 2.3, shows the Barker loss with asymmetric tolerances or multipliers. From this section onwards, all these three types together will be referred as Taguchi-type loss.

After all, the Barker loss is unbounded like Taguchi loss, the expected value under Barker loss is also similar to that of Taguchi loss.

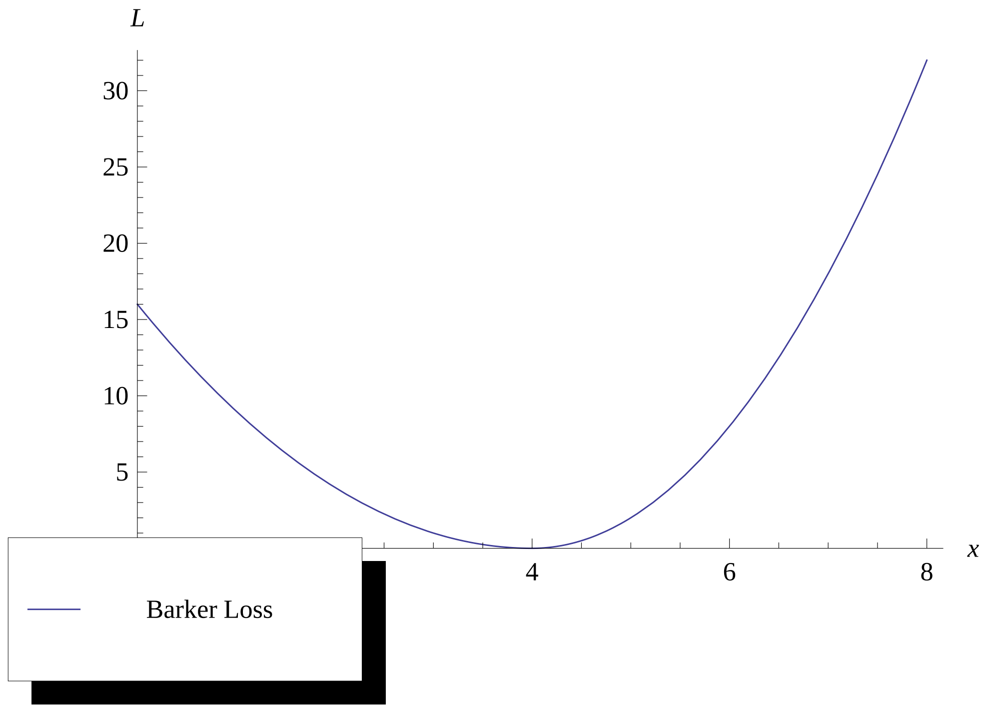


Figure 2.3: A Barker loss with $(B_1, B_2) = (1, 3)$ and $T = 4$

2.2 Spiring Inverted Probability Loss Function

2.2.1 Spiring Loss

Due to the problems aforementioned in Taguchi loss function, some started to explore another route. Spiring (Spiring, 1993) proposed a new concept of loss function by using Gaussian distribution as the general form, as he thought that most experiments or processes follows Gaussian distribution. The general form of this Reflected Normal loss function is

$$L(x, T) = K \left\{ 1 - \exp\left(-\frac{(x-T)^2}{2\gamma^2} \right) \right\} \text{ and } \gamma = \Delta/4 \tag{2.4}$$

where Δ is the Euclidean distance from the target to the point capturing the 99.97% of the maximum loss from a Gaussian distribution.

Spiring Loss Function is conceptually different from the Taguchi loss. It is

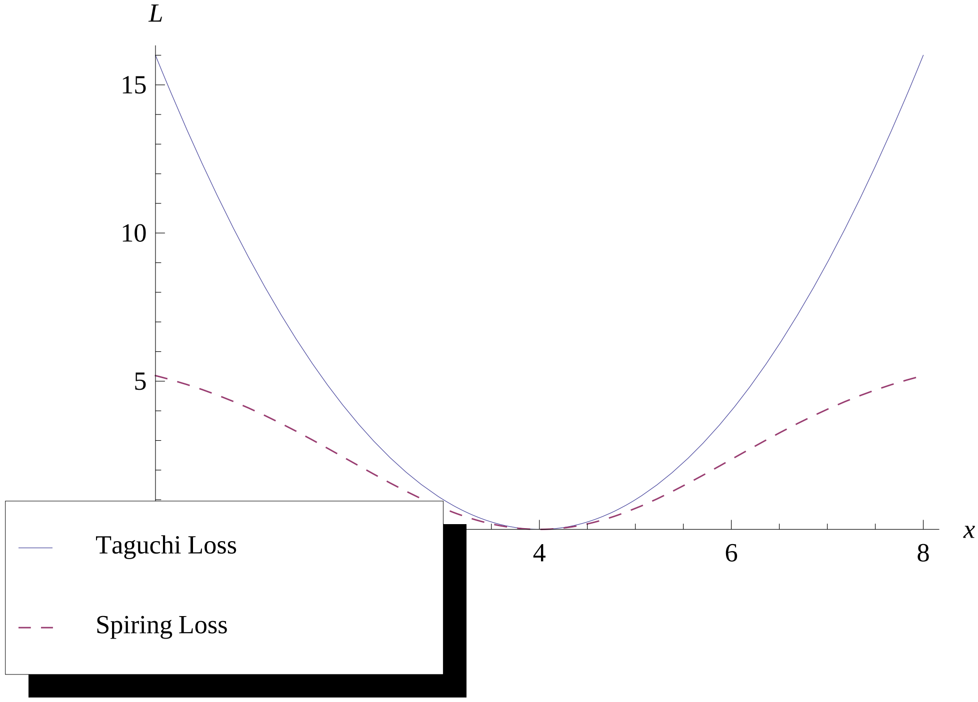


Figure 2.4: A Spiring loss with $(K, \gamma, T) = (6, 2, 4)$ and a Taguchi loss with $(B, T) = (1, 4)$

quite interesting that this enjoys, by default, the properties of boundedness and the freeness of the target value, T.

In Figure 2.4, it is shown that a Taguchi loss with the same target will generate the larger loss when compared with a Spiring loss with the same target and the maximum loss of a Taguchi loss is infinite. Moreover, the rate of approaching the maximum loss in Spiring loss is rather slow and smooth enough. In the meanwhile, compared with Ryan loss of the same target and same maximum loss, Spiring loss produces a smaller loss with the extreme deviations, as illustrated in Figure 2.5. In addition, the rate of approaching the maximum loss is also slower in Spiring loss than in Ryan loss.

Compared with Taguchi loss, the Spiring loss is also smaller and even less than the maximum loss. As Ryan loss is also bounded, the Spiring loss is comparable with Ryan loss, and whatever smaller depends on both parameters.

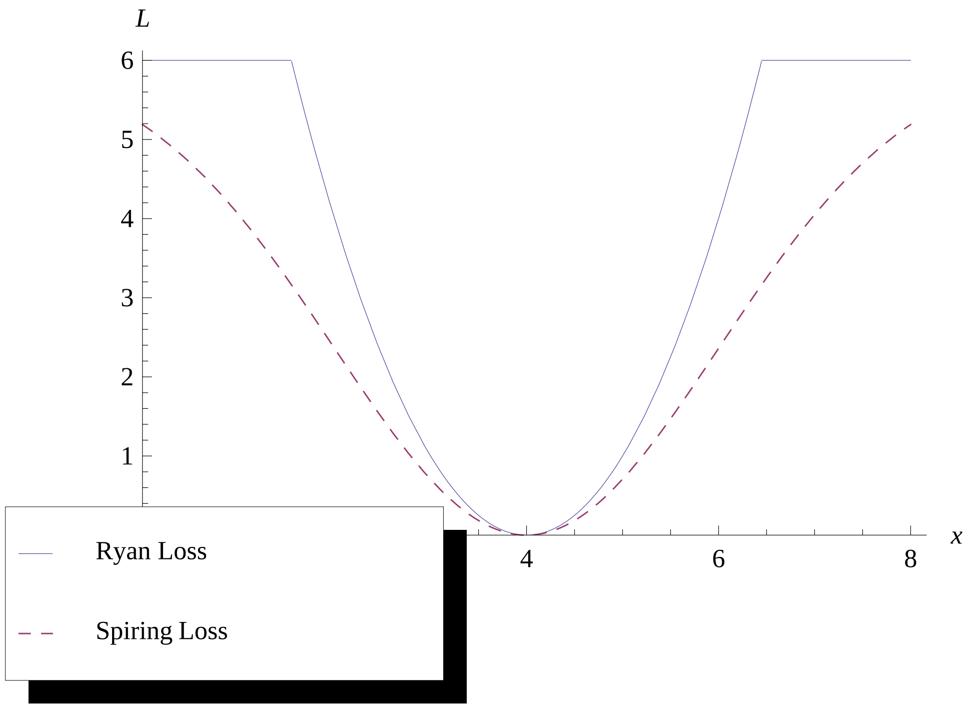


Figure 2.5: A Spiring loss with $(K, \gamma, T) = (6, 2, 4)$ and a Ryan loss with $(K, B, T) = (6, 1, 4)$

2.2.2 Sun *et al* Loss

Sun, Laramée and Ramberg (Sun et al., 1996) modified the Spiring Loss Function and proposed a revised one, Modified Reflected Normal loss function, by freeing $\Delta/\gamma = 4$ to $\Delta/\gamma \in (0, \infty)$. Obviously, Spiring loss function becomes a special case of this type. As Leung and Spiring (Leung and Spiring, 2002) indicated, the Modified Inverted Normal Loss Function "was an important step", because Sun et al. also figured out a method to fit the actual loss via a nonlinear least squares method. The modified form proposed by Sun et al. is

$$L(x, T) = \frac{K_\Delta}{1 - \exp\left\{-\frac{1}{2}(\Delta/\gamma)^2\right\}} \left\{1 - \exp\left(-\frac{(x - T)^2}{2\gamma^2}\right)\right\} \tag{2.5}$$

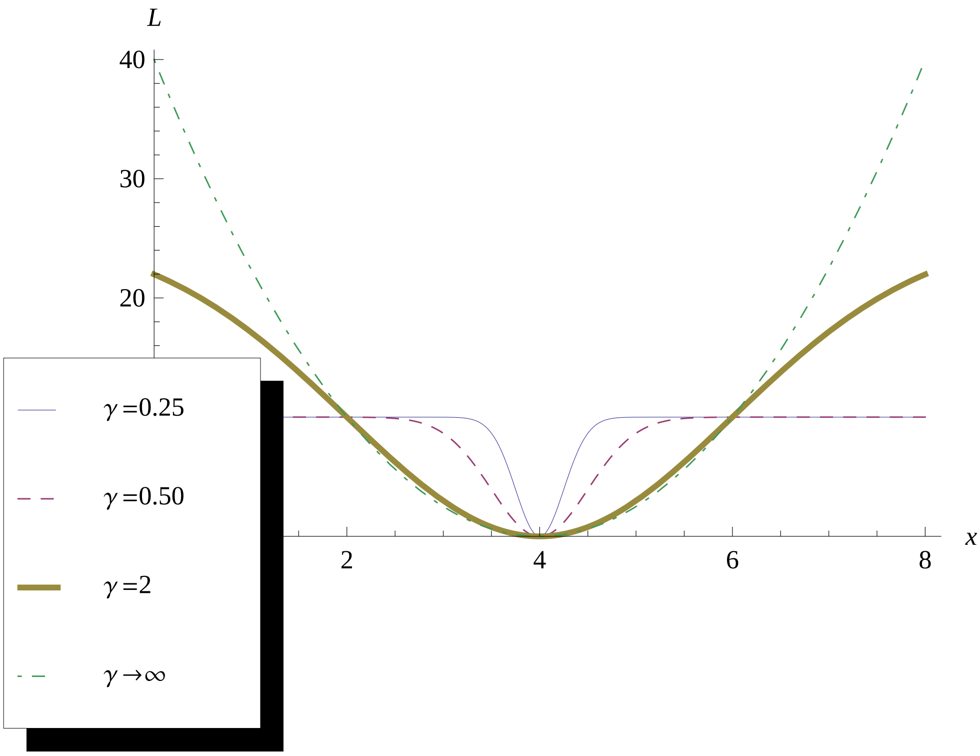


Figure 2.6: Sun *et al* losses of $(K_\Delta, \Delta, T) = (10, 2, 4)$ with different γ's

where $0 < \exp\left\{-\frac{1}{2}(\Delta/\gamma)^2\right\} < 1$ and $K_\Delta > 0$ is fixed. By a simple manipulation, we have

$$\lim_{\gamma\to\infty} L(x, T) = \frac{K_\Delta}{\Delta^2}(x-T)^2 = B(x-T)^2$$

Therefore, it also includes Taguchi quadratic loss as a limiting case.

Obviously, a new parameter Δ is added and K_Δ is the value depending on the ratio between γ and Δ. In the Figure 2.6, we fix the Δ being equal to 2 and the Modified Reflected Normal loss function behaves more like a Taguchi loss as γ tends to infinity. As γ tends to 0.50, the ratio Δ/γ becomes 4 and the loss function is really the Spiring loss. The advantage of this loss functions is that it bridges the bounded Spiring loss and the conventional unbounded Taguchi loss by switching K.

The reason behind the device that the Modified Reflected Normal can work is

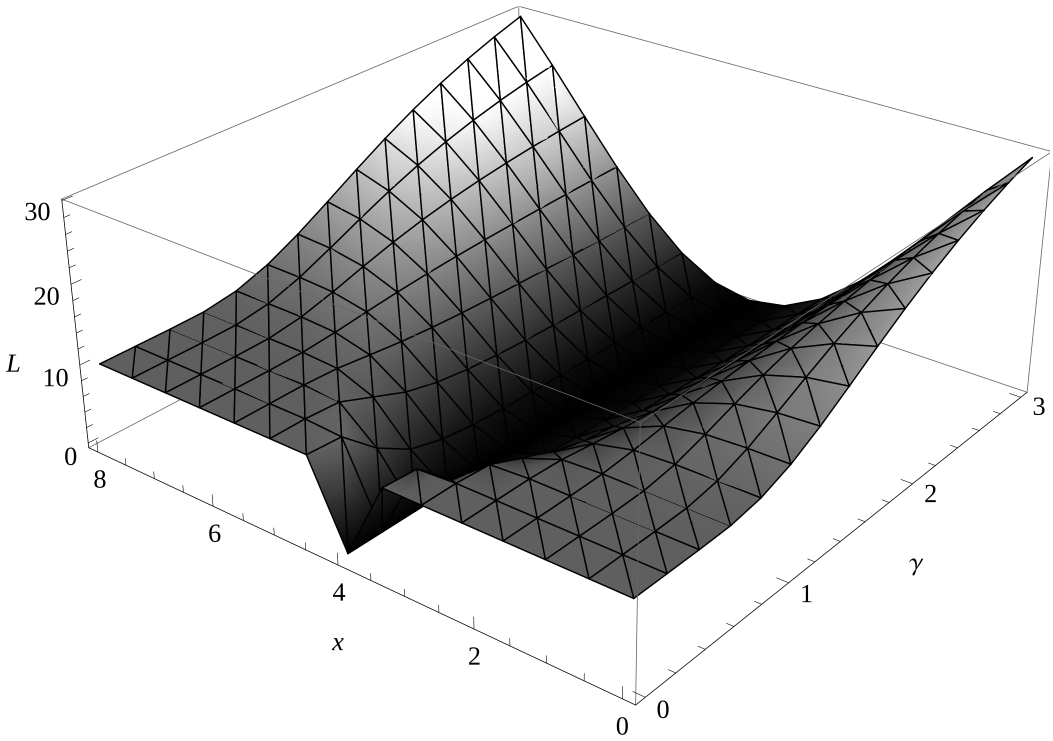


Figure 2.7: The 3D plot of Sun *et al* loss of $(K_\Delta, \Delta, T) = (10, 2, 4)$ against x and γ

the symmetry property and unimodality. The loss functions from either Spiring or Sun et al. also assumed the underlying process as Gaussian distributed. How the device works is also shown in the Figure 2.7.

On the grounds that Sun *et al.* loss is seemed as a bridge between Taguchi loss and Spiring loss, the expected loss depends on the parameter γ if Δ is fixed. Hence the expected loss may be smaller than Spiring loss or as significant as Taguchi loss.

2.2.3 Spiring–Yeung Framework of Loss Functions

Based on the his (Spiring, 1993) idea, Spiring furthered proposing a general class of loss functions (Spiring and Yeung, 1998), referred as Inverted Probability Loss Functions (IPLF), and tried to use the loss functions with *unimodal* distributions other than Gaussian to fit the need for asymmetric loss.

The general class of such Inverted Probability Loss Functions can be easily described as here: Let $f(x,\theta)$ be a probability density function (pdf) with a unique mode at $\hat{x}$. T is the target value where $T = \hat{x}$ should be matched and in general T is a function of other parameters or constants. If we further let

$$m = \sup_{x \in \mathcal{X}} f(x,\theta) = f(T,\theta),$$

then the general form of the Inverted Probability Loss Functions (IPLF) is proposed as

$$L(x,T) = K\left[1 - \frac{f(x,\theta)}{m}\right] \quad \forall x \in \mathcal{X} \tag{2.6}$$

where $\mathcal{X}$ is the support of the distribution $f(x,\theta)$ and $K > 0$ is a constant. Here, a remark is needed: the pdf $f(x,\theta)$ is irrelevant to the distribution of the characteristic or the random variable under examined.

2.2.4 Spiring piecewise INLF

If the Gaussian Distribution is considered to create the Inverted Probability Loss Function, Spiring and Yeung in the same paper (Spiring and Yeung, 1998) even suggested the following form for the situation with asymmetric loss:

$$L(x,T) = \begin{cases} K_1\left[1 - \exp\left(-\frac{(x-T)^2}{2\sigma_1^2}\right)\right] & x < T \\ K_2\left[1 - \exp\left(-\frac{(x-T)^2}{2\sigma_2^2}\right)\right] & x \geq T \end{cases} \tag{2.7}$$

where $K_i > 0$ are constants and σ_i^2 are parameters; but each of them is not necessarily the same. The Figure 2.8 compares the piecewise Taguchi-type loss, Barker loss with the same target and the Spiring piecewise Inverted Normal loss (INLF)

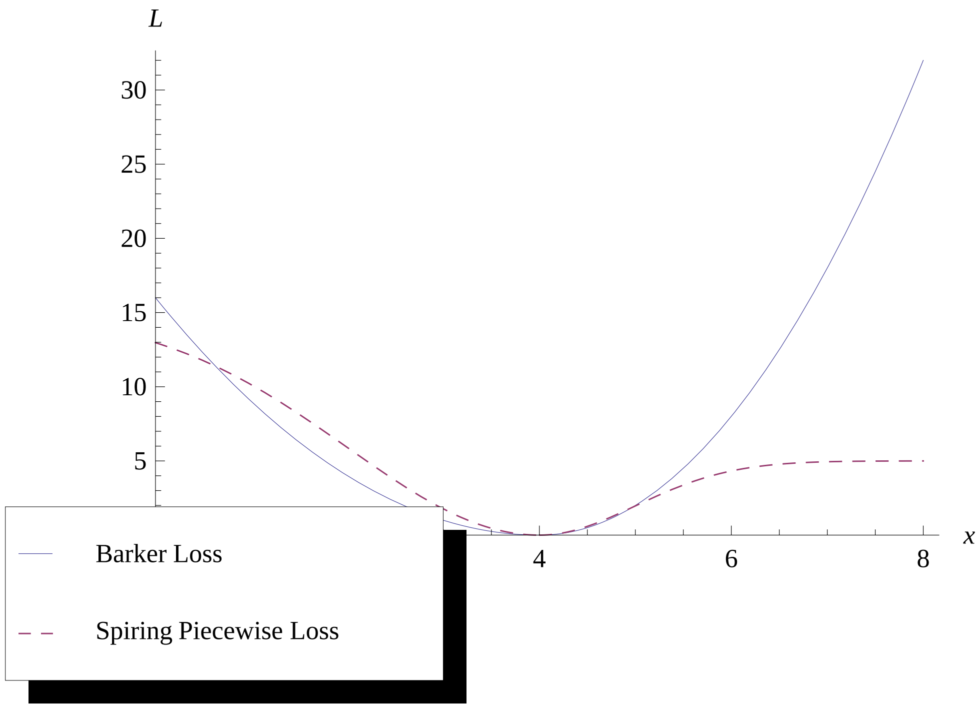


Figure 2.8: A Barker loss with $(B_1, B_2, T) = (1, 2, 4)$ and Spiring piecewise INLF with $(K_1, K_2, \sigma_1^2, \sigma_2^2, T) = (15, 5, 4, 1, 4)$

with the same target. Seeing that Barker loss is always unbounded and convex, Spiring piecewise INLF provides a choice of loss function with bounded maximum loss. As Figure 2.8 shows, the Spiring loss may give a higher loss than the Barker loss due to its quasiconvexity, for which Gaussian distribution is logconcave and so quasiconcave and strongly unimodal (Barndorff-Nielsen, 1978; Dharmadhikari and Joag-Dev, 1988; Bertin et al., 1997).

The Spiring piecewise INLF is the first loss with 2 different maximum losses and this result shows that the expected loss from Spiring piecewise INLF is less than $\max\{K_1,\ K_2\}$, in turn also less than that by any unbounded losses like Taguchi loss or Barker loss.

2.2.5 First Modified Spiring–Yeung IPLF

To generalise the Spiring piecewise INLF, Equation 2.7, Spiring and Yeung (1998) and proposed and Leung and Spiring (2002) reaffirmed one more generalised version of the IPLF, which is

$$L(x,T) = \begin{cases} K_1 \left[1 - \frac{f_1(x,T)}{m_1}\right] & x < T \\ K_2 \left[1 - \frac{f_2(x,T)}{m_2}\right] & x \geq T \end{cases} \tag{2.8}$$

where K_i are two constants and $m_i = \sup_{\mathcal{X}} f_i$. This modification can even provide a convenience in fitting the asymmetric loss, rather than finding a suitable probability density to fit the more restrictive Spiring–Yeung IPLF. According to Sun et al. (1996), K_i may be chosen as a function instead of constant, and therefore we suggest that the novelty of this generalised IPLF is greatly enhanced if K_i have more varieties.

In the literature, f_1 and f_2 are normally the same in form but with different parameters. For instance, Spiring and Yeung (1998) used both Gaussian densities and Leung and Spiring (2002) used both Beta densities. By Equation 2.8, f_1 and f_2 can be two distinct densities and the only requirement is that both densities have the same mode. More clearly, the left side can be a Gaussian, while the right side can be a Beta. For more details about the Beta densities applied in the IPLF, it is formally introduced later.

Figure 2.9 depicts the combination of two different losses from the two developed in the literature, which is unseen before to the best of my knowledge. The left side before the target is a Gaussian density and the right side after the target is a Beta density. It is noted that even the maximum loss in both side is different and

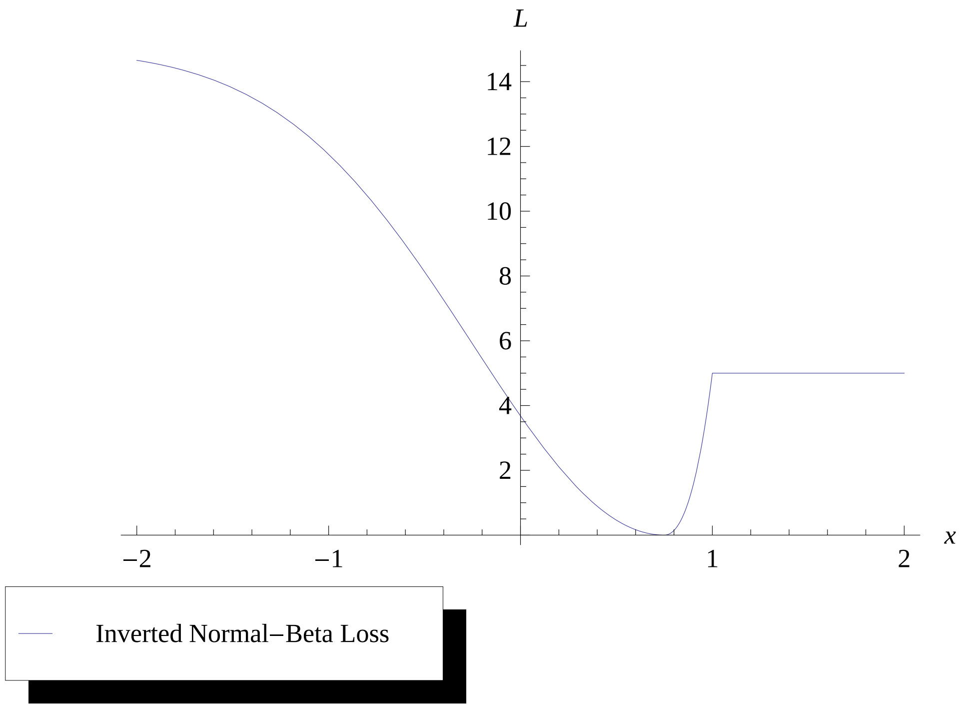


Figure 2.9: A modified IPLF with the left side as INLF with $(K_1, \sigma_2, T) = (15, 1, 0.75)$ and the right side as Inverted Beta loss with $(K_2, \alpha, T) = (5, 4, 0.75)$

hence this is also used for asymmetric loss as Spiring piecewise INLF.

2.2.6 Pan–Wang Loss

Pan and Wang (2000) later studied the Reflected Normal Loss Functions and thought deeply with the results, Equation 2.7 of Spiring and Yeung in 1998 (Spiring and Yeung, 1998). They proposed another more general one with two different modes L' and U' for asymmetric loss with Gaussian distribution:

$$L\left(x,\ \{L', U'\}\right) = \begin{cases} K_1\left[1 - \exp\left(-\frac{(x-L')^2}{2\sigma_1^2}\right)\right] & x < L' \\ 0 & L' \leq x \leq U' \\ K_2\left[1 - \exp\left(-\frac{(x-U')^2}{2\sigma_2^2}\right)\right] & x > U' \end{cases} \tag{2.9}$$

where the conditions are the same as Equation 2.7. L' and U' are the lower and the upper specification limits respectively. An example of Pan–Wang loss is plotted in Figure 2.10 where $L' = 3$ and $U' = 5$ and the maximum loss is 14 and 8 from the left and right respectively.

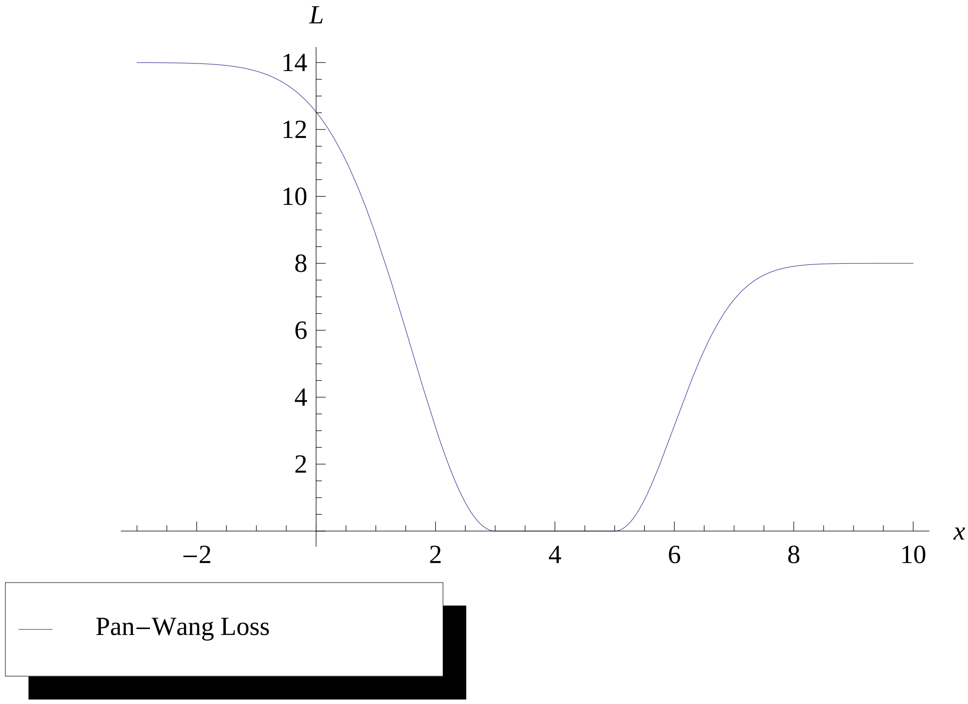


Figure 2.10: A Pan–Wang Revised Inverted Normal loss function with $(K_1,\, K_2,\, \sigma^2_{L_1},\, \sigma^2_{L_2},\, L',\, U') = (14,\, 8,\, 2,\, 1,\, 3,\, 5)$

Pan believed that the quality loss does not fall within the acceptable range of target value. K_i denotes as the maximum loss if the quality departs from the target and the lower and upper limit of the acceptable range respectively for $i = 1, 2$. σ_i^2 are the parameters for the shape of the loss functions. This Pan–Wang loss is modifying Spiring piecewise INLF by revising T into a pair of lower specification limits and upper specification limits (L', U'). Definitely with the Gaussian distribution, Pan–Wang loss is a supplement of the conventional process capability indices, such as C_p and C_{pk}.

Pan (Pan and Li, 2001; Pan and Pan, 2006; Pan, 2007; Pan and Pan, 2009) in his later papers also compared the effectiveness of his Revised Inverted Normal loss functions with Spiring piecewise INLF and Barker loss. Since the criteria are to compare the bounded loss in quality control and to allow the tolerances inside two limits, Pan–Wang loss is certainly preferred.

After all, the expected loss under Pan–Wang loss is less than that under Spiring piecewise INLF, for which a certain interval between L' and U' provides zero loss under Pan–Wang loss but not Spiring piecewise INLF.

2.2.7 IPLF of other distributions

Since Spiring–Yeung framework of loss, referred as Inverted Probability loss functions (IPLFs) is proposed, other distributions can be applied to describe the particular loss in different contexts. The general information about the IPLF is already summarised in the Subsection 2.2.3. We will try to introduce which distributions were once applied in the literature. However, mainly there were only two literature, Spiring and Yeung (1998) and Leung and Spiring (2002), studying the framework other than INLF or piecewise INLF.

For asymmetric loss, Gamma distribution was suggested and hence under the Spiring–Yeung IPLF, the Inverted Gamma loss function was created. The Inverted Gamma loss function has the form:

$$\begin{aligned} L(x,T) &= K\left\{1-\left[\frac{x}{T}\exp\left(1-\frac{x}{T}\right)\right]^{\alpha-1}\right\}\mathbb{1}_{x\in[0,\infty)}+K\mathbb{1}_{x\notin[0,\infty)} \\ &= K\left\{1-\exp\left[(1-\alpha)\left(\frac{x}{T}-\log\frac{x}{T}-1\right)\right]\right\}\mathbb{1}_{x\in[0,\infty)}+K\mathbb{1}_{x\notin[0,\infty)} \end{aligned} \tag{2.10}$$

The following figure, Figure 2.11, shows two different shapes with different α. Therefore, it apparently inherit the boundedness of IPLF and the shape may be accommendated by choosing the most suitable α.

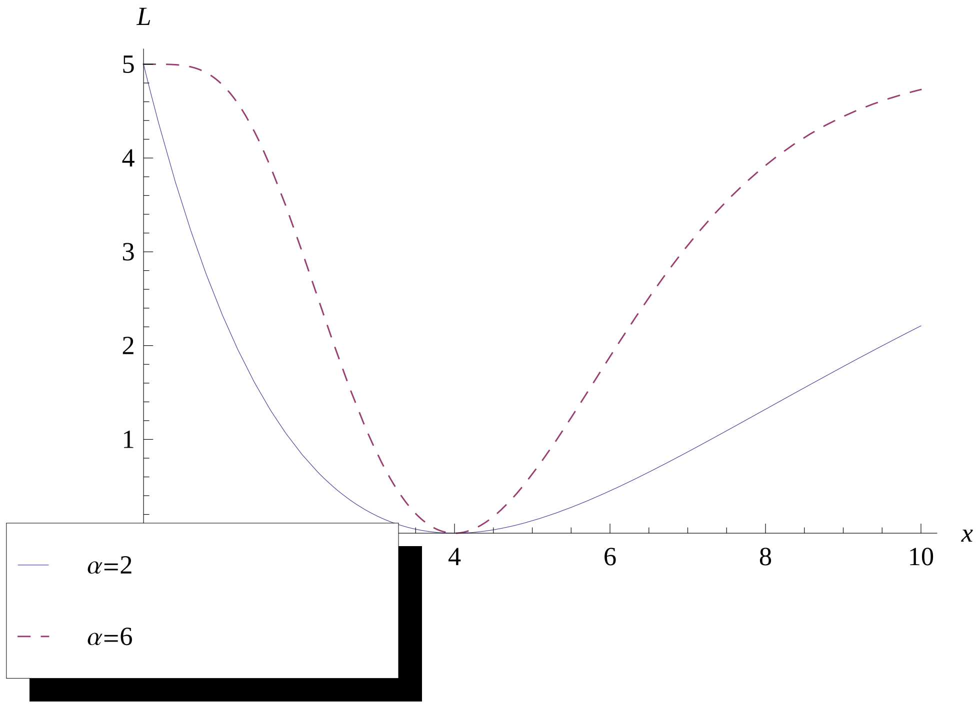


Figure 2.11: Inverted Gamma losses of $(K, T) = (5, 4)$ with different α's

If the α is well chosen, the Inverted Gamma loss can substitute the INLF and the quadratic loss but with the flexibility of the shape in depicting the loss. One of the properties of IPLF is that the preservation of some hierarchies of the underlying distribution. Considering the fact that the chi-squared distribution is a special case of gamma distribution and is unimodal, we can slightly reparametrise the Inverted Gamma loss to have an Inverted Chi-squared loss with degrees of freedom $d = 2\alpha$.

Leung and Spiring (2002) suggested to model both symmetric and asymmetric loss with Beta distribution and so they proposed an Inverted Beta loss function. Analogously, this Inverted Beta loss allows the change of the shapes. Unlike Inverted Gamma loss, it can further provide a loss with faster rate in either side of the deviation from the target within a bounded unit support, $[0, 1]$. The Inverted Beta loss has the following general form:

$$L(x,T) = K\left\{1 - \left[T(1-T)^{\frac{1-T}{T}}\right]^{1-\alpha}\left[x(1-x)^{\frac{1-T}{T}}\right]^{\alpha-1}\right\}\mathbb{1}_{x\in[0,1]} + K\mathbb{1}_{x\notin[0,1]} \tag{2.11}$$

However, this Inverted Beta loss has a serious drawback in flexibility of choosing parameters. Since the corresponding distribution is Beta distribution, Beta distribution has a strict conditions to be unimodal: both parameters have to be greater than 1.

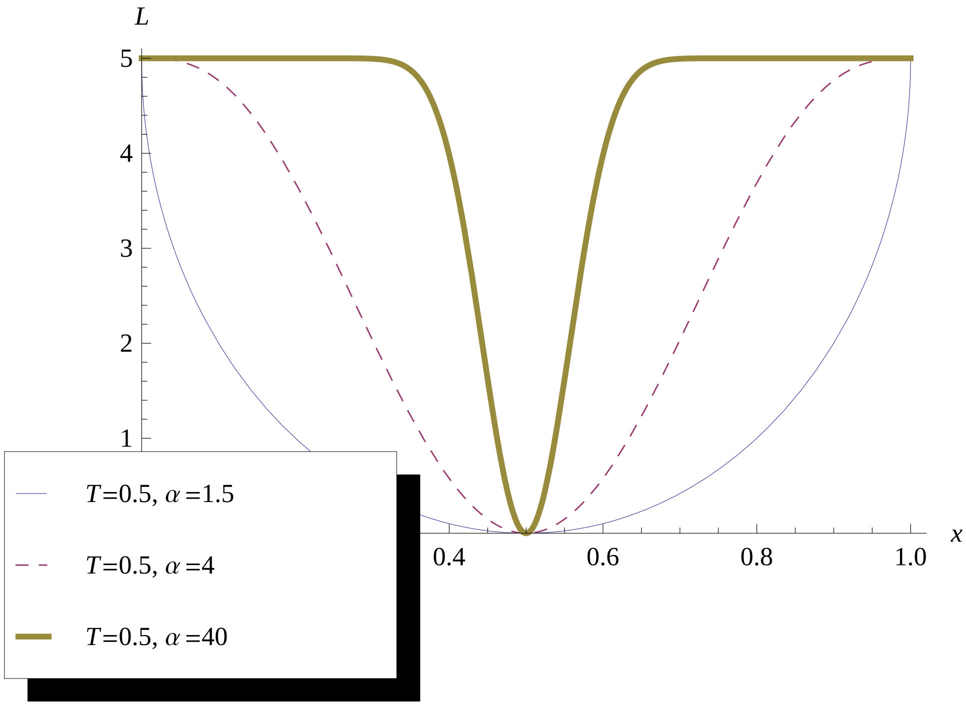


Figure 2.12: Inverted Beta losses of $(K,T) = (5, 0.5)$ with different α's

Using the framework of IPLF and the conditions of unimodality of Beta distribution, suppose (α, β) is the pair of parameters,

$$\begin{aligned} T &= \frac{\alpha - 1}{\alpha + \beta - 2} & \alpha, \beta > 1 \\ \alpha &= \left(\frac{T}{1-T}\right)\beta + \frac{1-2T}{1-T} & \\ \alpha - 1 &= \frac{T}{1-T}(\beta - 1) & 0 < T < 1 \end{aligned} \tag{2.12}$$

That is, if the target T and the parameter α are fixed, there is no freedom in adjusting the shape, unlike the Beta distribution that both parameters are able to reform the shape. As a result, the figure 2 in Leung and Spiring (2002) is not true. The following figure, Figure 2.12 is the corrected one with the parameter value. Figure 2.13 and Figure 2.14 further shows the difference of the shape as the target T is set away from 0.5.

Since the shape of the loss function deviates too much from an inverted form of Gaussian distribution or quadratic form, the expected loss under the Inverted Beta loss functions is also less significant than that under Spiring loss or Taguchi-type loss.

Further, it only allows to work in the unit range and we have to transform the data to unit range beforehand, even though this loss is scale invariant against the generalised Beta distribution. That is, in general we have a data set larger than the unit support and $T > 1$, the maximum and minimum of the data has to be estimated in order to make a transformation

$$X = \frac{Y - \hat{Y}_{\min}}{\hat{Y}_{\max} - \hat{Y}_{\min}}, \qquad Y \text{ is the sampled data} \tag{2.13}$$

Then the problem comes on which an extremum occurs. The dilemma has to be faced: the extremum is a part of the population itself and the loss has to be totally

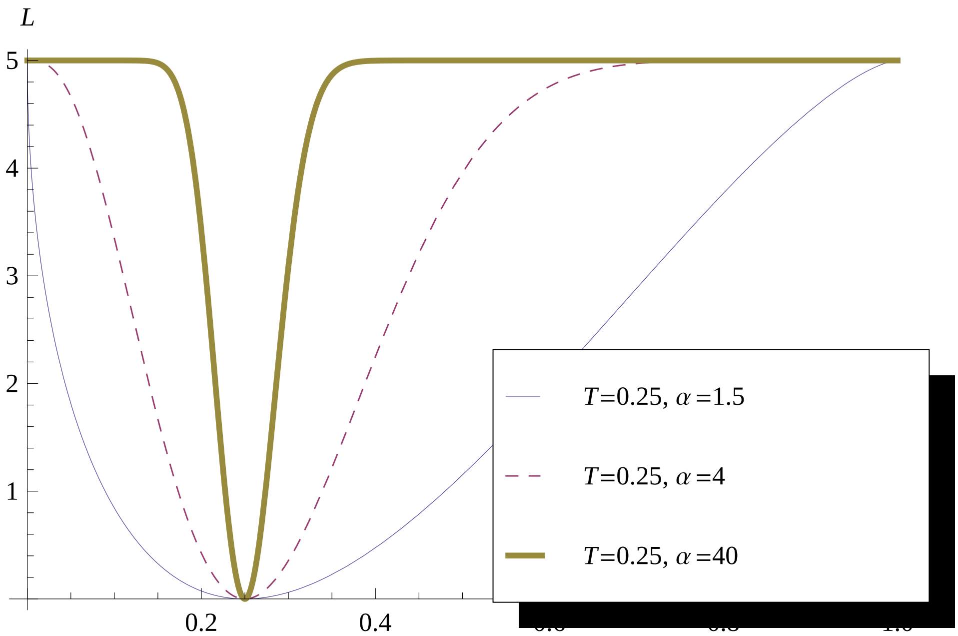


Figure 2.13: Inverted Beta losses of $(K, T) = (5, 0.25)$ with different α's

remade by choosing α or the extremum should be neglected.

Up till now, all the loss is only suitable for one characteristic. If there are more than one characteristics needing to be met, then the loss function has to be multivariate. Correspondingly, both Taguchi-type loss and IPLF are allowed to extend the loss to describe more than one dimension of loss. Since Taguchi-type loss is quadratic and simple, it is easy to vectorise all targets in the meanwhile. One of the particular example of IPLF by a multivariate Gaussian density, which Spiring (1993) also discussed but in a bivariate form only.

By definition, the (nondegenerate) multivariate Gaussian density $f_{\boldsymbol{X}}(\boldsymbol{x})$ with $rank(\boldsymbol{x}) = k$ is

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^k \det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{T}\right)^{\intercal} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{x} - \boldsymbol{T}\right)\right\} \tag{2.14}$$

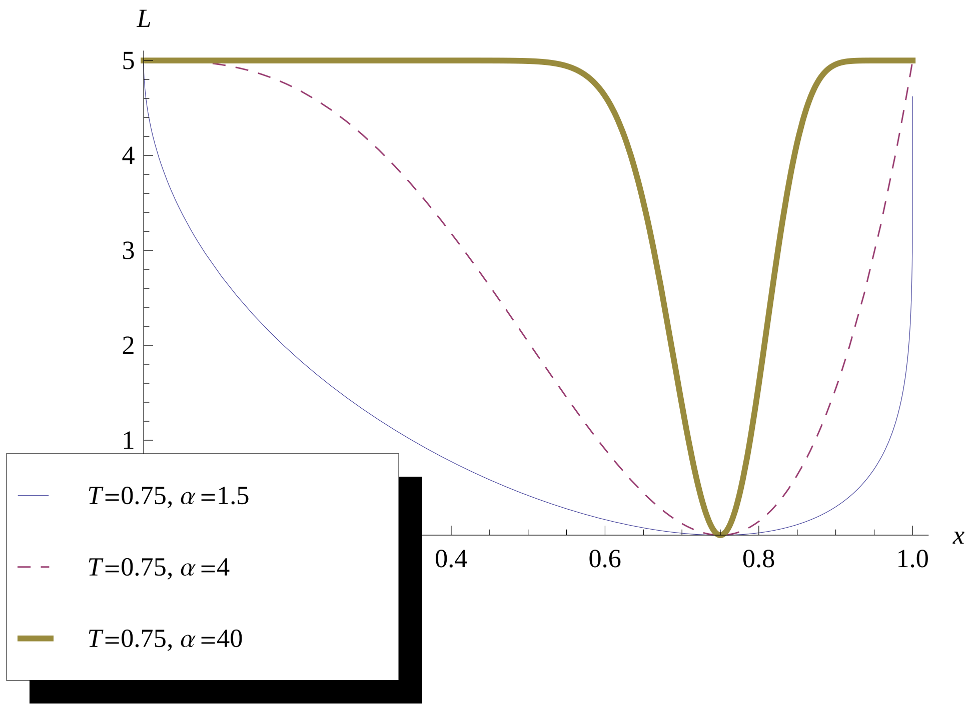


Figure 2.14: Inverted Beta losses of $(K, T) = (5, 0.75)$ with different α's

and so

$$m = \sup_{\boldsymbol{X} \in \mathcal{X}^k} f_{\boldsymbol{X}}(\boldsymbol{x}) = \boldsymbol{T} \tag{2.15}$$

By Equation 2.6, we have the following loss where $\boldsymbol{T} = (T_1, T_2, \cdots, T_k)^\intercal$ is the column vector consisting of target for each characteristic:

$$L(\boldsymbol{x}, \boldsymbol{T}) = K\left(1 - \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{\boldsymbol{T}}\right) = K\left(1 - \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{T})^\intercal \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{T})\right\}\right) \tag{2.16}$$

Hence, two types of losses can work with both univariate and multivariate situations. In this sense, Taguchi-type loss generally requires the independence of each characteristics while IPLF is allowed that any combinations of characteristic is correlated. It results that IPLF has more flexibilities in the multivariate case.

2.3 Fathi and Poonthanomsook Loss Function

Fathi and Poonthanomsook (2007) used the same method as Taguchi and the Loss Function is the function approximated by the Taylor series expansion about the target, but up to the quartic term to overcome all the problems of Taguchi loss. It is therefore called quartic loss function. The general form is

$$L\left(x,T\right)=B_2\left(x-T\right)^2+B_3\left(x-T\right)^3+B_4\left(x-T\right)^4 \tag{2.17}$$

where $B_2,\ B_3,\ B_4$ are all constants. This type of loss functions can capture the asymmetric and symmetric case. Fathi and Poonthanomsook also attempted to apply the analogous thought of Pan (Pan and Wang, 2000; Pan and Li, 2001) and used two modes L' and U' to fit for the asymmetric and symmetric case. If we replace $U'=\mu+k\sigma$ and $L'=\mu-l\sigma$ with μ being the mean, σ being the standard deviation, k and l are some constants,

$$\begin{cases} B_2=\dfrac{L\left(U',\ T\right)\Psi_1^3+\Psi_2^3\left(L\left(L',\ T\right)-B_4\Psi_1^3\left(\Psi_1+\Psi_2\right)\right)}{\Psi_1^2\Psi_2^2\left(\Psi_1+\Psi_2\right)} \\ B_3=\dfrac{L\left(U',\ T\right)\Psi_1^2-\Psi_2^2\left(L\left(L',\ T\right)+B_4\Psi_1^2\left(-\Psi_1^2+\Psi_2^2\right)\right)}{\Psi_1^2\Psi_2^2\left(\Psi_1+\Psi_2\right)} \end{cases} \tag{2.18}$$

where $\Psi_1=T-L'$, $\Psi_2=U'-T$ and $\partial_{xx}L\left(x,T\right)>0$ for all x if $B_4>0$ and $B_3^2<\frac{8}{3}B_2B_4$.

B_4 controls the shape of the quartic loss functions and Taguchi quadratic loss function is a special case of Fathi–Poonthanomsook loss under the conditions of $\Psi_1=\Psi_2$, $L(U',\ T)=L(L',\ T)$ and $B_4=0$. However, it also inherits some defects of Taguchi one, the quartic Loss Function is unbounded and increases in a more serious and faster way than Taguchi loss outside the acceptable range.

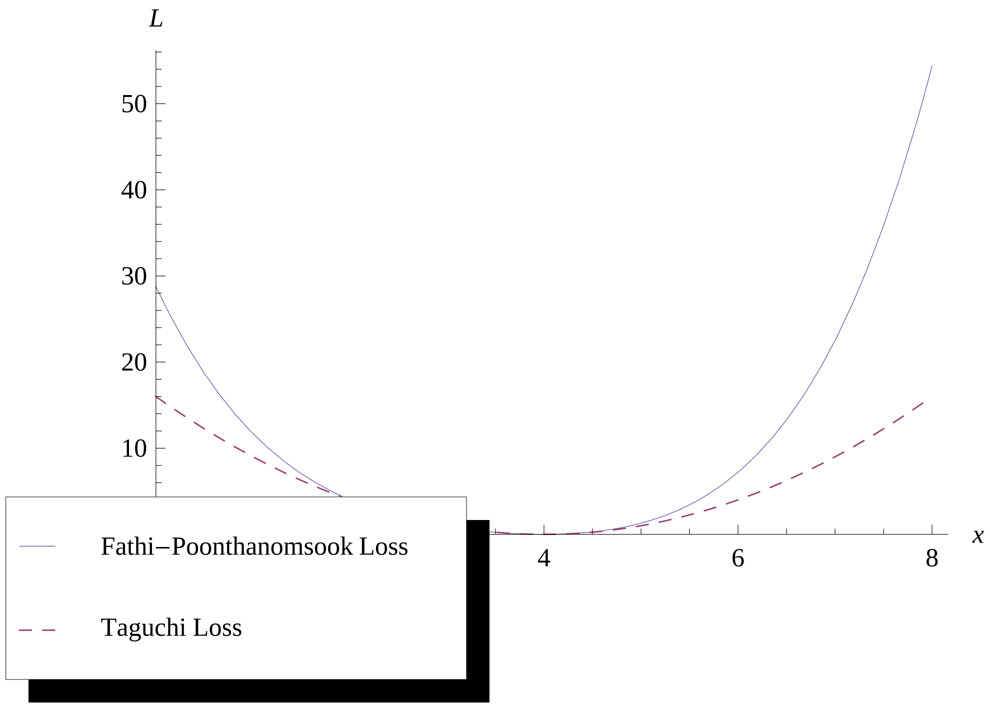


Figure 2.15: A Fathi–Poonthanomsook loss with $(B_2, B_3, B_4, T) = (1, 0.2, 0.3, 4)$ and a Taguchi loss with $(B, T) = (1, 4)$

The comparison between Fathi–Poonthanomsook loss and Taguchi loss with the same target is illustrated in the Figure 2.15. The conditions that $B_4 > 0$ and $-\sqrt{\frac{8}{3}B_2B_4} < B_3 < \sqrt{\frac{8}{3}B_2B_4}$ are both met, but it is observed that even a very small B_4 has led to an enormous loss while deviating from the target.

In contrast with Taguchi loss, the Fathi-Poonthanomsook loss often generates a very heavy expected loss.

2.4 Miscellaneous Concepts

There are some miscellaneous concepts when discovering those IPLF. Spiring and Yeung (1998) introduced the concept of "conjugate distribution" for a particular kind of Inverted Probability Loss Functions. That is, the risk function of the loss functions with respect to the "conjugate" distribution is in a closed form.

Spiring and Yeung (1998) also used some examples to provide an illustration that the "conjugate distribution" for Inverted Normal loss is the Gaussian density, while the "conjugate distribution" for Inverted Gamma loss is the Gamma density.

Leung and Spiring (2002) utilised the Spiring–Yeung framework and made use of Beta density to form the Inverted Beta loss. Since the two parameters of Beta density are also used to describe the shape, this type of loss functions can also fully be adjusted to the needs of symmetric and asymmetric losses. Besides, the Inverted Beta loss does not need to truncate at some points to fit for the asymmetric shape incurred by losses. Both authors tried to show that the "conjugate distribution" for Inverted Beta loss is the Beta density.

Leung and Spiring (2004) gave a summary of some properties of the family of Inverted Probability Loss Functions so far researched in the papers of Spiring (Spiring, 1993; Spiring and Yeung, 1998; Leung and Spiring, 2002). They focused their main points at the Loss Inversion Ratio (LIR), $\frac{f(x,T)}{m}$ and capture some properties of LIR. Actually, LIR properties are just simple properties of pdf and they reconfirmed the results among the earlier IPLF made by Gaussian, Gamma and Beta densities.

Properties of IPLF in Leung and Spiring (2004) when IPLF may be seemed as a random variable:

A. **Boundedness** $0 \leq L(x,T) \leq K;$

B. **Scale-Invariance** $L(x,T) = L(kx,kT), \qquad \forall k.$

Pan (Pan and Pan, 2006; Pan, 2007; Pan and Pan, 2009) related the Loss Function to the Process Capability Indices, and it created the new opportunities for research the application side of the loss functions in quality assurance. Further, Spiring and Leung (2009) also connected the monetary loss, Process Capability Indices and Taguchi-type loss.

In fact, the standard deviation has become synonymous with the dispersion, the physical meaning is not necessarily equivalent in either situations of inter-families of distributions, or intra-family of distributions. Therefore, the actual process spread and the Process Capability Indices may not provide a coherent indication over different distributions. It also leads to a new series of questions about the non-normality circumstances.

Returning to the original Taguchi loss, it is the approximation of Taylor series expansion up to its quadratic term. Some researchers like the anonymous referee of (Leung and Spiring, 2002) prefer the Taguchi form to the parametric form. Leung and Spiring (2002) also cited the conservative view that "no distributional assumptions are necessary" and "...the quadratic approach requires only the determination of a constant and estimates of the process mean and variance". That is why we conjecture that the Taguchi-type loss is just a special case of the general class of the IPLF.

2.5 Numerical Examples

To illustrate the concepts and some nice properties of some losses described in the previous sections, we refer to the data collected in Section 1.2 for the following discussion.

The following table shows the different associated risks or expected losses, which is the average loss to the customers or society when the target is not aimed with different particular chosen loss. We have two different distributions for the same process characteristics, while Normal $(49.40, 4.69^2)$ is our estimated pdf and the other Beta $(2.0994, 2.3184, 40, 60)$ is true. In particular, all the following calculations in Table 2.1 use the same settings: $K = K_1 = 0.3$, $K_2 = 0.2$, $K_\Delta = 0.3(1-e^{-8}) = 0.2999$, $B = B_1 = 0.1$, $B_2 = 0.15$, $B_3 = -0.02$, $B_4 = 0.003$, $\gamma = \gamma_1 = 2$, $\gamma_2 = 1$, $\Delta = 2$, $L' = 50$, $U' = 57.5$, $T = 55$.

	Beta	Normal	% Change
Taguchi Loss	4.8615	5.3327	9.69%
Ryan Loss	0.2661	0.2709	1.83%
Barker Loss	4.8840	5.3861	10.28%
Spiring Loss	0.2287	0.2356	3.01%
Sun *et al* with $\gamma \to 0$ Loss	0.2999	0.2999	1.3×10^{-10}%
Sun *et al* with $\gamma = 2$ Loss	0.2287	0.2356	3.01%
Sun *et al* with $\gamma \to \infty$ Loss	3.6450	3.9982	9.69%
Spiring Piecewise Loss	0.2314	0.2361	2.01%
Pan–Wang Loss	0.1069	0.1093	2.26%
Inverted Gamma-$(\alpha = 5)$ Loss	0.0105	0.0116	9.82%
Inverted Beta-$(\alpha = 3)$ Loss	0.1200	0.1262	5.13%
Fathi–Poonthanomsook Loss	31.176	38.602	23.82%

Table 2.1: (Frequentist) risk associated with different losses

Obviously, the boundedness and the shape of the loss control the robustness of the risk associated with the loss. In general, the Inverted Probability loss function with Normal distribution is more robust than others. However, the most robust one is Ryan loss, which is bounded and does not penalise the off-targets very seriously. As expected, the worst one is Fathi–Poonthanomsook loss for its higher order expansion. That is, if the distribution of the process characteristic has a long tail or it is mostly off-target, the Fathi–Poonthanomsook loss will provide an enormous expected loss.

2.6 Comments

The uses of loss functions are more and more popular in quality assurance while the industry needs a more flexible and realistic loss functions. Although in general conventional quadratic loss or Taguchi loss are adopted, more thoughts on "exotic" losses such as Spiring–Yeung Inverted probability loss framework (IPLF) have gained some attentions from some researchers.

This survey provides a short summary on the loss function being used and examined in the quality assurance. Although it is not exhaustive, most major well-discussed and current research results have been included. Some comparisons have also been conducted to examine different risks associated with losses and different distributions of process characteristics. If the loss is bounded, the expected loss will be more robust and in general the expected loss from Spiring–Yeung IPLF does not deviate too much.

Chapter 3

Prospects and Developments

There are very few literature comparingg Taguchi–type losses with Spiring–type losses. Even if there are some discussions between two, they were judged and considered on a different platform, as if deciding between an apple and an orange. In this chapter, we introduce some new concepts from the ingredients of the materials whatsoever. In particular, we will place more attention on the Inverted Student-t loss.

3.1 Ryan–Barker Loss

To have a fair comparison with the Spiring piecewise INLF in Subsection 2.2.4, we need to find another loss with similar properties. Hence, we introduce a slight modification for Barker loss with the techniques of Ryan loss. We will refer this loss as Ryan–Barker loss which is also suitable for asymmetric loss or symmetric loss. This loss can be seen as a generalisation of all Taguchi-type loss. The form of the Ryan–Barker loss is:

$$L(x,T) = \begin{cases} K_1 & x \leq T - \sqrt{K_1/B_1} \\ B_1(x-T)^2 & T - \sqrt{K_1/B_1} < x \leq T \\ B_2(x-T)^2 & T < x \leq T + \sqrt{K_2/B_2} \\ K_2 & x > T + \sqrt{K_2/B_2} \end{cases} \tag{3.1}$$

where $K_1,\ K_2,\ B_1,\ B_2$ are all constants greater than 0, but not necessarily equal. In the following figure, Figure 3.1, both Spiring piecewise INLF and Ryan–Barker loss are bounded and monotonically increasing while the process is off target, but it does not show any advantages over others.

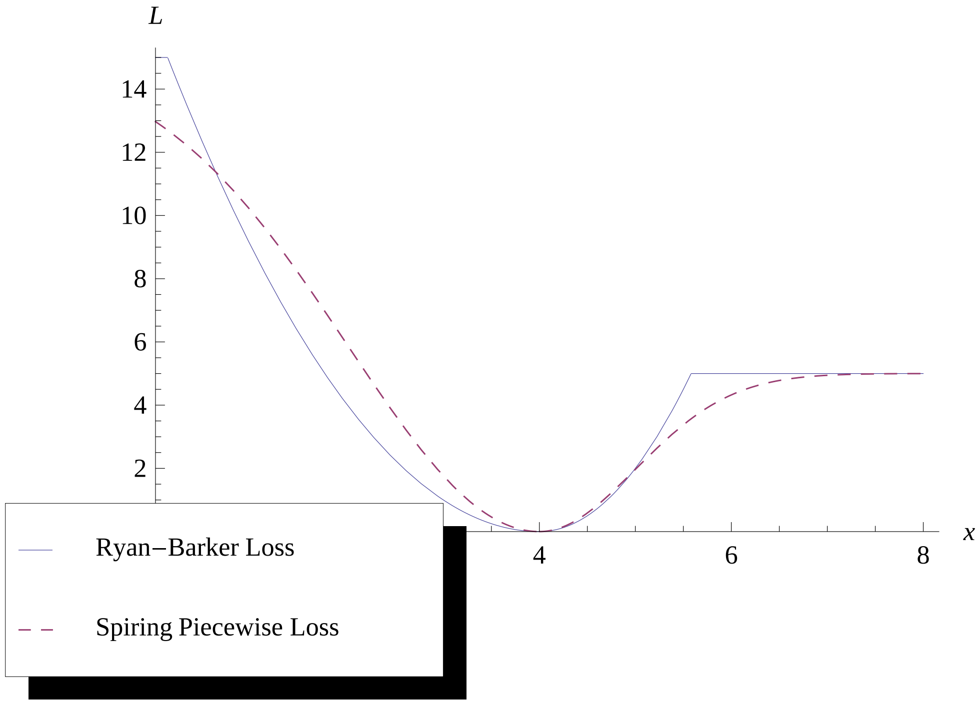


Figure 3.1: A Ryan–Barker loss with $(K_1, K_2, B_1, B_2, T) = (15, 5, 1, 2, 4)$ and Spiring piecewise INLF with $(K_1, K_2, \sigma^2_{L_1}, \sigma^2_{L_2}, T) = (15, 5, 4, 1, 4)$

This loss enjoys both the merits of boundedness from Ryan loss and asymmetry from Barker loss, resulting in a expectation of a value less than the maximum of K_1 and K_2. When comparing with Spiring piecewise INLF, both shape and expected value are also very close and it is possible to replace Spiring piecewise INLF with this loss.

3.2 Second Modified Spiring–Yeung IPLF

Similarly, a more generalised version than Subsection 2.2.5 can be proposed by adding the modifications from Pan–Wang loss in Subsection 2.2.6. the following generalised IPLF can be

$$L\left(x,\ \{a_1, a_2\}\right) = \begin{cases} K_1 \left[1 - \frac{f_1(x, a_1)}{m_1}\right] & x < a_1 \\ 0 & a_1 \leq x \leq a_2 \\ K_2 \left[1 - \frac{f_2(x, a_2)}{m_2}\right] & x > a_2 \end{cases} \tag{3.2}$$

where the conditions are the same as Equation 2.8 and $a_1 \leq a_2$. It is reminded that $f_i \geq 0$ and the loss L is a constant outside the support of f_i. Both f_i can also be two distinct densities and we require that $m_i < \infty$. Without this further condition on m_i, the loss will be kept as maximum loss except at the target. As aforementioned, to have a just and fair comparison between two losses, the major properties of the two have to be matched. The Pan–Wang loss has the properties:

- the loss is bounded
- the loss may be asymmetric
- the loss may be zero for an interval.

Therefore, Pan–Wang loss is now included in this framework, Equation 3.2 where $a_1 = L$ and $a_2 = U$. It is also possible that $a_1 = a_2 = T$ in Equation 2.8. As a result, this second modified IPLF is the most generalised version so far.

3.3 Ryan–Barker–Pan Loss

Therefore, using the analogous methods to produce a Ryan–Barker loss, we also propose another slight modification from Equation 3.1 to form a new loss. This loss is referred as Ryan–Barker–Pan loss to give all the credibility to these three discoverers.

$$L(x, \{a_1, a_2\}) = \begin{cases} K_1 & x \leq a_1 - \sqrt{K_1/B_1} \\ B_1(x - a_1)^2 & a_1 - \sqrt{K_1/B_1} < x < a_1 \\ 0 & a_1 \leq x \leq a_2 \\ B_2(x - a_2)^2 & a_2 < x \leq a_2 + \sqrt{K_2/B_2} \\ K_2 & x > a_2 + \sqrt{K_2/B_2}. \end{cases} \tag{3.3}$$

in which all conditions are the same as Equation 3.1 and $a_1 \leq a_2$. We also transform the parametrisation such that $L' = a_1$ and $U' = a_2$. Evidently, this loss is also a generalisation of Ryan–Barker loss in Equation 3.1. In consequence, we can have a comparison with the same ground.

Clearly, both losses have the same amount of parameters and also satisfy both criteria of boundedness and the allowance of the tolerance limits. Definitely, the IPLF has more flexibilities in changing the rate of approaching the maximum loss. Without this advantage, both are equally acceptable.

When comparing with Pan–Wang loss, both shape and expected value are also very close. In consequence, this loss is also possible to replace Pan–Wang loss.

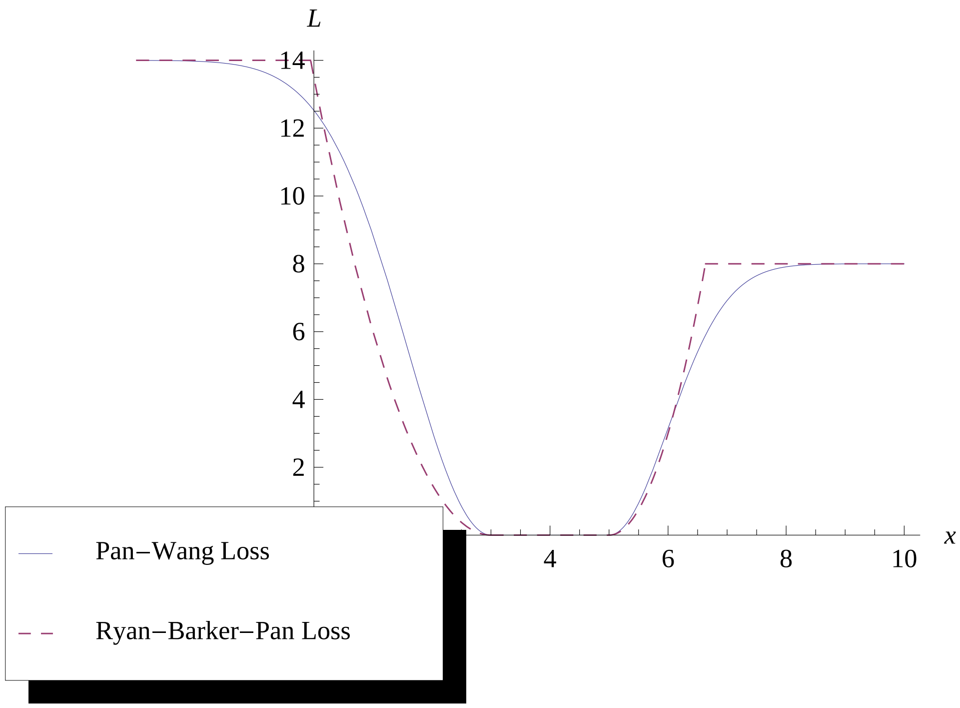


Figure 3.2: A Pan–Wang loss with $(K_1, K_2, \sigma_1^2, \sigma_2^2, L', U') = (14, 8, 2, 1, 3, 5)$ and Ryan–Barker–Pan loss with $(K_1,\, K_2,\, B_1,\, B_2,\, L',\, U') = (14,\, 8,\, 1.5,\, 3,\, 3,\, 5)$

3.4 Inverted Student-t Loss

The loss function approach in quality assurance was initiated by Taguchi (1986) to assess and monitor the losses associated with the process characteristic deviating from the target preset. However, due to the undesirable performance of the Taguchi loss, many practitioners and researchers have tried to propose other loss functions to meet their needs.

Since Spiring (1993) initiated the Spiring loss or Spiring piecewise INLF, Spiring–Yeung framework of loss (IPLF) has been promoted and mostly INLF is adopted. It still does not know whether the modifications from Sun et al. (1996) to relate the Taguchi loss in Section 2.1 only works on INLF. In this section, we try to develop a more general loss than INLF by applying the Spiring–Yeung IPLF. While preserving the nice properties of Spiring–Yeung IPLF, a family of symmetric loss based on an inverted Student-t density is proposed. For asymmetric loss, the most generalised framework, Equation 3.2, can be applied to adjust the choice of a particular loss. Some statistical properties of this Inverted Student-t loss will be discussed and the results are illustrated.

Similar to the INLF (Spiring, 1993; Spiring and Yeung, 1998), the Student-t density has many new properties like Gaussian density. With the Spiring–Yeung IPLF framework, some properties associated with the family of Student-t distributions are needed to be checked for the development of Inverted Student-t loss.

The process target is needed to be the same as the unimodal point of the density to be inverted in Spiring–Yeung IPLF. Student-t density is symmetric in the full real line $\mathbb{R}$ and always have the mode at 0, then the ideal target must be 0 as well. For a more general case with target equal to other values, we have to consider non-central Student-t density or nonstandard Student-t density. It is noted that no matter which Student-t density is chosen, Student-t density is not a member of exponential family. As the mode of a non-central Student-t density is not analytically solvable, we choose the nonstandard Student-t density as the one applied in the Equation 2.6.

Student-t density was proposed by William Sealy Gosset in 1909 under a pseudonym "Student". In the case of a nonstandard Student-t pdf with $\nu > 0$ and $\sigma \neq 0$, the functional form for all $x \in \mathbb{R}$ is

$$f(x \mid T, \sigma, \nu) = \frac{1}{\sqrt{\nu\sigma^2}\mathrm{B}\left(\frac{\nu}{2}, \frac{1}{2}\right)} \left(1 + \frac{1}{\nu}\left(\frac{x-T}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}. \tag{3.4}$$

where $\mathrm{B}(\cdot, \cdot)$ is a Beta function.

The nonstandard Student-t density has a form as $X = T + \sigma Y$, where Y follows Student-t distribution, resulting in the mode as T. Therefore, by setting T equal to

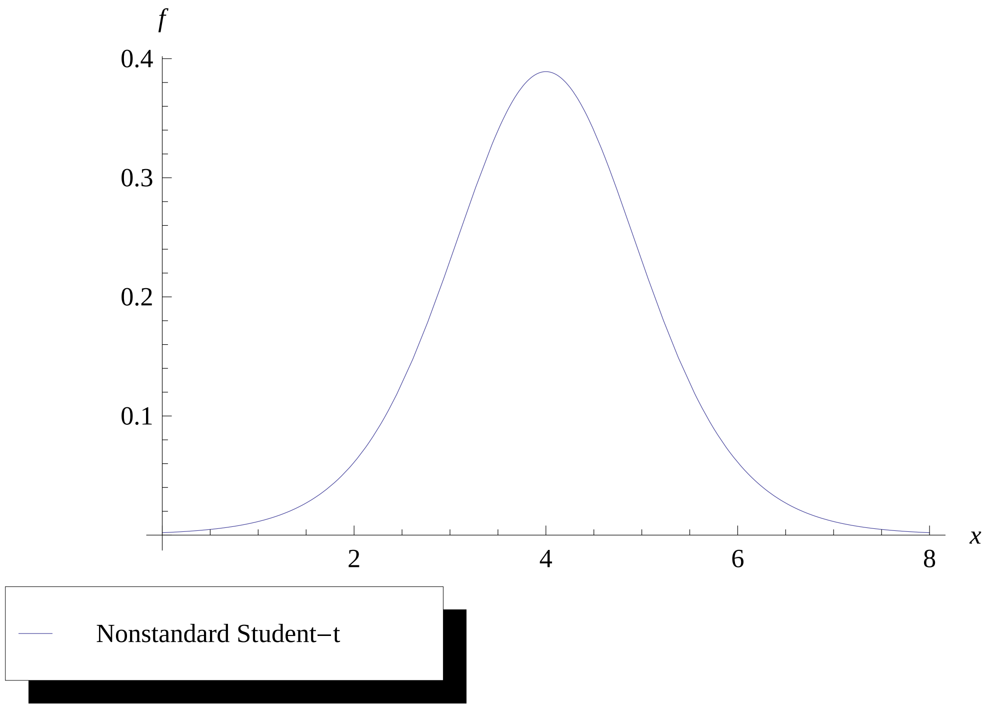


Figure 3.3: A nonstandard Student-t with $(T, \sigma, \nu) = (4, 1, 10)$

the target,

$$\begin{aligned} m =& \sup_x f(x \mid T, \sigma, \nu) = f(T \mid T, \sigma, \nu) \\ =& \frac{1}{\sqrt{\nu\sigma^2}\mathrm{B}\left(\frac{\nu}{2}, \frac{1}{2}\right)} \end{aligned} \tag{3.5}$$

and the resulting Inverted Student-t loss is

$$\begin{aligned} L(x, T) &= K\left\{1 - \frac{f(x \mid T, \sigma, \nu)}{f(T \mid T, \sigma, \nu)}\right\} \\ &= K\left\{1 - \left(1 + \frac{1}{\nu}\left(\frac{x-T}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}\right\}, \qquad x \in \mathbb{R}. \end{aligned} \tag{3.6}$$

Figure 3.3 and Figure 3.4 clearly describe the relationships between the nonstandard Student-t density and the Inverted Student-t loss with $K = 1$. Further, we know that when $\nu \to 1$ the Student-t density will be a nonstandard Cauchy density with a mode and a median at T and scale σ. When $\nu \to \infty$, the Student-t density will be a Gaussian distribution, $N(T, \sigma^2)$.

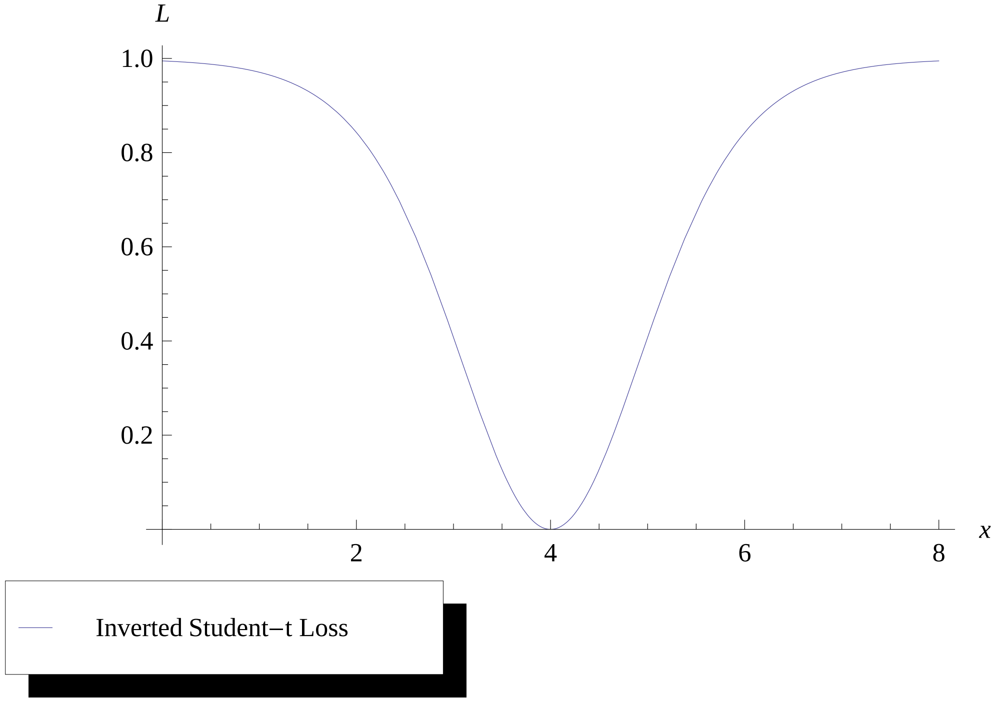


Figure 3.4: An Inverted Student-t loss with $(T, \sigma, \nu) = (4, 1, 10)$

This Inverted Student-t density can be seemed as a generalisation of Spiring INLF. For different degrees of freedom ν, we can have a look at the Figure 3.5. When ν is closer to 1, the rate of approaching the maximum loss of 1 is slower. When the ν is close to ∞, it is closer to be an INLF, which is shown in Figure 2.4.

If we further accept the modification of Sun et al. to relate the Taguchi loss, we need to add the parameter Δ to amend the loss. The new loss function is defined as

$$L(x, T) = \frac{K_\Delta}{1 - \exp\left(-1/2(\Delta/\sigma)^2\right)} \left\{ 1 - \left(1 + \frac{1}{\nu} \left(\frac{x - T}{\sigma} \right)^2 \right)^{-\frac{\nu+1}{2}} \right\}, \qquad x \in \mathbb{R}, \tag{3.7}$$

where K_Δ is not the maximum loss, but the value of the loss at a certain ratio deviating from the target. Figure 3.6 shows the influences of different Δ/σ and illustrates the flexibilities of this new generalised Inverted Student-t loss. The user

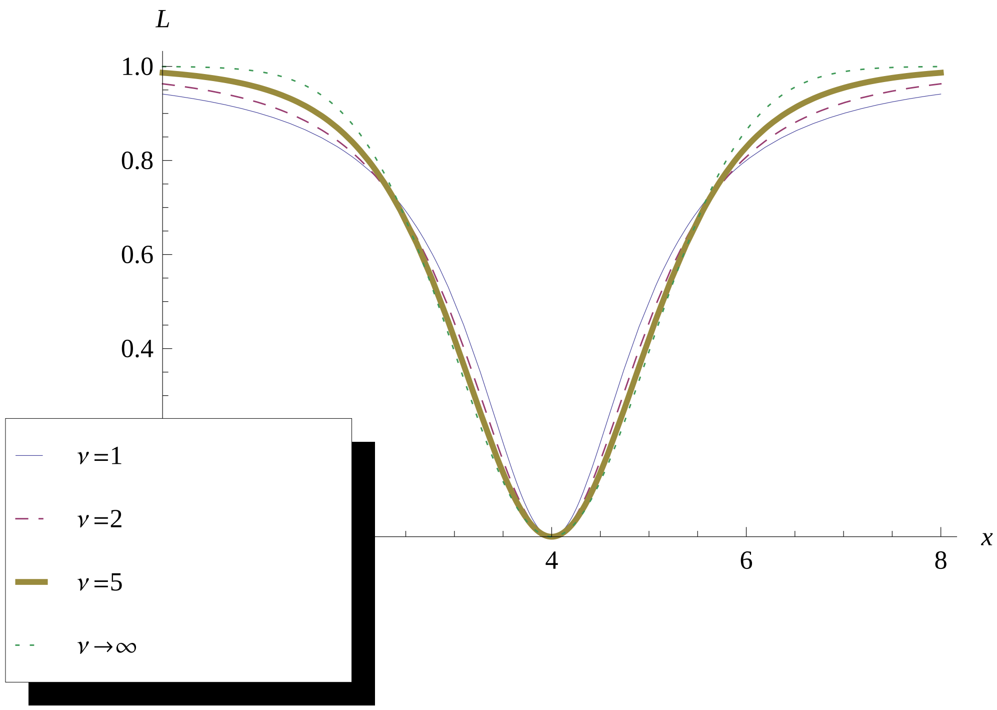


Figure 3.5: Inverted Student-t losses of $(T, \sigma) = (4, 1)$ with different ν's

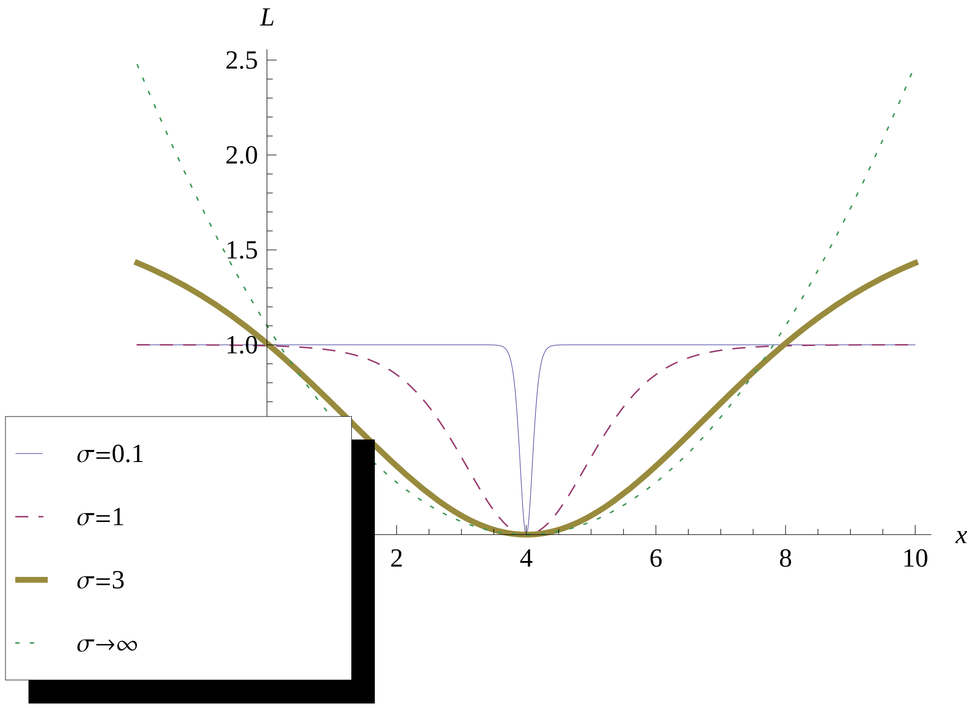


Figure 3.6: Generalised Inverted Student-t losses of $(T, \Delta, \nu) = (4, 2, 10)$ with different σ's

can adjust the ratio according to their own experience and own knowledge of the loss. As σ increases, the loss will close to be a Taguchi loss and ignores the effect of the maximum value. In conclusion, we have some properties of this generalised Inverted Student-t loss:

When σ tends to 0,

$$\begin{aligned}\lim_{\sigma\to 0} L(x,T) &= \lim_{\sigma\to 0}\frac{K_\Delta}{1-\exp\left(-1/2(\Delta/\sigma)^2\right)}\left\{1-\left(1+\frac{1}{\nu}\left(\frac{x-T}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}\right\}\\ &= K_\Delta(1-\mathbb{1}_{x=T}) = K_\Delta\mathbb{1}_{x\neq T}\end{aligned} \tag{3.8}$$

where $\mathbb{1}_{x=T}$ is the indicator function that $x = T$. That is, it converges to a uniform loss equal to the maximum loss except at the discontinuity at T.

When σ tends to ∞,

$$\begin{aligned}\lim_{\sigma\to \infty} L(x,T) &= \lim_{\sigma\to \infty}\frac{K_\Delta}{1-\exp\left(-1/2(\Delta/\sigma)^2\right)}\left\{1-\left(1+\frac{1}{\nu}\left(\frac{x-T}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}\right\}\\ &= \frac{K_\Delta(1+\nu)}{\Delta^2\nu}(x-T)^2\end{aligned} \tag{3.9}$$

Apparently, it is the Taguchi loss in Equation 2.1 with $B = \dfrac{K_\Delta(1+\nu)}{\Delta^2\nu}$.

When ν tends to 0,

$$\begin{aligned}\lim_{\nu\to 0} L(x,T) &= \lim_{\nu\to 0}\frac{K_\Delta}{1-\exp\left(-1/2(\Delta/\sigma)^2\right)}\left\{1-\left(1+\frac{1}{\nu}\left(\frac{x-T}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}\right\}\\ &= \frac{K_\Delta}{1-\exp\left\{-\frac{1}{2}(\Delta/\sigma)^2\right\}}\end{aligned} \tag{3.10}$$

It is another constant loss however the process meets the target.

When ν tends to ∞,

$$\begin{aligned}\lim_{\nu\to\infty} L(x,T) &= \lim_{\nu\to\infty} \frac{K_\Delta}{1-\exp\left(-1/2(\Delta/\sigma)^2\right)} \left\{1-\left(1+\frac{1}{\nu}\left(\frac{x-T}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}\right\} \\ &= \frac{K_\Delta}{1-\exp\left\{-\frac{1}{2}(\Delta/\sigma)^2\right\}}\left(1-\exp\left\{-\frac{1}{2}\left(\frac{x-T}{\sigma}\right)^2\right\}\right)\end{aligned} \tag{3.11}$$

As predicted, the Inverted Student-t loss will be the Spiring INLF as $\nu \to \infty$. Analogously, the generalised Inverted Student-t loss will be Sun *et al.* loss as $\nu \to \infty$ with $\gamma = \sigma$.

Apparently, a question may be raised why the modification of Sun et al. (1996) only works on Spiring INLF and the Inverted Student-t loss, but not in general. If we have a deep investigation on the amendment of K, the term $\exp\left[-1/2(\Delta/\sigma)^2\right]$ is a smooth function. It has the effect when the loss being applied is also smooth enough. In addition, another stop of the modifications is a Taguchi loss, a symmetric loss in the full real plane $\mathbb{R}$. Hence, it is also necessary that all the losses under modification have some common properties. In other words, this modification works like a bridge to link two sides together, but the two sides have to be close enough. Such as the half-plane asymmetric Inverted Gamma loss or the bounded asymmetric Inverted Beta loss cannot have such a method to relate to the Taguchi loss.

The loss function is a tool to depict the loss incurred from the process characteristics when it is not on target. In most of the times, we need to consider the average loss associated with a chosen loss, which is easily evaluated and compared.

The risk function is

$$\mathbb{E}[L(x,\ T)] = \int_{\mathbb{R}} \frac{K_\Delta}{1 - \exp\left(-1/2(\Delta/\sigma)^2\right)} \left\{ 1 - \left(1 + \frac{1}{\nu} \left(\frac{x - T}{\sigma} \right)^2 \right)^{-\frac{\nu+1}{2}} \right\} \mathbf{d}F_X \tag{3.12}$$

but generally not a closed form, even if the process characteristic X has a Student-t distribution. Fortunately, in general, it is still easily computable in numerical.

In practice, if the loss function is chosen, the parameters of the loss are needed to estimate to reflect the decision about the loss from the process. In many cases, only partial information is provided where K_Δ is set to the loss when the associated loss is at $T \pm \Delta$. Similar to the Sun *et al* Loss in Sun et al. (1996) and IBLF in Leung and Spiring (2002), additional "secondary information" on the losses at a set of some additional points are required to solve the whole representation of the loss function to meet the objective to accurately depict the losses with the target in mind. To determine the shape parameters ν and σ, we suggest to use the similar methods by Sun *et al* by using a nonlinear least square search procedure by

$$\min_{\nu>0,\,\sigma>0} \sum_{i=1}^{n} \left[L_i - L(y_i, T)\right]^2 \tag{3.13}$$

where L_i are some additional points of $\{(y_1, L_1), \cdots, (y_n, L_n)\}$.

The shape of the Inverted Student-t loss by definition is already scale invariant under a linear transformation. That is, if the Inverted Student-t loss is based on another nonstandard Student-t distribution, a similar shape with different scales and locations of an Inverted Student-t loss will be obtained with the relationship of the original $y = a + bx$. As a result, the risk is also scale invariant under a linear transformation.

3.5 IPLF as a random variable

In probability and statistics, sometimes we need to calculate the expected value of a function of $g(X)$, but we only know the probability density of X. Correspondingly, many statisticians or practitioners want to find the expected loss or the (frequentist) expected risk of a loss function. Hence, the law of the unconscious statistician can be applied by simply calculating

$$\mathbb{E}[L(X,T)] = \int_{\mathcal{X}} L(x,T)\, \mathbf{d}F_X \tag{3.14}$$

and to be more correctly speaking for a estimated density, the equation is

$$\widehat{\mathbb{E}}[L(X,T)] = \int_{\mathcal{X}} L(x,T)\, \mathbf{d}\widehat{F_X} \tag{3.15}$$

However, as the risk function is used as a tool to examine the statistical procedure, we will abuse the notation to use the former equation to represent the above equation if there is an ambiguity.

The general form of the expected value of the IPLF related to Equation 2.6 is

$$\begin{aligned} \mathbb{E}\left[L\left(x,T\right)\right] &= \int_{\mathcal{X}} K\left\{1 - \frac{f\left(x,T\right)}{m}\right\}\; F_X(\mathbf{d}x) \\ &= K\left\{1 - \frac{1}{m}\int_{\mathcal{X}} f\left(x,T\right)\; F_X\left(\mathbf{d}x\right)\right\} \\ &= K\left\{1 - \frac{1}{m}\,\mathbb{E}\left[f\left(x,T\right)\right]\right\} \end{aligned} \tag{3.16}$$

Since $0 \leq f(x,T) \leq m$, $0 \leq \mathbb{E}[f(x,T)/m] \leq 1$ and so $0 \leq \mathbb{E}[L(x,T)] \leq K$. The central moments of order r of the IPLF with respect to the distribution of the underlying process, utilised to find the variance, skewness and kurtosis of the loss functions, related to Equation 2.6 is

$$
\begin{aligned}
&\mathbb{E}\left[\{L(x,T)-\mathbb{E}\left[L(x,T)\right]\}^{r}\right]\\
&=\int_{\mathcal{X}}\left\{K\left(1-\frac{f(x,T)}{m}\right)-\mathbb{E}\left[L(x,T)\right]\right\}^{r}F_{X}(\mathbf{d}x)\\
&=K^{r}\int_{\mathcal{X}}\left\{\mathbb{E}\left[\frac{f(x,T)}{m}\right]-\frac{f(x,T)}{m}\right\}^{r}F_{X}(\mathbf{d}x)\\
&=K^{r}\int_{\mathcal{X}}\sum_{i=0}^{r}\binom{r}{i}\left\{\mathbb{E}\left[\frac{f(x,T)}{m}\right]\right\}^{r-i}\left\{-\frac{f(x,T)}{m}\right\}^{i}F_{X}(\mathbf{d}x)\\
&=K^{r}\sum_{i=0}^{r}(-1)^{i}\binom{r}{i}\left\{\mathbb{E}\left[\frac{f(x,T)}{m}\right]\right\}^{r-i}\mathbb{E}\left(\left\{\frac{f(x,T)}{m}\right\}^{i}\right)
\end{aligned}
\tag{3.17}
$$

provided that $\mathbb{E}\left[\{f(x,T)\}^{r}\right]$ exists finitely.

Lemma 3.5.1. *Every r^{th} central moment of IPLF is bounded below by $K^{r}(1-2^{r-1})\{r \mod 2\}$ and above by $K^{r}(2^{r-1}-1)$.*

Proof. Suppose Y denote $f(x,T)/m$ and μ_r be the r-th central moment of Y. As shown in Equation 3.17 and $0\le\mathbb{E}\big[Y^{r}\big]\le 1$ for all r,

$$
\begin{aligned}
\mu_r &= K^{r}\sum\nolimits_{i=0}^{r}(-1)^{i}\binom{r}{i}\{\mathbb{E}\left[Y\right]\}^{r-i}\mathbb{E}\left[Y^{i}\right]\\
&\le K^{r}\left(\sum\nolimits_{i=0}^{\lfloor r/2\rfloor}\binom{r}{2i}-1\right)=\left(2^{r-1}-1\right)K^{r}
\end{aligned}
\tag{3.18}
$$

and

$$
\begin{aligned}
\mu_r &= K^{r}\sum\nolimits_{i=0}^{r}(-1)^{i}\binom{r}{i}\{\mathbb{E}\left[Y\right]\}^{r-i}\mathbb{E}\left[\left\{\frac{f(x,T)}{m}\right\}^{i}\right]\\
&\ge K^{r}\left(-\sum\nolimits_{i=0}^{\lfloor r/2\rfloor}\binom{r}{2i}+1\right)\{r \pmod 2\}\\
&= K^{r}\left(1-2^{r-1}\right)\{r \pmod 2\}
\end{aligned}
\tag{3.19}
$$

where $r \pmod 2 = 0$ is even and $=1$ when r is odd. $\square$

From Lemma 3.5.1, some bounds of the variance, skewness and kurtosis of the IPLF can be found.

Corollary 3.5.2. *Variance of the IPLF is bounded and Skewness of the IPLF is bounded below by -3.*

Proof. By Lemma 3.5.1, substituting $r = 2$,

$$0 \leq \mathbb{V}\left[L(X,T)\right] = \mu_2 \leq K^2(2-1) = K^2 \tag{3.20}$$

Substituting $r = 2$ and $r = 3$,

$$Skew[L(X,T)] = \frac{\mu_3}{\mu_2^{3/2}} \geq \frac{K^3(1-2^2)}{K^3} = -3 \tag{3.21}$$

while the upper bound of skewness may be infinite. □

Apparantly, this Corollary 3.5.2 gives the same special result in Leung and Spiring (2004).

However, without the knowledge of probability density of X, it is impossible to use the law of the unconscious statistician. It is unlike the Taguchi loss that the associated risk function is always

$$\mathbb{E}[L(X,T)] = B[\sigma^2 + (\mu - T)^2] \tag{3.22}$$

provided that the distribution of X has a finite second moment. As a result, Spiring and Yeung (1998) introduced the concept of "conjugate loss" that the loss has to be chosen with respect to the density of X.

Under IPLF, the support of the density with respect to loss has to be considered carefully. Outside the support of the associated density in creating the loss for interests, the IPLF attains the maximum loss rather than zero. In other words, if the suppport of the chosen associated density is less than the process density, it may give a great trouble in calculation. For example, in Leung and Spiring (2002), it is claimed that the risk from Inverted Beta loss has a closed form for all distributions, but in fact it only works if the fixed support of the distribution is equivalent to that of the Inverted Beta loss. The following illustration shows the more general case where we retain the parameter α.

$$\begin{aligned}
\mathbb{E}[L(X,T)] &= \int_{\mathcal{X}} K\left\{1 - \frac{f(x,T)}{m}\right\} \mathbf{d}F_X \\
&= K\left\{1 - \frac{1}{m}\int_{\mathcal{X}} f(x,T)\,\mathbf{d}F_X\right\} \\
&= K\left\{1 - C\int_{[0,1]} \left(x(1-x)^{(1-T)/T}\right)^{\alpha-1}\,\mathbf{d}F_X\right\} \\
&= K\left\{1 - C\int_{[0,1]} x^{\alpha-1}(1-x)^b\,\mathbf{d}F_X\right\} \\
&\neq K\left\{1 - C\int_{\mathbb{R}} x^{\alpha-1}(1-x)^b\,\mathbf{d}F_X\right\}
\end{aligned} \tag{3.23}$$

where $b = (\alpha-1)(1-T)/T$.

3.6 Numerical Examples

To illustrate the concepts and some nice properties of some losses described in the previous sections, we refer to the data collected in Section 1.2 for the following discussion.

The following table shows the different associated risks or expected losses, which is the average loss to the customers or society when the target is not aimed with different particular chosen loss. We have two different distributions for the same process characteristics, while Normal $(49.40, 4.69^2)$ is our estimated pdf and the other Beta $(2.0994, 2.3184, 40, 60)$ is true. In particular, all the following calculations in Table 2.1 use the same settings: $K = K_1 = 0.3$, $K_2 = 0.2$, $K_\Delta = 0.3(1-e^{-8}) = 0.2999$, $B = B_1 = 0.1$, $B_2 = 0.15$, $B_3 = -0.02$, $B_4 = 0.003$, $\gamma = \gamma_1 = 2$, $\gamma_2 = 1$, $L' = 50$, $U' = 57.5$, $T = 55$.

	Beta	Normal	% Change
Ryan–Barker Loss	0.2622	0.2657	1.35%
Spiring Piecewise Loss	0.2314	0.2361	2.01%
Ryan–Barker–Pan Loss	0.1386	0.1418	2.25%
Pan–Wang Loss	0.1069	0.1093	2.26%
Spiring Loss	0.2287	0.2356	3.01%
Sun *et al* with $\gamma \to 0$ Loss	0.2999	0.2999	1.3×10^{-10}%
Sun *et al* with $\gamma = 2$ Loss	0.2287	0.2356	3.01%
Sun *et al* with $\gamma \to \infty$ Loss	3.6450	3.9982	9.69%
Inverted Student-t $(\nu \to 0,\ \sigma = 2)$ Loss	0.7622	0.7622	1.6×10^{-8}%
Inverted Student-t $(\nu = 1,\ \sigma = 2)$ Loss	0.5658	0.5801	2.53%
Inverted Student-t $(\nu = 10,\ \sigma = 2)$ Loss	0.5775	0.5943	2.91%
Inverted Student-t $(\nu \to \infty,\ \sigma = 2)$ Loss	0.5810	0.5985	3.01%
Inverted Student-t $(\nu = 10,\ \sigma \to 0)$ Loss	0.2999	0.2999	0.001%
Inverted Student-t $(\nu = 10,\ \sigma \to \infty)$ Loss	4.009	4.398	9.69%

Table 3.1: (Frequentist) risk associated with different new proposed losses

From Table 3.1, it shows that the first two pairs of losses are very close and sometimes the new proposed ones have lower changes when the distribution changes. A slight modification of Taguchi-type losses makes them also enjoy the nice properties of IPLF. Hence, it is still hard to judge which loss is better and is up to the taste of the decision makers. For the final Inverted Student-t loss, its flexibility makes it able to mimic both a conventional Taguchi-type and a Spiring–Yeung IPLF. Further, Inverted Student-t Loss also generalises some cases such as constant loss, but there is a trade-off on the impossibility to have a risk in closed form to reduce the computational demand.

3.7 Conclusion

This chapter develops some new losses with some new flexibilities, such as Ryan–Barker loss, and clarifies that most modifications for either side can also use in another side. One of the most eminent is that the Spiring–Yeung framework is also generalised to include other kind of losses in the literature as its special case. Different directions and some cautions are issued in finding expected losses with IPLF.

Chapter 4

Scope of IPLF

This chapter illustrates that many different loss functions are just a special case of IPLF, though they are well established in their own field. It also enlarges the scope of IPLF extensively from some losses of Taguchi-type to a more general class of losses.

4.1 Ryan loss and bounded Taguchi-type loss

Beforehand it is no harm to understand the definition of generalised IPLF before any further discussions. IPLF was proposed in Spiring and Yeung (1998) as a framework of some losses and bounded losses on fitting the asymmetric and symmetric loss. For simplicity and convenience, we will restate the the definition of generalised IPLF in Equation 3.2.

Definition 4.1.1 (Inverted Probability Loss Function)**.** *Suppose* $f_i(x, \theta_i)$ *be the probability density function (pdf) with a* <u>*unique*</u> *mode at* $\hat{x}_i$ *and* a_i *be the target* *value. Then, let $a_i = \hat{x}_i$ in making a transformation such that*

$$m_i = \sup_{x \in \mathcal{X}_i} f_i\left(x, \theta_i\right) = f_i(a_i, \theta_i) < \infty \qquad \forall i$$

The form of the Inverted Probability Loss Functions (IPLF) is proposed as

$$\forall x \in \mathcal{X}_i,\ L\left(x,\ \{a_1, a_2\}\right) = \begin{cases} K_1\left[1 - \frac{f_1(x,\theta_1)}{m_1}\right] & x < a_1 \\ 0 & a_1 \leq x \leq a_2 \\ K_2\left[1 - \frac{f_2(x,\theta_2)}{m_2}\right] & x > a_2 \end{cases} \tag{4.1}$$

where $\mathcal{X}_i$ is the support of the distribution $f_i(x, \theta_i)$ and $K_i > 0$ may be a constant or a function, $i = 1, 2$.

To show that the bounded Taguchi-type is a special case of IPLF, we first have to find an appropriate distribution having the form of $(x-T)^2$. In the IPLF framework, it is simpler to set $a_1 = a_2 = T$ and $f_1 = f_2$ without loss of generality. We propose a new statistical distribution, Inverted-U quadratic distribution, in short IUQuad(a, b). This distribution has the following probability density function (pdf) with two parameters, a and b:

$$f(x) = \frac{6(x-a)(b-x)}{(b-a)^3} \tag{4.2}$$

where the support is $x \in [a, b]$ and 0 elsewhere.

Notice that neither of a or b are scale or location parameters, but $b-a$ is a scale parameter, and a is the location parameter. The next theorem will establish this fact.

Theorem 4.1.2. *If a random variable X follows IUQuad$(0, b-a)$, then the random* *variable $X + a$ follows IUQuad(a, b).*

Proof. If X is IUQuad$(0, b-a)$, by Equation 4.2, we have

$$f_X(x) = \frac{6x(b-a-x)}{(b-a)^3}, \qquad \forall x \in [0, b-a]. \tag{4.3}$$

By simple manipulation,

$$f_{X+a}(x) = F'_{X+a}(x) = F'_X(x-a) = f_X(x-a), \tag{4.4}$$

so

$$f_{X+a}(x) = \frac{6(x-a)(b-x)}{(b-a)^3} \qquad \forall x \in [a, b]. \tag{4.5}$$

This proves that $X + a$ is a IUQuad(a, b) random variable. □

The parameters a and b can be any real numbers with the condition of $a < b$. With some investigations, this IUQuad(a, b) is found to be a Beta distribution with 4 parameters. Hence, we have the following theorem.

Theorem 4.1.3. *IUQuad(a, b) is a linear transformation of Beta(2,2).*

Proof. Suppose X follows IUQuad(a, b). Considering the density of $Y = (X - a)/(b-a)$, by Theorem 4.1.2, then we obtain

$$\begin{aligned} f_Y(y) &= f_{X/(b-a)}\left(y + \frac{a}{b-a}\right) = f_X(a + (b-a)y) \\ &= 6y(1-y) = \frac{y^{2-1}(1-y)^{2-1}}{\mathrm{B}(2,2)} \qquad y \in [0,1] \end{aligned} \tag{4.6}$$

where $\mathrm{B}(\cdot,\cdot)$ is the Beta function. Owing to the transformation being linear and one-to-one, these two distributions have the same properties. □

As a result, this distribution is a transformed Beta, and so it is absolutely continuous and has a finite support. The mode is also easy to be known. Here are some summary of some basic properties:

Support	$x \in [a, b]$
Mean	$\frac{a+b}{2}$
Median	$\frac{a+b}{2}$
Mode	$\frac{a+b}{2}$
Variance	$\frac{(b-a)^2}{20}$
Skewness	0
Excess Kurtosis	$-\frac{6}{7}$
pdf	$\frac{6(x-a)(b-x)}{(b-a)^3}$
CDF	$\frac{(x-a)^2(3b-a-2x)}{(b-a)^3}$
MGF	$\frac{6\left(e^{at}(-2+at-bt)+e^{bt}(2+at-bt)\right)}{(a-b)^3t^3}$
CF	$\frac{6\left(e^{iat}(-2i-at+bt)+e^{ibt}(2i-at+bt)\right)}{(a-b)^3t^3}$

Table 4.1: Distributional properties of Inverted-U quadratic distribution

It is undeniable that this Inverted-U quadratic is absolutely continuous and unimodal. Now, we can apply the Spiring–Yeung IPLF framework and let $T = \dfrac{a+b}{2}$ be the ideal target and fixed. Then $m = \sup_{x\in[a,b]} f(x) = f(T) = \dfrac{3}{2(b-a)}$. Thus,

$$\begin{aligned} L(x, T) &= K\left\{1 - \frac{f(x,T)}{m}\right\} \\ &= K\left\{\frac{(a+b-2x)^2}{(a-b)^2}\right\} \qquad \forall x \in [a, b] \\ &= K\left\{\frac{2}{b-a}\right\}^2\left(x - \frac{a+b}{2}\right)^2 \\ &= B'(x-T)^2\mathbb{1}_{x\in[a,b]} + K\mathbb{1}_{x\notin[a,b]} \end{aligned} \tag{4.7}$$

where $B' = \frac{4K}{(b-a)^2}$ and K are both constants.

Evidently, this particular loss from IPLF is the same as Ryan loss in Equa-

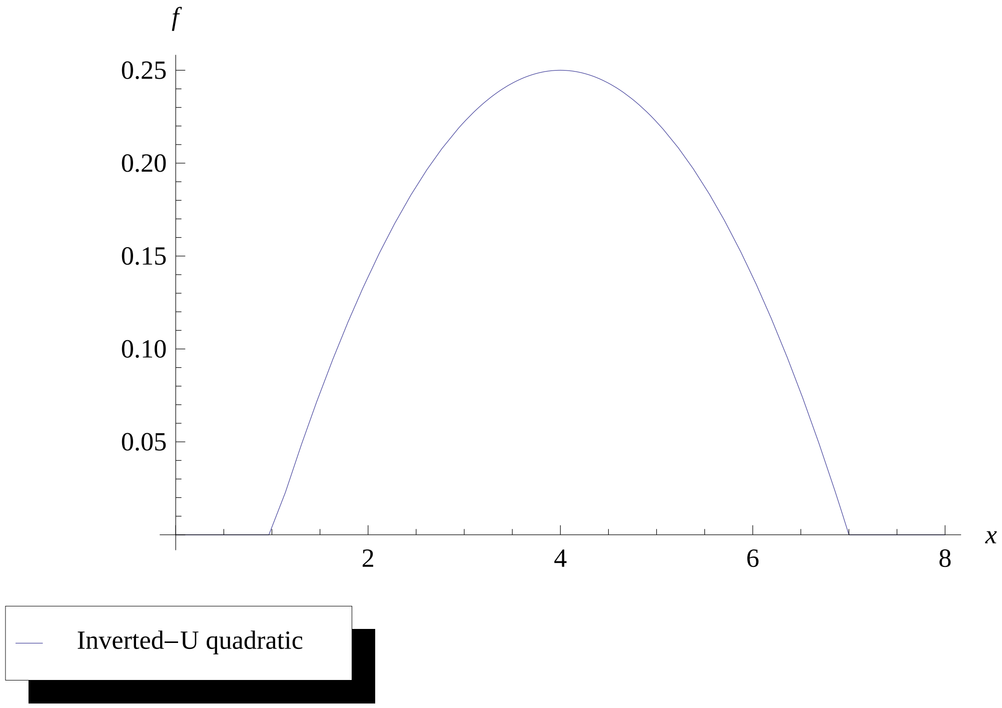


Figure 4.1: Inverted-U Quadratic density of $(a, b) = (1, 7)$

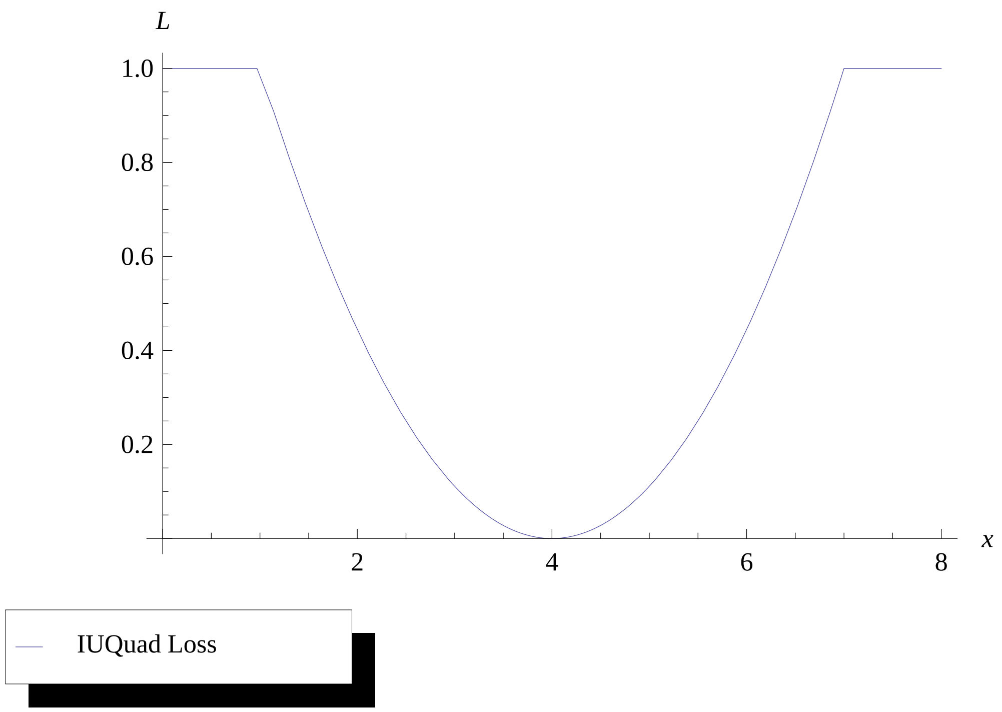


Figure 4.2: IPLF from Inverted-U Quadratic density of $(a, b) = (1, 7)$ with target at 4

tion 2.2 by setting $a = T - \sqrt{K/B'}$ and $b = T + \sqrt{K/B'}$ respectively. This loss is also a particular case of Inverted Beta loss (Leung and Spiring, 2002). Figure 4.1 and Figure 4.2 plot the density and the loss made from IPLF. However, for the general unbounded Taguchi loss, or the conventional quadratic loss, it cannot be fully described in IPLF.

Seeing that IPLF is based on the probability density, the associated probability density having a bounded support will cause the IPLF bounded. Apparently, there is no density for the squared term $(X - T)^2$ maintaining the support of the full real plane $\mathbb{R}$. Additionally, the shortcoming of IPLF is that it is also scale invariant (Leung and Spiring, 2004). It is only possible to modify K to be a piecewise constant like

$$K = \frac{(b-a)^2}{4} K' \mathbb{1}_{x \in [a,b]} + K'' \mathbb{1}_{x \notin [a,b]} \tag{4.8}$$

where K' and K'' are constants and not necessarily the same.

Taking the expected value of the loss for a particular random variable X is one of the way to represent the usefulness of the loss. The expected value of a loss function is also called the frequentist risk. However, the frequentist risk of this loss is seemed rather simple but the result is not user-friendly and not solvable by hand. We let X follow Gaussian distribution with mean μ and variance σ^2 with $\sigma > 0$,

$$
\begin{aligned}
\mathbb{E}[L(x,T)] &= \int_{\mathcal{X}} L(x,T)\mathbf{d}F_X = K\left\{1-\int_{\mathcal{X}}\frac{f(x,T)}{m}\mathbf{d}F_X\right\}\\
&= K\left\{1-\int_{[a,b]}\frac{f(x,T)}{m}\mathbf{d}F_X\right\}\\
&= \frac{K}{\sqrt{\pi}(a-b)^2}e^{-\frac{a^2+b^2+\mu^2}{2\sigma^2}}\left(\sqrt{\pi}e^{\frac{a^2+b^2+\mu^2}{2\sigma^2}}\left\{2\left((a-\mu)(b-\mu)+\sigma^2\right)\right.\right.\\
&\left.\left\{\mathrm{erf}\left(\frac{b-\mu}{\sqrt{2}\sigma}\right)-\mathrm{erf}\left(\frac{a-\mu}{\sqrt{2}\sigma}\right)\right\}+(a-b)^2\right\}\\
&\left.+2\sqrt{2}\sigma\left((a-\mu)e^{\frac{a^2+2b\mu}{2\sigma^2}}+(\mu-b)e^{\frac{2a\mu+b^2}{2\sigma^2}}\right)\right)
\end{aligned}
\tag{4.9}
$$

where $\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\,\mathbf{d}t$. As proved in Theorem 4.1.3 that the IUQuad(a,b) is a transformed Beta$(2,2)$ density.

Referring to the data set in Section 1.2, if $T=55$, $K=0.3$, $a=50$ and $b=60$, the expected loss $\mathbb{E}[L(X,T)]=0.214$, which is more or less similar to the expected loss under Ryan loss.

4.2 Bounded Fathi–Poonthanomsook Loss

Recall from the Section 2.3, the Fathi–Poonthanomsook-type form is

$$
L\left(x,\ T\right)=B_2(x-T)^2+B_3(x-T)^3+B_4(x-T)^4
$$

where $B_2,\ B_3$ and B_4 are all constants. However, in order to make $L(x,T)$ fit for their purpose, Fathi and Poonthanomsook (2007) later needed $\partial_{xx}L>0$ and required $B_4>0$ and $B_3^2<\frac{8}{3}B_2B_4$. In general, all constants are assumed as real and therefore $B_2>0$, $B_3\in\left\{x\in\mathbb{R}\mid x^2<\frac{8}{3}B_2B_4\right\}$ and $B_4>0$.

This loss is a quartic polynomial and hence absolutely continuous and unbounded in the real plane. Here, we try to connect IPLF with this type in a reverse process. Taking the differentiation and substituting $x + T$ by x, then we need to check the determinant of a cubic equation Δ with the conditions such that it has only one real root.

$$\begin{aligned}
\partial_x L &= 2B_2 x + 3B_3 x^2 + 4B_4 x^3 \\
\Delta &= 0 - 0 + 36B_2^2 B_3^2 - 128 B_4 B_2^3 - 0 \\
&= 4B_2^2 \left(9B_3^2 - 32B_2B_4\right) \\
&< 4B_2^2 \left(24B_2B_4 - 32B_2B_4\right) \\
&< 0
\end{aligned} \tag{4.10}$$

Hence, the loss function has only one real root, T, which is the minimum point. In other words, the related Inverted probability distribution is also unimodal at T and absolutely continuous. By the familiar techniques to deal with bounded Taguchi-type in Section 4.1, we can also suggest a suitable but more complex distribution having the functional form of $(x-T)^2 + (x-T)^3 + (x-T)^4$ to get a bounded Fathi–Poonthanomsook loss. For simplicity, we would use A, B and F as the constants and call this distribution as "Inverted U-quartic distribution", in short IUQuartic (a, b, d, A, B, F) . The following Table 4.2 collects some properties of this distribution.

The Inverted-U quartic distribution is similar to the Inverted-U quadratic distribution, not having any possibilities with a support of an infinite range. Indeed, $\int_a^\infty (x-T)^p \,\mathbf{d}x = \infty$ if $p \geq -1$ and $a \leq 0$. Hence, both distributions only exist with a finite support. The main difference between the IUQuad (a, b) and IUQuartic (a, b, d, A, B, F) is that the parameter d is required to ensure the associated density

Support	$x \in [a,b]$
Mean	$\dfrac{-6(a-b)^2B + 20A(a+b)d + 3(a-b)^2(a+b)dF}{40Ad + 6(a-b)^2dF}$
Mode	$\dfrac{a+b}{2}$
Variance	$\frac{28(a-b)^2\left(20A^2d(5d-4)-27B^2(a-b)^2\right)+24Ad(35d-34)F(a-b)^4+9d(7d-8)F^2(a-b)^6}{84(3dF(a-b)^2+20Ad)^2}$
PDF	$\frac{d+1}{d(b-a)} + \left(\frac{240/d}{20A(a-b)^3+3(a-b)^5F}\right)\left[A\left(x-\frac{a+b}{2}\right)^2 + B\left(x-\frac{a+b}{2}\right)^3 + F\left(x-\frac{a+b}{2}\right)^4\right]$ $A>0,\ F>0,\ B^2<\frac{8}{3}AF,\ b>a,\ d\geq 4$

Table 4.2: Distributional properties of Inverted-U quartic distribution

fulfilling the condition of nonnegativity.

Due to the fact that this distribution IUQuartic (a, b, d, A, B, F) under the conditions aforementioned is a polynomial, it is absolutely continuous and unimodal. We can then follow the steps required in Spiring–Yeung IPLF framework.

Let $T = \dfrac{a+b}{2}$ be the ideal target and fixed. Then $m = \sup_{x\in[a,b]} f(x) = f(T) = \dfrac{d+1}{d(b-a)} < \infty$. Thus,

$$\begin{aligned}
&L(x,T)\\
&= K\left\{1-\frac{f(x,T)}{m}\right\}\\
&= K\left\{\frac{240\left[A\left(x-\frac{a+b}{2}\right)^2 + B\left(x-\frac{a+b}{2}\right)^3 + F\left(x-\frac{a+b}{2}\right)^4\right]}{(d+1)[20A(b-a)^2+3(b-a)^4F]}\right\}\\
&= \left[\frac{240AK}{L}\right]\left(x-\frac{a+b}{2}\right)^2 + \left[\frac{240BK}{L}\right]\left(x-\frac{a+b}{2}\right)^3\\
&\quad + \left[\frac{240CK}{L}\right]\left(x-\frac{a+b}{2}\right)^4\\
&= \left[B_2'(x-T)^2 + B_3'(x-T)^3 + B_4'(x-T)^4\right]\mathbb{1}_{x\in[a,b]} + K\mathbb{1}_{x\notin[a,b]}
\end{aligned} \tag{4.11}$$

where $B_2',\ B_3',\ B_4'$ and K are all constants and $L = (d+1)[20A(b-a)^2 + 3(b-a)^4F]$. Other parameters $a,\ b,\ d,\ A,\ B$ and F are predetermined and have to

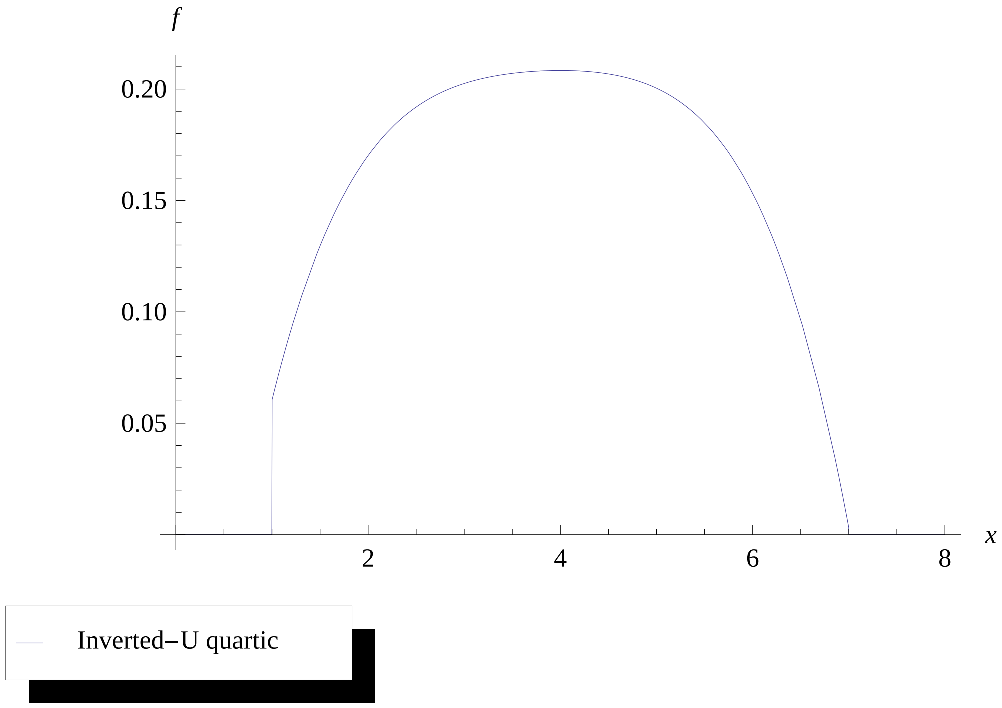


Figure 4.3: Inverted-U Quartic density of $(a, b, d, A, B, F) = (1, 7, 4, 1, 0.2, 0.3)$

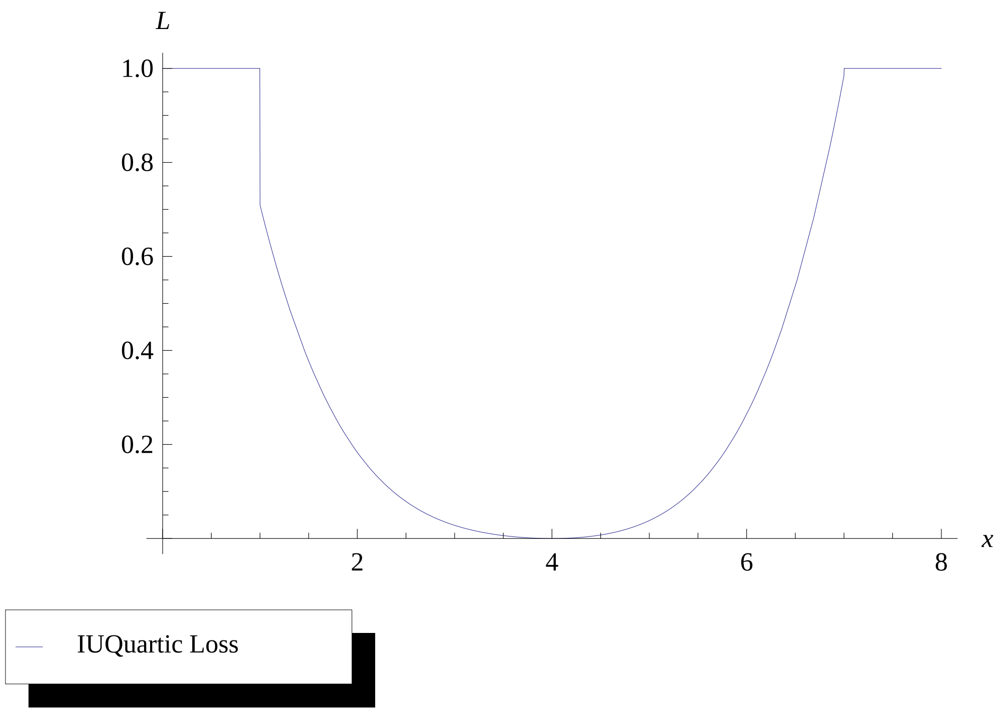


Figure 4.4: IPLF from Inverted-U Quartic density of $(a, b, d, A, B, F) = (1, 7, 4, 1, 0.2, 0.3)$

follow the requirements in Table 4.2. Evidently, $L > 0$ normally in most cases.

Referring to the data set in Section 1.2, if $T = 55$, $K = 0.3$, $a = 50$, $b = 60$, $d = 4$, $A = 1$, $B = -0.2$ and $F = 0.3$, the expected loss $\mathbb{E}[L(X, T)] = 0.199$. This bounded loss is now tamed such that the expected loss is also below the maximum loss.

Since we have a bounded Fathi–Poonthanomsook loss, then we also have the associated density IUQuartic and IUQuartic loss from Spiring–Yeung framework. Figure 4.3 and Figure 4.4 illustrates the associated density and the loss respectively. If all have an alert, then all notice that under IPLF, every loss function made is bounded and fits the requirement of the boundedness what Berger (Berger, 1985) also discussed about an appropriate loss function. The above result is the same as Equation 2.17 when K is set extremely large.

Summing up these two sections, IPLF is a novel framework and covers most cases as long as the loss function has a unique minimum point. Now, we conclude with a theorem:

Theorem 4.2.1. *For any absolutely continuous loss function with a unique minimum, there exists at least a corresponding unimodal probability density such that it fits the Spiring–Yeung framework of loss function.*

Proof. Trivial if the loss function L is bounded. If L is unbounded, then it is possible to approximate $L(x, T) = K\,[1 - f(x)/f(T)] \implies f(x) = f(T)\left[1 - \frac{L}{K}\right]$ with a suitable K. $\square$

4.3 Bounded LINEX loss

Various loss functions has been developed in the literature to suit different needs. Upon different loss, the most popular is still the unbounded symmetric loss, such as squared error loss, that is, Taguchi loss in quality control. However, many authors, such as Berger (1985); Robert (1996, 2001) criticise the usage of the squared error loss, which is nice in mathematical convenience rather than appropriateness of the true loss representation.

In some decision problems, some types of asymmetric losses are proposed. One of the most eminent examples is LINEX, which was proposed by Varian (1975) and populated by Zellner (1986). It was also described in Press (2002).

Suppose X be the variable that needs to meet a target from a decision. The asymmetric LINEX loss is defined by:

$$L(x,T) = b\{e^{a(x-T)} - a(x-T) - 1\}. \tag{4.12}$$

for $b > 0$ and $a \neq 0$. This loss is very flexible in capturing asymmetric loss and the shape changes according to the parameter a, because it controls the weights in exponential side and the linear side. Evidently, both exponential and linear parts are unbounded and so the overall loss is unbounded.

To adjust the LINEX loss being bounded, we try to use the Spiring–Yeung framework to tame the loss. Without loss of generality, we let $b = K$, the multiplier. Suppose we would like to have a maximum loss as K. Then the associated density is

$$\begin{aligned}\int_{\mathcal{X}}\left[f(x)/f(T)\right]\mathbf{d}x = 1/f(T) &= \int_{\mathcal{X}}\left[1 - L(x,T)/K\right]\mathbf{d}x \\ &= \int_c^d \left[2 - e^{a(x-T)} + a(x-T)\right]\mathbf{d}x \\ &= \frac{e^{-aT}}{2a}\left(2e^{ac} - 2e^{ad} - a(c-d)e^{aT}(4 + a(c+d-2T))\right)\end{aligned} \tag{4.13}$$

and

$$f(x) = \frac{1}{G}(2 - e^{a(x-T)} + a(x-T) + |a|\, e) \tag{4.14}$$

where $G = \frac{e^{-aT}\left(2e^{ac}-2e^{ad}-a(c-d)e^{aT}(4+a(c+d-2T))\right)}{2a} + (-c+d)e\,|a|$ and the support is $[c,d]$. Since $f(x)$ requires to be a pdf, the parameter e is needed to make sure that f is nonnegative.

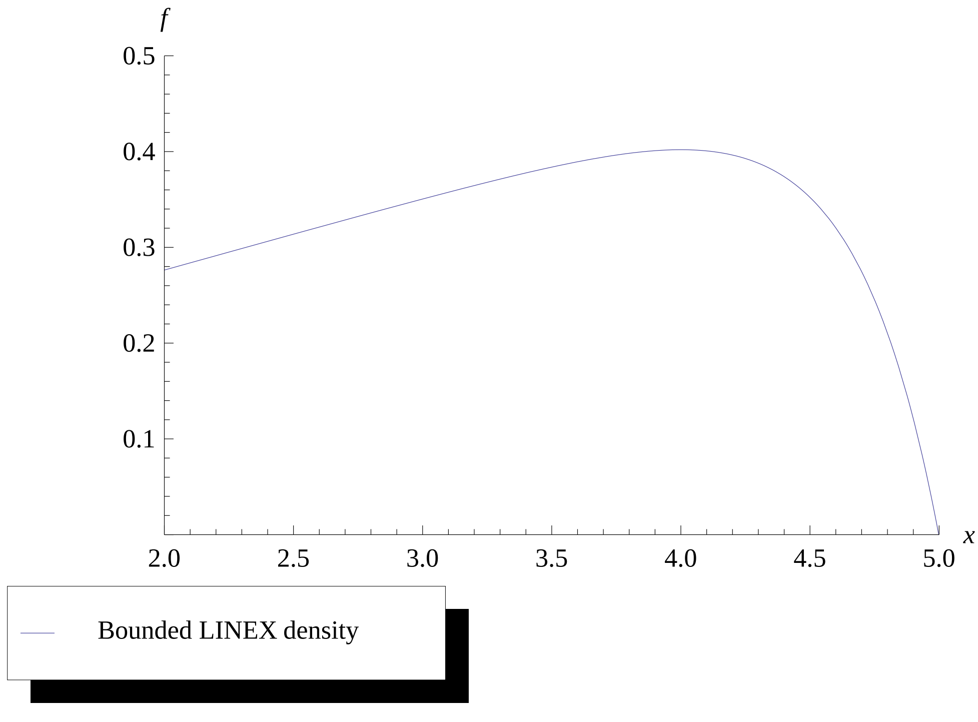


Figure 4.5: Density from bounded LINEX of $(a,c,d,e,T) = (3,2,5,5,4)$

Without loss of generality, we also assume $K = 1$. Then, it is very easy to check that $m = \sup_{\mathcal{X}} f = f(T) < \infty$. Therefore, by the Spiring–Yeung IPLF, we have the inverted loss from the density above.

$$
\begin{aligned}
L(x,T) &= K\left\{1-\frac{f(x)}{f(T)}\right\} \\
&= K\left\{1-\frac{2-e^{a(-T+x)}+a(-T+x)+e\,|a|}{1+e\,|a|}\right\} \\
&= K\left\{\frac{e^{a(x-T)}-a(x-T)-1}{1+e\,|a|}\right\}\mathbb{1}_{x\in[c,d]}+K\mathbb{1}_{x\notin[c,d]}
\end{aligned}
\tag{4.15}
$$

When comparing the IPLF loss from the bounded LINEX density, it is probably worse than the one proposed by Wen and Levy (2001a,b). The BLINEX by Wen and Levy (2001b) has a support of the full plane, but under IPLF, our loss only works on a compact set.

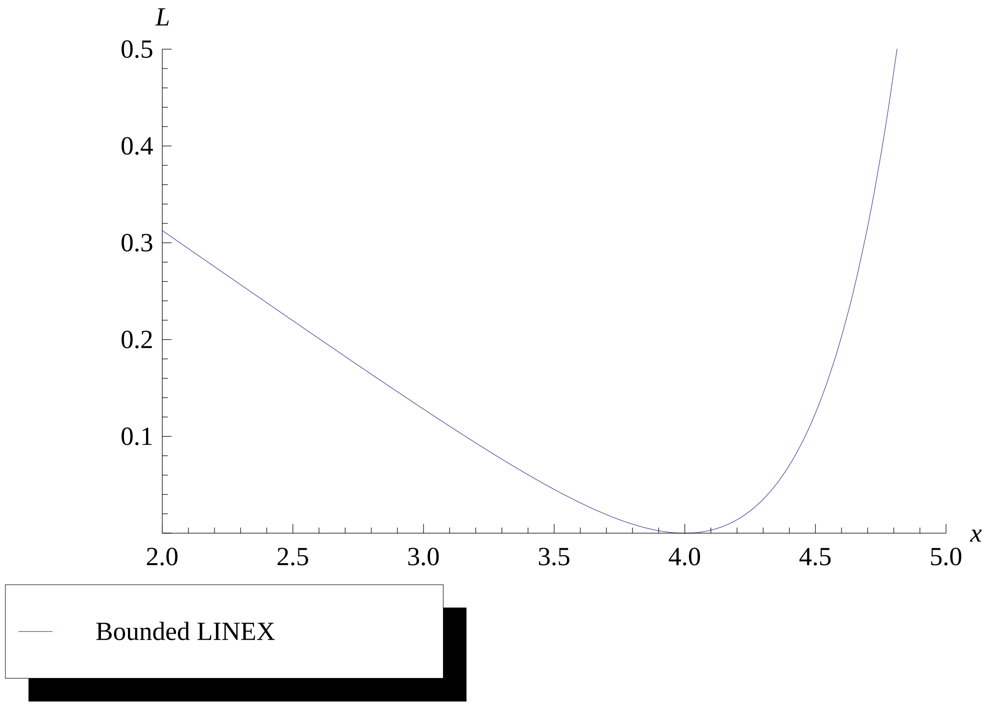


Figure 4.6: Bounded LINEX IPLF of $(a,c,d,e,T)=(3,2,5,5,4)$

The shortcoming of IPLF is due to the fact that it is based on a density, whereas a density has to be greater than zero and defined on a compact set mostly. On the contrary, a random variable may have an infinite support that the losses from

IPLF may not have. In consequence, it adds a lot of difficulties in calculating the frequentist risk and makes higher chances prone to error. The method of Wen and Levy (2001b) is described as below:

Let the initial unbounded loss be $L'(x, T)$,

$$\begin{aligned} L(x,T) &= \frac{L'(x,T)}{1 + L'(x,T)/K} \\ &= K\left(1 - \frac{1}{1 + K^{-1}L'(x,T)}\right) \end{aligned} \tag{4.16}$$

where K can change the shape and the maximum loss simultaneously, like the modification of Sun et al. (1996). This method also has the capability of keeping the support of the initial unbounded loss, which is more superior than IPLF.

Therefore, we should have a rethink whether it is possible to have any new methods to fit the condition of a density to enlarge the scope of IPLF.

Referring to the data set for illustration in Section 1.2, if $T = 55$, $K = 0.3$, $a = 0.5$, $c = 40$, $d = 60$ and $e = 16$, the expected loss for Equation 4.15 is $\mathbb{E}[L(X,T)] = 0.090$. Under some investigations, the expected loss under Taguchi loss is probably larger.

4.4 IPLF from Natural Regular Exponential Family

This section mainly focuses on the regular exponential family. In the previous studies of Spiring (Spiring and Yeung, 1998; Leung and Spiring, 2002, 2004), he and his research team tried to study one by one IPLF from the distribution from regular exponential family: Gaussian, Gamma and finally Beta. However, there are a lot of rooms for further studying other regular exponential family members and an overview of the whole regular exponential family.

In this section, we will study the exponential family first, then deal with the Spiring–Yeung IPLF to discuss which members of the exponential family can be utilised. Finally, we will also try to explain why certain properties for Spiring–Yeung IPLF work.

4.4.1 A Brief Introduction of Exponential Family

A family $\{\mathbb{P}_\theta\}$ of distributions in exponential family is usually written as the densities of the form (Barndorff-Nielsen, 1978; Brown, 1986; Lehmann and Casella, 1998; Liese and Miescke, 2008) with

$$\mathbf{d}\,\mathbb{P}\left(x \mid \theta\right) = \exp\Big\{\eta(\theta)^\intercal l(x) - \psi\left(\eta(\theta)\right)\Big\}\mu(\mathbf{d}x) \tag{4.17}$$

where μ is the appropriate dominating measure. In general, it is either Lebesgue measure or counting measure. θ denotes the parameter in scalar or vector form. ψ, l and η are some appropriate functions, perhaps they are continuous or not. $\eta(\theta)^\intercal$ denotes the transpose of $\eta(\theta)$ if $\eta(\theta)$ is a vector or matrix.

From what we learnt, this form can be further simplified with choosing a suitable dominating measure, re-parameterizing and reducing the information needed by sufficiency. Hence, the most minimal form is

$$\mathbf{d}\,\mathbb{P}\left(x \mid \eta\right) = \exp\Big\{\eta^\intercal l(x) - \psi\left(\eta\right)\Big\} \tag{4.18}$$

The natural parameter space N is an important concept in exponential family, and it can further decompose the whole exponential family

$$N = \left\{\eta \in \mathbb{R}^k \mid 0 < \int \exp\Big\{\eta^\intercal l(x)\Big\}\,\mathbf{d}\nu(x) < \infty\right\}$$

If there exists a $\mathbb{P}_\eta$ for each $\eta \in N$, the whole family $\{\mathbb{P}_\eta\}$ is *full* by definition. With an additional condition of the openness of N in $\mathbb{R}^k$, *id est*, $\forall \eta \in N : \exists \varepsilon > 0 \ni B_\varepsilon(\eta) \subseteq N$, then it is called as the *regular* exponential family. If $l(x) = x$, then it is a *natural* exponential family. $\psi(\cdot)$ is a normalising real-valued function while its importance and uses can be shown below by using the characteristic function $\varphi(\cdot)$ of the general form of the regular exponential family:

$$\begin{aligned} 1 &= \int_{\mathcal{X}} \mathbf{d}\,\mathbb{P}(x \mid \eta) \\ &= \int_{\mathcal{X}} \exp\Big\{\eta^\intercal l(x) - \psi(\eta)\Big\}\,\mathbf{d}\nu \\ \psi(\eta) &= \log\left[\int_{\mathcal{X}} \exp\Big\{\eta^\intercal l(x)\Big\}\,\mathbf{d}\nu\right] \end{aligned} \tag{4.19}$$

$$\begin{aligned} \varphi(u) &= \mathbb{E}[\exp(iu^\intercal X) \mid \eta] \\ &= \int_{\Omega} \exp\Big\{iu^\intercal l(x) + \eta^\intercal l(x) - \psi(\eta)\Big\}\,\mathbf{d}\nu \\ &= \exp\Big\{\psi(\eta + iu) - \psi(\eta)\Big\} \int_{\mathcal{X}} \exp\Big\{(\eta + iu)^\intercal l(x) - \psi(\eta + iu)\Big\}\,\mathbf{d}\nu \\ &= \exp\Big\{\psi(\eta + iu) - \psi(\eta)\Big\} \end{aligned} \tag{4.20}$$

Hence, since the characteristic function $\varphi(\cdot)$ is one of the generating function of all the moments of the random variable X, the differentiability and the smoothness of $\psi(\cdot)$ controls the existence of all moments and the shape of the density of X.

From now on, two main assumptions have to be made. The parameter set is open and nonempty and the regular exponential family has full rank. The results of Barndorff-Nielsen (1978) and Bar-Lev et al. (1992) also indicated that full natural regular exponential family and even full regular exponential family also contains the infinite-divisible elements and self-decomposable elements. Under the conditions in the Theorem 3.2 in Bar-Lev et al. (1992) and Yamazato result (Lukacs, 1983), some distributions, no matter whether they are univariate or multivariate, in the family, such as Gaussian, Gamma, Beta, Hyperbolic, Pareto, are unimodal. Hence, all can be used to produce the relevant loss function with the Spiring–Yeung general class. Let T be the unique mode less than infinity,

$$\begin{aligned} L(x,T) &= K\left\{1 - \frac{\mathbf{d}\,\mathbb{P}(x \mid \eta)}{\sup_{\mathcal{X}} \mathbf{d}\,\mathbb{P}(x \mid \eta)}\right\} \\ &= K\left\{1 - \frac{f(x)}{f(T)}\right\} \\ &= K\left\{1 - \frac{\exp\Big\{\eta^{\intercal} l(x) - \psi\,(\eta)\Big\}}{\exp\Big\{\eta^{\intercal} l(T) - \psi\,(\eta)\Big\}}\right\} \\ &= K\left\{1 - \exp\Big\{\eta^{\intercal}[l(x) - l(T)]\Big\}\right\} \end{aligned} \tag{4.21}$$

However, with the implicit assumption of unimodality, not all members can work under IPLFs. Due to different definition of unimodality in continuous type distribution and discrete type distribution, we have to divide into two cases.

If $\upsilon = -\log \mathbf{d}\,\mathbb{P}$ is quasi-convex, the distribution is unimodal; if υ is convex, the distribution will be strongly unimodal. Since convexity implies quasi-convexity but not the converse, it is apparently that strongly unimodality automatically implies unimodality. Since unimodality is a very weak property that only a limited of fruitful results are obtained. In the latter part, we only focus on the strongly unimodal members, even if unimodality is already enough for setting IPLFs.

One of the major remarkable results for studying strongly unimodality is shown in the next theorem, which is well proven in many literature, such as Barndorff-Nielsen (1978); Brown (1986); Dharmadhikari and Joag-Dev (1988); Bertin et al. (1997).

Theorem 4.4.1. *Any marginal distributions or convolutions of a strongly unimodal* *distributions are again strongly unimodal.*

That is, if the distribution is strongly unimodal, then the marginal is also strongly unimodal. Hence, Dirichlet distribution with all parameters ≥ 1 is strongly unimodal, the marginal distribution of Dirichlet, Beta distribution with all parameters ≥ 1, is also strongly unimodal.

For the strongly unimodal members, they have some common properties:

A. All moments exist. (Bertin et al., 1997)

B. Marginal distribution and convolutions of strongly unimodal of same type are strongly unimodal again. (Barndorff-Nielsen, 1978)

C. $[f(i)]^2 \geqslant f(i-1)\, f(i+1)\,,\ i \in \mathbb{Z}$ for discrete members. (Barndorff-Nielsen, 1978; Bertin et al., 1997)

4.4.2 Spiring–Yeung IPLF Framework with Exponential Family

In consequence, only those members having the strongly unimodal properties is suitable for setting IPLFs without any problems for fitting the requirements of unimodality and boundedness. All members belong to the natural regular exponential family. That is, if the density is from other families, such as curved exponential family, it is not guaranteed that the density can be applied in Spiring–Yeung IPLF framework.

The followings are those continuous members in the exponential family suitable for IPLFs:

- Multivariate Normal
- Gamma with shape parameter ≥ 1
- Laplace with known mean
- Generalized inverse Gaussian with power parameter of 1
- Multivariate Hyperbolic
- Dirichlet with all parameters ≥ 1
- Wishart

Because there is still no process following discrete distribution, the loss function as a discrete function is very limited. This result opens a new door to the uses of discrete pdfs. Until now, the study on Spiring–Yeung IPLF is quite limited, there is also none of the research about the discrete distribution in the loss functions as well. Under the framework of IPLF, it is possible to make a discrete loss with the following strongly unimodal members in the exponential family:

- Poisson
- Negative Binomial with shape parameter ≥ 1
- k-Negative Multinomial with shape parameter $\geq k$
- Multinomial
- Multivariate Hyperbolic
- Multivariate hypergeometric

Another interesting investigation is that the Spiring–Yeung framework does not have any explicit requirements in the normalising real-valued function $\psi(\cdot)$. On the contrary, this normalising real-valued function controls whether all moments exist and therefore the density has a unique mode or not. Some restrictions are implicit that $\psi(\cdot)$ is able to infinitely often differentiable.

4.4.3 Rationale behind Spiring–Yeung IPLF

The Spiring–Yeung IPLF framework was discovered by Spiring and Yeung in an *ad-hoc* way by just plucking some common densities from full regular exponential family. There are certain restrictions and rooms for further study.

In general, the Spiring–Yeung IPLF framework requires to find the mode as the target and change the parameter space to include the mode. It is not trivial to assure that it works most of the time. The form will transform to a different expression of parametrising an exponential family to get the mean equivalent to the mode. For example, to construct the Inverted Beta loss in Leung and Spiring (2002) has to change the Beta distribution with parameters (α, β) to that with parameters (T, α).

I would like to call this re-parametrisation as “modal value parametrisation”. Similar to the mean value parametrisation, we need the following theorem to characterise for regular exponential family, which is proven in Brown (1986); Liese and Miescke (2008).

Theorem 4.4.2. *The mapping*

$$g : \eta \mapsto \nabla\psi(\eta) = \mathbb{E}_\eta \left[l(x) \right] \tag{4.22}$$

is a diffeomorphism of N onto the open set $g(N^\circ)$, where N° is the interior of N.

In regular exponential family, we can always represent the exponential family in the modal value parametrisation, for all $\eta \in N^\circ$, by $\mathbb{P}_{g'(l)}, T \in g'(N^\circ)$ to change the parametrisation such that $\mathbb{E}_{g'(l)} = T$ by using the theorem above. That is, it is always possible to change the parametrisation to be that one of them is T where the number of parameters is unchanged. On the contrary, if the density does not belong to regular exponential family, the re-parametrisation may fail and it may not be used in Spiring–Yeung IPLF framework, even though the density is unimodal.

4.4.4 Re-examining the "Conjugate Distribution"

As the form is explicit now, we can switch our focus to study the concept of "conjugate distribution" in Section 2.4. From Equation 3.16,

$$\mathbb{E}[L(x,T)] \text{ is in closed form } \iff \mathbb{E}[f(x,T)] \text{ is in closed form}$$

Hence, the focus may be changed to look at $\mathbb{E}[f(x,T)]$. Meanwhile, $f(x,T)$ is also a pdf with reparameterization and so have its own support, $\mathcal{X}_p$. Suppose the underlying distribution is a member in the exponential family,

$$\begin{aligned}
\mathbb{E}[f(x,T)] &= \int_{\mathcal{X}} f(x,T)\,\mathbf{d}\,\mathbb{P}(x \mid \eta) \\
&= \int_{\mathcal{X}\cap\mathcal{X}_p} f(x,T)\exp\Big\{\eta^\intercal l(x) - \psi(\eta)\Big\}\,\mathbf{d}\nu \\
&\quad + \int_{\mathcal{X}\cap\mathcal{X}_p^\complement} \exp\Big\{\eta^\intercal l(x) - \psi(\eta)\Big\}\,\mathbf{d}\nu
\end{aligned} \tag{4.23}$$

That is, to check that $\mathbb{E}[f(x,T)]$ is in closed form, two conditions have to be met:

A. $f(x, T)$ has to match the mathematical form of the exponential family such that it can be written as the same density with some new parameters;

B. The second term needs to vanish and thus it is necessary to have $\mathcal{X}_p$ being larger than or equal to $\mathcal{X}$.

In summary, if the expected loss is in closed form and $\{\mathbb{P}_\eta\}$ is full and regular, $f(x, T)$ has to be in the same member of exponential family as well. It is also the reason why an IPLF from a particular density with respect to its density gives the expected loss in closed form.

4.5 Conclusion

This chapter studied whether it is possible to get the associated density with the given loss functions. The shortcoming of IPLF is formed by a density function, but the density function may be only valid in a bounded support. This situation occurs in such as bounded Taguchi-type, bounded Fathi loss and bounded LINEX as some particular examples. IPLF can mimic them when these are bounded, but not in the most available form. In the final part, the common IPLF in literature is formed from the exponential family member. Many IPLFs from the exponential family members are well studied and the conditions whether which members from exponential family are examined. Finally, this chapter also gives some explanations why some will give an expected loss in closed form.

Chapter 5

Generalised Lambda Distribution and IPLF

The Generalised Lambda distribution is a distribution based on the quantiles instead of the realisation of a random variable and hence the support is in a variable range from bounded range to infinite range. In general, it is also very powerful and extensive such that it can approximate any univariate probability density. This chapter discusses the Generalised Lambda distribution and follows the practice of Spiring and Yeung (1998) in a more restrictive Tukey lambda distribution in IPLF.

5.1 Introduction

In Spiring and Yeung (1998), Tukey lambda distribution was also discussed, but this distribution is completely contrast to the ones frequently used, which are some members of exponential family. Tukey lambda distribution is based on quantile function, which is the inverse of the common probability function. The Tukey lambda distribution was first initiated in Hastings et al. (1947) and studied in depth by Tukey (1960). After various research, the Tukey lambda distribution was generalised from a 3-parameters distribution to a Generalised Lambda distribution of 4 parameters by Ramberg and Schmeiser (1972, 1974) and Ramberg et al. (1979). This Generalised Lambda distrbution was also studied extensively in the excellent monographs by Karian and Dudewicz (2000, 2011) and we will use them as the main references. In the meanwhile, a different parametrisation of 4-parameter GLD was suggested in Freimer et al. (1988). With a few discussions in Gilchrist (2002), a 5-parameter GLD was proposed.

The Generalised Lambda distribution, in short GLD, is absolutely continuous distribution. Since it is based on quantiles rather than the realisation of a random variable, we simply have a review on those basic concepts on quantiles.

For any random variable X, the cumulative distribution function (cdf) of X is

$$F_X(x) = \mathbb{P}[X \leq x] = \int_{(-\infty,x]} \mathbf{d}F_X \tag{5.1}$$

Obviously, if we further know that the distribution is absolutely continuous with respect to another measure, say μ, then

$$F_X(x) = \int_{(-\infty,x]} \frac{\mathbf{d}F_X}{\mathbf{d}\mu}\,\mu(\mathbf{d}x) = \int_{(-\infty,x]} f(x)\,\mu(\mathbf{d}x) \tag{5.2}$$

where $f(x)$ is referred as probability density function (pdf) if μ is Lebesgue measure or probability mass function (pmf) if μ is counting measure. Then we will further define the quantile function as follows:

$$Q(p) = F_X^{-1}(p) := \inf_{x\in\mathbb{R}} \{x \mid F_X(p) \geq y\}, \qquad 0 \leq p \leq 1. \tag{5.3}$$

where the domain of $Q(p)$ is bounded and the range may be infinite. If X has an absolutely continuous distribution (with respect to Lebesgue measure), $Q(F_X(x)) = x$ and $F_X(Q(p)) = p$.

One particular example is the standard Uniform distribution. The following shows the distribution, density and quantile respectively. If $Z \sim U(0,1)$, then

$$\begin{aligned} f_Z(z) &= \mathbb{1}_{z\in[0,1]} \\ F_Z(z) &= z\mathbb{1}_{z\in[0,1]} + \mathbb{1}_{z\geq 1} \\ Q_Z(y) &= y\mathbb{1}_{y\in[0,1]} \end{aligned} \tag{5.4}$$

where $\mathbb{1}_{\cdot}$ is the indicator function.

In this chapter, we propose to extend the IPLF with the Generalised Lambda distribution which is a natural generalisation of Tukey symmetric lambda distribution used in Spiring and Yeung (1998). This distribution with new parametrisation also gives a reasonable interpretation why it is defined such a way to cover any 4-parameter Generalised Lambda distribution. Numerical examples are also provided to demonstrate the applications of how to select the parameters and calculate the expected loss.

5.2 Generalised Lambda Distribution

The most general expression for the Generalised Lambda distribution with 5 parameters, λ_1, λ_2, $\lambda_3, \lambda_4, \lambda_5$, in short, GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$, is stated in terms of a quantile function of the form:

$$Q(p \mid \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = \left[\lambda_1 + \frac{1-\lambda_5}{\lambda_2\lambda_3}(p^{\lambda_3} - 1) - \frac{1+\lambda_5}{\lambda_2\lambda_4}((1-p)^{\lambda_4} - 1)\right] \mathbb{1}_{p\in[0,1]} \tag{5.5}$$

Evidently, when $\lambda_1 = 0$ and $\lambda_2 = 1$, $\lambda_3 = \lambda_4 = \lambda$ and $\lambda_5 = 0$, this GLD $(0, 1, \lambda, \lambda, 0)$ is the Tukey symmetric lambda distribution described in Tukey (1960) and Spiring and Yeung (1998). This form of 5-parameter GLD is briefly described as a natural generalisation of 4-parameter in Gilchrist (2002), but we use a different parametrisation. So far in the literature, there are two different parameterisations for 4-parameter GLD, both also defined by a quantile function as well. The first one was proposed by Ramberg and Schmeiser (1972, 1974) and the form is

$$Q(p \mid \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2}, \qquad 0 \leq p \leq 1, \tag{5.6}$$

whereas the second one was proposed by Freimer et al. (1988).

$$Q(p \mid \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{p^{\lambda_3} - 1}{\lambda_2\lambda_3} + \frac{(1-p)^{\lambda_4} - 1}{\lambda_2\lambda_4}, \qquad 0 \leq p \leq 1. \tag{5.7}$$

The 4 parameters control the shape, the location and the scale. In addition, GLD is an absolutely continuous distribution. It seems that this GLD is also a member of location-scale family, which is proved in the next theorem.

Theorem 5.2.1. *GLD* $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ *belongs to the location-scale family, where* λ_1 *is the location parameter and* λ_2 *is the scale parameter.*

Proof. Let $X \sim GLD\,(0, 1, \lambda_3, \lambda_4, \lambda_5)$, by Equation 5.5,

$$Q_X(p) = \left[\frac{1-\lambda_5}{\lambda_3}\left[p^{\lambda_3}-1\right] - \frac{1+\lambda_5}{\lambda_4}[(1-p)^{\lambda_4}-1]\right]\mathbb{1}_{p\in[0,1]} \tag{5.8}$$

Considering $X' = \lambda_1 + \frac{X}{\lambda_2}$, by the change of measure,

$$F_{X'}(x) = \mathbb{P}[X' \le x] = \mathbb{P}\left[X \le \big(x-\lambda_1\big)\lambda_2\right] = F_X\big((x-\lambda_1)\lambda_2\big) \tag{5.9}$$

and therefore by the absolutely continuity of GLD, let $F_X\big((x-\lambda_1)\lambda_2\big) = p$, we have

$$\begin{aligned} Q_X(p) &= \lambda_2(x-\lambda_1) \\ &= \left[\frac{1-\lambda_5}{\lambda_3}\left[p^{\lambda_3}-1\right] - \frac{1+\lambda_5}{\lambda_4}[(1-p)^{\lambda_4}-1]\right]\mathbb{1}_{p\in[0,1]} \\ x &= \left[\lambda_1 + \frac{1-\lambda_5}{\lambda_2\lambda_3}\left[p^{\lambda_3}-1\right] - \frac{1+\lambda_5}{\lambda_2\lambda_4}[(1-p)^{\lambda_4}-1]\right]\mathbb{1}_{p\in[0,1]} \end{aligned} \tag{5.10}$$

Similarly, taking the quantile function with respect to X', we have

$$Q_{X'}(p) = x = \left[\lambda_1 + \frac{1-\lambda_5}{\lambda_2\lambda_3}\left[p^{\lambda_3}-1\right] - \frac{1+\lambda_5}{\lambda_2\lambda_4}[(1-p)^{\lambda_4}-1]\right]\mathbb{1}_{p\in[0,1]} \tag{5.11}$$

This proves that X' also follows GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$, where λ_1 is the location parameter and λ_2 is the scale parameter. □

By Theorem 5.2.1, it is easy to know that λ_2 cannot be 0, because it is a scale parameter. The λ_1, λ_3 and λ_4 has no restrictions. In a special case of GLD, Tukey 3-parameter symmetric lambda distribution will converge to a standard logistic distribution when $\lambda_3 = \lambda_4 = \lambda \to 0$.

In this most general GLD, λ_1 is the location parameter and λ_2 is the scale parameter. For convenience, we only allow $\lambda_2 > 0$. While the left tail is controlled by λ_3, the right tail is controlled by λ_4. λ_5 is used to control some part of skewness and has to be bounded by -1 and 1. Without loss of generality, we set $\lambda_1 = 0$ and $\lambda_5 = -1$ and $\lambda_2 = 1$ to see whether λ_3 and λ_4 can be zero. Since

$$\lim_{\lambda_3 \to 0} Q(p) = \lim_{\lambda \to 0} \frac{p^{\lambda_3} - 1}{\lambda_3} = \log p,$$

we can remove the discontinuity at 0 by setting $(p^{\lambda_3} - 1)/p = \log p$ when $\lambda_3 = 0$. Similarly, this is also applied for λ_4. In sum, $\lambda_1, \lambda_3, \lambda_4 \in \mathbb{R}$, $\lambda_2 > 0$ and $\lambda_5 \in (-1, 1)$.

For being used in IPLF, the density is inevitable and necessary, but it is impossible to write the density in full for GLD. We also need the density form of this GLD and it is needed to have a further discussion on choosing the density as unimodal and bounded.

Lemma 5.2.2. *For the GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$, the probability density function for a random variable X with both $\lambda_3, \lambda_4 \neq 0$ is*

$$f_X(x) = \frac{\lambda_2}{(1 - \lambda_5) p^{\lambda_3 - 1} + (1 + \lambda_5)(1 - p)^{\lambda_4 - 1}} \mathbb{1}_{p \in (0,1)} \tag{5.12}$$

where $p = F_X(x)$.

Proof. Since by Theorem 5.2.1, GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ is a location-scale family. Hence, by the properties of location-scale family, we can only consider GLD $(0, 1, \lambda_3, \lambda_4, \lambda_5)$ without loss of generality as

$$f(x) = af(a(x + t))$$

where a and t are the scale and location parameter respectively.

Hence, for $X' \sim GLD\,(0, 1, \lambda_3, \lambda_4, \lambda_5)$, we have, by the Inverse Function Theorem and Equation 5.5 that,

$$\begin{aligned} f_{X'}(x) = F'_{X'}(x) &= \frac{1}{Q'_{X'}(F_{X'}(x))} \\ &= \frac{1}{(1-\lambda_5)\,[F_{X'}(x)]^{\lambda_3-1} + (1+\lambda_5)\,[1-F_{X'}(x)]^{\lambda_4-1}} \\ &= \frac{1}{(1-\lambda_5)p^{\lambda_3-1} + (1+\lambda_5)\,(1-p)^{\lambda_4-1}}\mathbb{1}_{p\in[0,1]} \end{aligned} \tag{5.13}$$

where $p = F_{X'}(x)$ and the general probability density function follows by multiplying the parameter λ_2 and transforming back to the original form of the random variable analogous to Theorem 5.2.1. □

Obviously, a function $f(x)$ is a probability density function if it satisfies the conditions of nonnegativity and almost surely boundedness and it integrates to 1 over the whole space. Since we require $\lambda_2 > 0$ and $\lambda_5 \in (-1, 1)$, under this parametrisation, GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ always have a vaild density and this situation is unlike Karian and Dudewicz (2000, 2011) that they have to specify which regions of the parameters (λ_3, λ_4) to make GLD valid.

Since $0 \leq p \leq 1$ in Equation 5.5, after considering the other cases and Lemma 5.2.2, we immediately have the following Theorem 5.2.3.

Theorem 5.2.3. *The GLD* $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ *of Equation 5.5 is valid in the parameter space where* $\lambda_1,\ \lambda_3,\ \lambda_4 \in \mathbb{R}$, $\lambda_2 > 0$ *and* $\lambda_5 \in (-1, 1)$. *In particular, the*

GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ *has a valid density as follows:*

$$f_X(x) = \begin{cases} \dfrac{\lambda_2}{(1-\lambda_5)p^{\lambda_3-1} + (1+\lambda_5)(1-p)^{\lambda_4-1}} \mathbb{1}_{p\in(0,1)} & \lambda_3, \lambda_4 \neq 0, \\ \dfrac{\lambda_2}{(1-\lambda_5)/p + (1+\lambda_5)(1-p)^{\lambda_4-1}} \mathbb{1}_{p\in(0,1)} & \lambda_3 = 0,\ \lambda_4 \neq 0, \\ \dfrac{\lambda_2}{(1-\lambda_5)p^{\lambda_3-1} + (1+\lambda_5)/(1-p)} \mathbb{1}_{p\in(0,1)} & \lambda_3 \neq 0,\ \lambda_4 = 0, \\ \dfrac{\lambda_2}{(1-\lambda_5)/p + (1+\lambda_5)/(1-p)} \mathbb{1}_{p\in(0,1)} & \lambda_3 = \lambda_4 = 0, \\ 0 & \textit{otherwise.} \end{cases} \tag{5.14}$$

where $0 \leq p = F_X(x) \leq 1$*.*

To find the density suitable for IPLF, the investigation of the possible shapes of the density is needed to ensure that there is only one relative extreme point. The point is also where the GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ has a relative maximum or relative minimum, but we also make sure that the point is a relative maximum.

Theorem 5.2.4. *The local extremes of the GLD* $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ *density occur at values of* p *where*

$$g(p) = \frac{p^{\lambda_3-2}}{(1-p)^{\lambda_4-2}} = \frac{(\lambda_4-1)(1+\lambda_5)}{(\lambda_3-1)(1-\lambda_5)} \tag{5.15}$$

for any $\lambda_3 \neq 1$ *and* λ_4 *in* $\mathbb{R}$ *without any restrictions.*

Proof. Since GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ is under location-scale family by Theorem 5.2.1, the location and shape parameter is irrelevant to the shape, we only consider the "standard" GLD $(0, 1, \lambda_3, \lambda_4, \lambda_5)$ without loss of generality. By Theorem 5.2.3, we have

$$f_X(x) = \begin{cases} \dfrac{1}{(1-\lambda_5)p^{\lambda_3-1} + (1+\lambda_5)(1-p)^{\lambda_4-1}} \mathbb{1}_{p\in(0,1)} & \lambda_3, \lambda_4 \neq 0, \\ \dfrac{1}{(1-\lambda_5)/p + (1+\lambda_5)(1-p)^{\lambda_4-1}} \mathbb{1}_{p\in(0,1)} & \lambda_3 = 0,\ \lambda_4 \neq 0, \\ \dfrac{1}{(1-\lambda_5)p^{\lambda_3-1} + (1+\lambda_5)/(1-p)} \mathbb{1}_{p\in(0,1)} & \lambda_3 \neq 0,\ \lambda_4 = 0, \\ \dfrac{1}{(1-\lambda_5)/p + (1+\lambda_5)/(1-p)} \mathbb{1}_{p\in(0,1)} & \lambda_3 = \lambda_4 = 0 \end{cases} \tag{5.16}$$

and 0 otherwise. Differentiating $f_X(x)$ with respect to x, we get

$$\begin{aligned} f'_X(x) &= \left(\frac{\mathbf{d}f}{\mathbf{d}p}\right)\left(\frac{\mathbf{d}p}{\mathbf{d}x}\right) = \left(\frac{\mathbf{d}f}{\mathbf{d}p}\right) f_X(x) \\ &= \begin{cases} -\frac{(1-\lambda_5)(\lambda_3-1)p^{\lambda_3-2}-(1+\lambda_5)(\lambda_4-1)(1-p)^{\lambda_4-2}}{\left[(1-\lambda_5)p^{\lambda_3-1}+(1+\lambda_5)(1-p)^{\lambda_4-1}\right]^3} \mathbb{1}_{p\in(0,1)} & \lambda_3, \lambda_4 \neq 0, \\ -\frac{-\frac{1-\lambda_5}{p^2}-(1-p)^{\lambda_4-2}(\lambda_4-1)(1+\lambda_5)}{\left[(1-\lambda_5)/p+(1+\lambda_5)(1-p)^{\lambda_4-1}\right]^3} \mathbb{1}_{p\in(0,1)} & \lambda_3 = 0,\ \lambda_4 \neq 0, \\ -\frac{-p^{\lambda_3-2}(-1+\lambda_3)(\lambda_5-1)+\frac{1+\lambda_5}{(1-p)^2}}{\left[(1-\lambda_5)p^{\lambda_3-1}+(1+\lambda_5)/(1-p)\right]^3} \mathbb{1}_{p\in(0,1)} & \lambda_3 \neq 0,\ \lambda_4 = 0, \\ -\frac{-\frac{1-\lambda_5}{p^2}+\frac{1+\lambda_5}{(1-p)^2}}{[(1-\lambda_5)/p+(1+\lambda_5)/(1-p)]^3} \mathbb{1}_{p\in(0,1)} & \lambda_3 = \lambda_4 = 0 \end{cases} \end{aligned} \tag{5.17}$$

and 0 otherwise.

Therefore, we have 4 different cases to be considered. Since $f_X(x) \geq 0$, we can get $f'_X(x) = 0$ by setting the numerator equal to 0.

Case A: $\lambda_3, \lambda_4 \neq 0$, $f'_X(x) = 0$ if and only if

$$g(p) = \frac{p^{\lambda_3-2}}{(1-p)^{\lambda_4-2}} = \frac{(\lambda_4-1)(1+\lambda_5)}{(\lambda_3-1)(1-\lambda_5)} \tag{5.18}$$

Case B: $\lambda_3 = 0,\ \lambda_4 \neq 0$, $f'_X(x) = 0$ if and only if

$$g(p) = \frac{1}{p^2(1-p)^{\lambda_4-2}} = -\frac{(\lambda_4-1)(1+\lambda_5)}{1-\lambda_5} \tag{5.19}$$

Case C: $\lambda_3 \neq 0,\ \lambda_4 = 0$, $f'_X(x) = 0$ if and only if

$$g(p) = p^{\lambda_3 - 2}(1-p)^2 = -\frac{1+\lambda_5}{(\lambda_3 - 1)(1-\lambda_5)} \tag{5.20}$$

Case D: $\lambda_3 = \lambda_4 = 0$, $f'_X(x) = 0$ if and only if

$$g(p) = \frac{(1-p)^2}{p^2} = \frac{1+\lambda_5}{1-\lambda_5} \tag{5.21}$$

Combining the results of the 4 cases, the theorem is obtained. □

Evidently, the shape of the GLD is controlled by the function g and we will refer g as the shape function of GLD. By investigating the shape function, we have the following theorem for the unimodality of GLD.

Before we can further discuss the shape function $g(\cdot)$, we have to rewrite $f_X(x)$ in terms of the quantile function $Q(p)$. It is easy to observe from Lemma 5.2.2 that

$$\begin{aligned} f'_X(x) &= \frac{\mathbf{d}}{\mathbf{d}x} f_X(x) = \frac{\mathbf{d}}{\mathbf{d}p}\left(\frac{1}{Q'_X(p)}\right)\left(\frac{1}{Q'_X(p)}\right) \\ &= -\frac{Q''_X(p)}{[Q'_X(p)]^3}\mathbb{1}_{p\in(0,1)} \end{aligned} \tag{5.22}$$

where $p = F_X(x)$. That is, $Q''_X(p) = 0 \iff g(p) = \frac{(\lambda_4-1)(1+\lambda_5)}{(\lambda_3-1)(1-\lambda_5)}$. Analogously, differentiating $f'_X(x)$ again with respect to x, we have by the chain rule,

$$\begin{aligned} f''_X(x) &= \frac{\mathbf{d}}{\mathbf{d}x} f'_X(x) = \frac{\mathbf{d}}{\mathbf{d}p}\left(-\frac{Q''_X(p)}{[Q'_X(p)]^3}\right)\left(\frac{1}{Q'_X(p)}\right) \\ &= \left(\frac{3\,[Q''_X(p)]^2\,[Q'_X(p)]^2 - [Q'_X(p)]^3\,Q'''_X(p)}{[Q'_X(p)]^6}\right)\left(\frac{1}{Q'_X(p)}\right) \\ &= \frac{3\,[Q''_X(p)]^2\,Q'_X(p) - [Q'_X(p)]^2\,Q'''_X(p)}{[Q'_X(x)]^6}\mathbb{1}_{p\in(0,1)} \end{aligned} \tag{5.23}$$

Thus, the quantile function $Q_X(p)$ and its derivatives can fully explain the shape, the convexity and the unimodality of GLD.

The focus is on a particular point x_0 such that $f'_X(x_0) = 0$ and$f_X(x_0) > 0$. Consequently, a corresponding point p_0 defined to be $p_0 = F_X(x_0)$ will have the condition that $g(p_0) = \frac{(\lambda_4-1)(1+\lambda_5)}{(\lambda_3-1)(1-\lambda_5)}$ and hence $Q''_X(p_0) = 0$. As a result, it can be simplified to

$$f''_X(x_0) = \left\{-\frac{Q'''_X(p_0)}{\left[Q'_X(p_0\right]^4}\right\} \mathbb{1}_{p\in(0,1)} \tag{5.24}$$

Apparently, f''_X has the opposite sign as Q'''_X when we want to know more about the stationary point x_0. For the convenience, we also show the third derivative of Q'''_X below.

$$Q'''_X(p) = \begin{cases} \begin{aligned} &-p^{-3+\lambda_3}\left(-2+\lambda_3\right)\left(-1+\lambda_3\right)\left(-1+\lambda_5\right) \\ &\quad +(1-p)^{-3+\lambda_4}\left(-2+\lambda_4\right)\left(-1+\lambda_4\right)\left(1+\lambda_5\right) \end{aligned} & \lambda_3, \lambda_4 \neq 0, \\ \frac{2-2\lambda_5}{p^3} + (1-p)^{-3+\lambda_4}\left(-2+\lambda_4\right)\left(-1+\lambda_4\right)\left(1+\lambda_5\right) & \lambda_3 = 0, \lambda_4 \neq 0, \\ -p^{-3+\lambda_3}\left(-2+\lambda_3\right)\left(-1+\lambda_3\right)\left(-1+\lambda_5\right) - \frac{2(1+\lambda_5)}{(-1+p)^3} & \lambda_3 \neq 0, \lambda_4 = 0, \\ 2\left(\frac{1}{(1-p)^3} + \frac{1}{p^3} + \left(\frac{1}{(1-p)^3} - \frac{1}{p^3}\right)\lambda_5\right) & \lambda_3 = \lambda_4 = 0. \end{cases} \tag{5.25}$$

where 0 otherwise and $Q'''_X(x)$ is only well defined while $0 < p < 1$.

We now can prove a theorem on the unimodality of GLD with both conditions from the properties of shape function g defined in Equation 5.15 and $Q'''_X(p)$.

Theorem 5.2.5. *The GLD* $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ *is unimodal when both* $\lambda_3, \lambda_4 > 2$ *or both* $\lambda_3, \lambda_4 < 1$.

Proof. First of all, we have to ensure that GLD has a unique local extremum. By Theorem 5.2.4, we have the shape function g in Equation 5.15 with the form

$$g(p) = \frac{p^{\lambda_3 - 2}}{(1-p)^{\lambda_4 - 2}} \mathbb{1}_{p \in (0,1)} \tag{5.26}$$

Differentiating g again with respect to p,

$$\begin{aligned} g'(p) &= (1-p)^{1-\lambda_4} x^{\lambda_3 - 3} \left[-2 + (p-1)\lambda_3 + p\lambda_4\right] \mathbb{1}_{x \in (0,1)} \\ &= C\left[(p-1)(\lambda_3 - 2) + p(\lambda_4 - 2)\right] \mathbb{1}_{p \in (0,1)} \end{aligned} \tag{5.27}$$

where $C \geq 1$. Apparently, when both $\lambda_3, \lambda_4 > 2$, $g'(p) > 0$ for all $p \in (0,1)$. It is obvious that g is a also continuous function.

$$\lim_{p \to 0+} g(p) = 0 \qquad \text{and} \qquad \lim_{p \to 1-} g(p) = \infty \tag{5.28}$$

Therefore, $g(x)$ is monotonic increasing when both $\lambda_3, \lambda_4 > 2$ and $g(p)$ pass through each point exactly once. $g(p)$ holds the only critical point of the density of the GLD once such that there exists only one point $\tilde{p}$ with $g(\tilde{p}) = \frac{(\lambda_4 - 1)(1+\lambda_5)}{(\lambda_3 - 1)(1-\lambda_5)}$.

However, it does not tell us that the relative extremum is a maximum or a minimum. By Equation 5.24 and Equation 5.25, $f''_X(x) < 0 \iff Q'''_X(p) > 0$. That is, suppose $Q_X(\tilde{p}) = \tilde{x}$, $f''_X(\tilde{x}) < 0$ for all p as $\lambda_3, \lambda_4 > 2$, this unique local extreme $\tilde{x}$ is the mode of the GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$.

Similarly, when both $\lambda_3, \lambda_4 < 2$, $g' < 0$ for all $p \in (0,1)$. In addition, $\lim_{x \to 0+} g(p) = \infty$ and $\lim_{x \to 1-} g(p) = 0$. Therefore, $g(x)$ is monotonic decreasing when both $\lambda_3, \lambda_4 < 2$ and leading to a unique solution of $g(p) - \frac{(\lambda_4 - 1)(1+\lambda_5)}{(\lambda_3 - 1)(1-\lambda_5)} = 0$. However, $Q'''_X(p) > 0$ as both $\lambda_3, \lambda_4 < 1$ such that $f''_X(\tilde{x}) < 0$ if $\tilde{x} = Q_X(\tilde{p})$ with

$\tilde{p}$ is the solution. Combining the two conditions, $\lambda_3, \lambda_4 < 1$ will also guarantee the unimodality of GLD. □

The Theorem 5.2.5 recognises which regions of parameters will get a unimodal shape for GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$. Moreover, by Gilchrist (2002), the tails of the density of X are also controlled by λ_3 and λ_4 respectively. If $\lambda_3 \leq 0$, the left tail range will be $(-\infty, \lambda_1)$, whereas $\lambda_4 > 0$, the left tail range will be $(\lambda_1 - \lambda_2/\lambda_3, \lambda_1)$. Analogously, if $\lambda_4 \leq 0$, the right tail range will be (λ_1, ∞), whereas $\lambda_4 > 0$, the right tail range will be $(\lambda_1, \lambda_1 + \lambda_2/\lambda_4)$. That is, a GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ with unique mode can capture both unbounded support and bounded support.

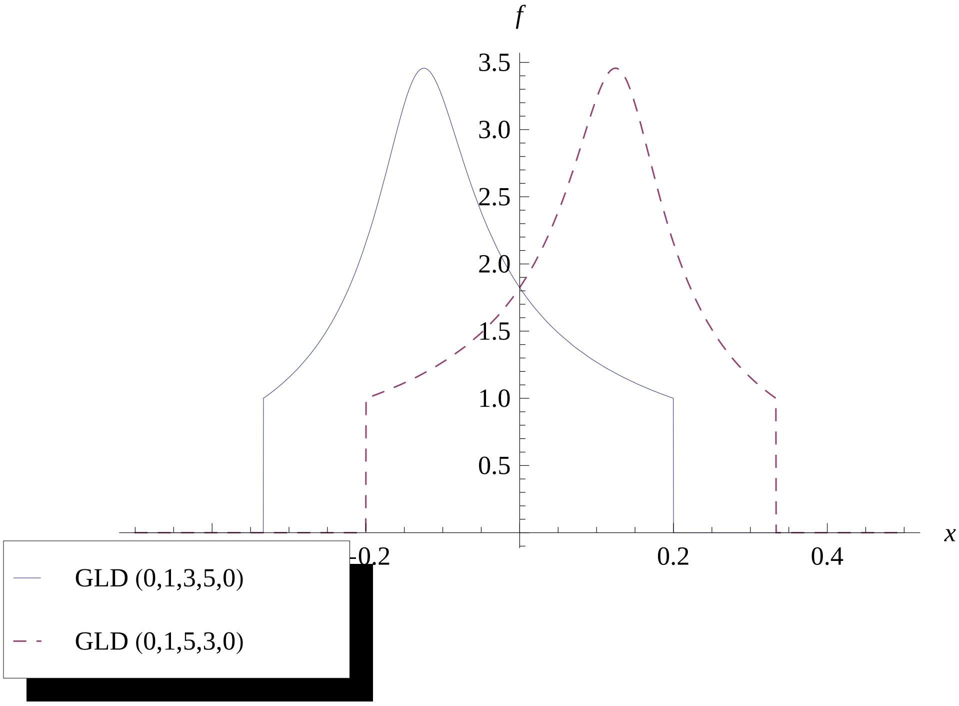


Figure 5.1: Densities of GLD $(0, 1, 3, 5, 0)$ and GLD $(0, 1, 5, 3, 0)$

We can further show our examination by using the reciprocal rule of the quantile function. Suppose $Y = 1/X$ and let the pth-quantile of Y be y_p with a corresponding value x_p

$$p = F_Y(y_p) = \mathbb{P}(Y \leq y_p) = \mathbb{P}(X > x_p) = 1 - F_X(x_p) \tag{5.29}$$

Clearly, $y_p = 1/x_p = 1/Q_X(1-p)$ and we obtain $Q_X(1-p)Q_{1/X}(p) = 1$.

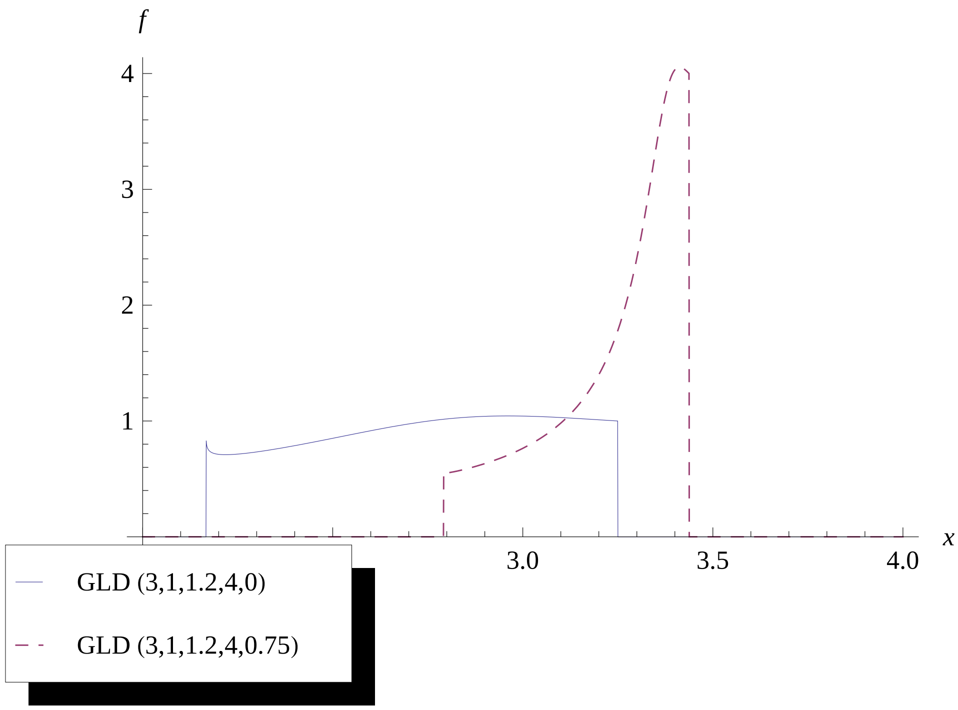


Figure 5.2: Densities of GLD $(3, 1, 1.2, 4, 0)$ and GLD $(3, 1, 1.2, 4, 0.75)$

In general, $\lambda_5 \neq 0$, the two tails of GLD is not symmetric, but if $\lambda_5 = 0$, then

$$Q_{1/X}(p) = 1/Q_X(1-p) = 1/\left[\lambda_1 + \frac{1}{\lambda_2\lambda_4}(p^{\lambda_4} - 1) - \frac{1}{\lambda_2\lambda_3}((1-p)^{\lambda_3} - 1)\right] \mathbb{1}_{p\in[0,1]} \tag{5.30}$$

which is the reciprocal of the symmetric image of quantile function where the pair (λ_3, λ_4) is reversed. Hence, the reciprocal of X is described by the quantile function. The Figure 5.1 shows clearly the symmetric properties about λ_3 and λ_4 where the pair of parameters is interchanged . When λ_5 is very close to -1 or 1, it can nullify the effect of either λ_3 or λ_4 and amplify the other tail. It is possible to make a non-unimodal density may become unimodal, which is illustrated in Figure 5.2.

For the shape of GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ distributions, it can also be well studied with the shape function g, the second derivative of the quantile function, Q''_X and the third derivative of the quantile function, Q'''_X. This also matches the table provided in Gilchrist (2002) for 4-parameter GLD. With the similar approach in Theorem 5.2.5, the following result is summarised in the Table 5.1:

λ_3	λ_4	Distributional Form
$(-\infty, 1)$	$(-\infty, 1)$	Unimodal
$[1, \infty)$	$(-\infty, 1)$	Monotone decreasing
1	1	Uniform
$(1, 2]$	$[1, 2]$	U-shaped
$(2, \infty)$	$[1, 2]$	S-shaped
$(2, \infty)$	$(2, \infty)$	Unimodal

Table 5.1: Shapes of Generalized Lambda distributions

5.3 Application with IPLF

After discussing the Generalised Lambda distribution, we will discuss how to use in the IPLF framework. IPLF was proposed by Spiring and Yeung (1998) for a general class of loss. The main procedure is as follows.

Definition 5.3.1 (Inverted Probability Loss Function). *Suppose $f_i(x, \theta)$ be the probability density function (pdf) with a unique mode at $\hat{x}_i$ and a_i be the target value. Then, let $a_i = \hat{x}_i$ to make a transformation such that*

$$m_i = \sup_{x \in \mathcal{X}_i} f_i\,(x, \theta_i) = f_i(a_i, \theta_i) < \infty \qquad \forall i$$

The form of the Inverted Probability Loss Functions (IPLF) is proposed as

$$\forall x \in \mathcal{X}_i,\ L\left(x,\ \{a_1, a_2\}\right) = \begin{cases} K_1\left[1 - \frac{f_1(x,\theta_1)}{m_1}\right] & x < a_1 \\ 0 & a_1 \le x \le a_2 \\ K_2\left[1 - \frac{f_2(x,\theta_2)}{m_2}\right] & x > a_2 \end{cases} \tag{5.31}$$

where $\mathcal{X}_i$ is the support of the distribution $f_i(x, \theta)$ and $K_i > 0$ may be a constant or a function.

Evidently, this definition is a more general form. We can set $K_1 = K_2 = 1$ and $a_1 = a_2 = T$ without loss of generality. For GLD $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ able to suit IPLF, it is necessary to be unimodal and hence both $\lambda_3, \lambda_4 < 1$ or both $\lambda_3, \lambda_4 > 2$ by Theorem 5.2.5.

In order to develop the loss from different target from this distribution, we need to characterise the target of interest with this distribution, the λ_1 may be needed to be modified by choosing the target value $x(\tilde{p})$. λ_1 is the location parameter by Theorem 5.2.1. That is,

$$x(p) = \left[\lambda_1 + \frac{1-\lambda_5}{\lambda_2\lambda_3}\left[p^{\lambda_3} - 1\right] - \frac{1+\lambda_5}{\lambda_2\lambda_4}[(1-p)^{\lambda_4} - 1]\right]\mathbb{1}_{p\in[0,1]} \tag{5.32}$$

Obviously, by Theorem 5.2.4, for any local extrema, they also satisfy the equality in Equation 5.15. The equation is not solvable by hand, since it is a nonlinear equation without knowing the degrees, which is a situation unlike the 3-parameter Tukey symmetric lambda density in Spiring and Yeung (1998). Although it is not solved algebraically without specifying λ_3, λ_4 and λ_5, it guarantees the existence of the solution and is solvable after realising the parameter values.

For instance, if we know that $(\lambda_2, \lambda_3, \lambda_4, \lambda_5) = (1, 0.5, -2, -0.1)$, then it is appropriate by Theorem 5.2.5 for being used in IPLF. It is easily figured out that the support of the density is $(-2, \infty)$. From Equation 5.15, we can also solve the optimal $\tilde{p}$. By machine solving, $\tilde{p} = 0.194479$. As a result, by Lemma 5.2.2,

$$\begin{aligned} \sup f(x(p) \mid \lambda_2, \lambda_3, \lambda_4, \lambda_5) &= \frac{\lambda_2}{(1-\lambda_5)\tilde{p}^{\lambda_3 - 1} + (1+\lambda_5)(1-\tilde{p})^{\lambda_4 - 1}} \\ &= 0.237177 \end{aligned} \tag{5.33}$$

Then, the target value has to be chosen to get $x(\tilde{p})$, say T, so

$$\begin{aligned} \lambda_1 &= x(\tilde{p}) - \frac{1-\lambda_5}{\lambda_2\lambda_3}\left[\tilde{p}^{\lambda_3} - 1\right] + \frac{1+\lambda_5}{\lambda_2\lambda_4}[(1-\tilde{p})^{\lambda_4} - 1] \\ &= T + 0.986287 \end{aligned} \tag{5.34}$$

This is also the one mistake often made to assume that the $\lambda_1 = T$.

By the Spiring–Yeung framework, he IPLF based on this distribution can be written as

$$\begin{aligned} L(x(p), T) &= K\left\{1 - \frac{f_X(x)}{0.237177}\right\} \\ &= K\left\{1 - \frac{1}{0.237177[1.1p^{-0.5} + 0.8(1-p)^{-3}]}\right\} \end{aligned} \tag{5.35}$$

Therefore, the corresponding distribution for this IPLF is GLD $(T + 0.986287,$ $1,$ $0.5,$ $-2,$ $-0.1)$. The associated graph, Figure 5.3 and Figure 5.4 illustrate the loss from the Generalised Lambda distribution.

Definitely, the Inverted Tukey loss in Spiring and Yeung (1998) is a special case of this IPLF from the Generalised Lambda distribution, as Tukey distribution

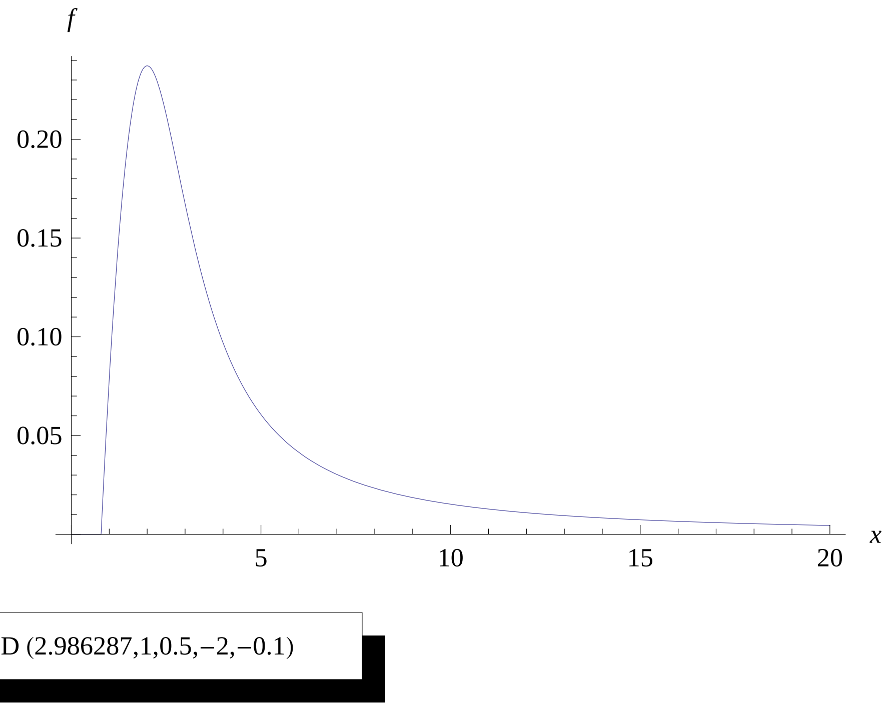


Figure 5.3: Density of GLD $(T + 0.986287,\, 1,\, 0.5,\, -2,\, -0.1)$ with $T = 2$

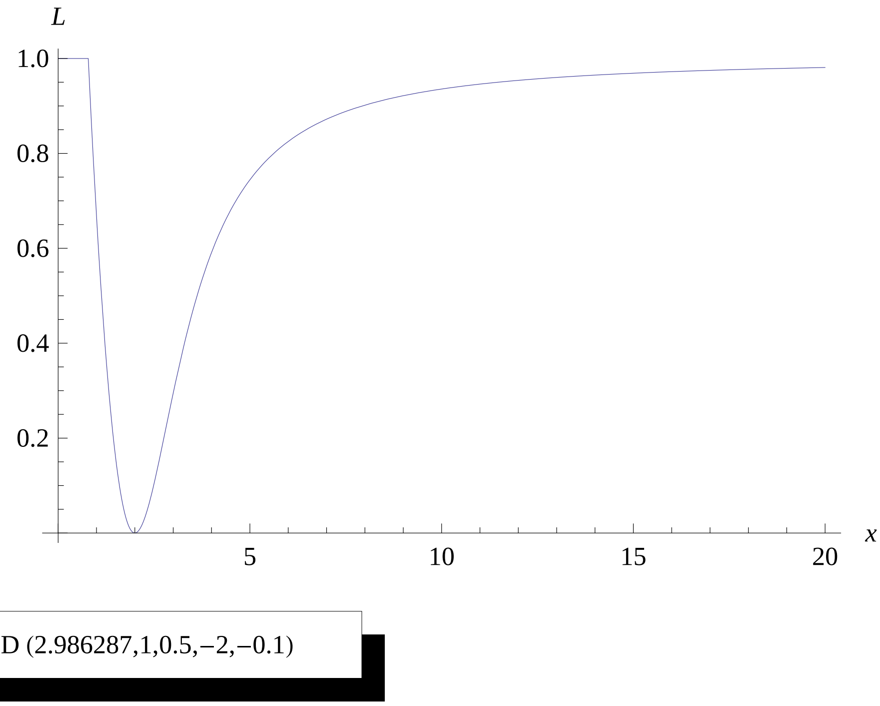


Figure 5.4: Associated IPLF from GLD $(T + 0.986287,\, 1,\, 0.5,\, -2,\, -0.1)$ with $T = 2$

is GLD $(\lambda_1, 1, \lambda, \lambda, 0)$ in the nonstandard form. With the unbounded support, λ_3, λ_4 has to be chosen both less than 0. With the bounded support, λ_3, λ_4 can be both chosen to be inside $(0, 1)$ or both greater than 2. This is one of the most flexible family of distributions so that it may ease the problem that the IPLF not able to extend to infinite support.

Referring to the data set for illustration in Section 1.2, if $T = 55$, the associated density for loss is GLD $(T + 0.986287,\ 1,\ 0.5,\ -2,\ -0.1)$ and $K = 0.3$, the expected loss $\mathbb{E}[L(X, T)] = 0.271$, with a similar value as Ryan loss, which are both generally much less than Taguchi loss.

5.4 Conclusion

This chapter introduced the new 5-parameter Generalised Lambda distribution and discussed some properties of this distribution. Moreover, the IPLF formed by this distribution extended the class of loss functions to a new level that IPLF can also work in a variable support.

Chapter 6

Applications of IPLFs

This chapter presents a few applications by using IPLFs to illustrate some related concepts. Under different contexts, the IPLF covers often too large for different problems. We would choose Spiring INLF, the IPLF derived from Gaussian density, as a particular example.

6.1 Process Capability Index with Exponential Squared Loss

6.1.1 Introduction

Loss function is a recent well-known useful tool in making a decision or evaluating a decision rule in situations where uncertainties are involved. Since the introduction of Taguchi philosophy (Taguchi, 1986; Taguchi et al., 1989), the loss function has been adopting by decision theoretic statisticians and economists for many years. We briefly introduce the relationships between both the loss functions and process capability indices.

Taguchi has introduced the quadratic loss function (Taguchi, 1986) to illustrate the need in consideration of target while assessing quality since 1986. He promotes the use of loss functions by suggesting that small deviations from the target result in a loss of quality. However, criticisms have been addressed to the Taguchi quadratic loss function by certain experts (Berger, 1985; Tribus and Szonyi, 1989; Box and Tiao, 1992). Some modifications by truncating the quadratic loss function at the points where the function intersects the maximum loss were also promoted (Tribus and Szonyi, 1989). Abdolshah et al. (2009) made good use of Taguchi loss function together with a capable process reject rates to develop a new process capability index, Taguchi-based Process Capability Index (TPCI).

Kane (1986) stated that capability indices were receiving increased usage in process assessments and purchasing decisions in the automotive industry, and the indices were of particular interest. Johnson (1992) mentioned that these indices were not related to cost failing to customer desires though these indices were simple to compute and are convenient for use by quality professionals because they were based on traditional specification limits. Chen and Chou (2001) extended the main work from Johnson (1992) and this work was the first time to explore the relationship between process capability indices and expected square error losses. Leung et al. (2012) even found out the relationship of all current PCIs with expected weighted squared error losses instead of just simple squared error losses.

In this study, we try to use the Spiring–Yeung framework of loss functions to create certain new PCIs, which can compare with TPCI. Numerical examples are also provided to demonstrate the applications of each loss function associated with each PCI used.

6.1.2 Loss-based Process Capability Index

In general, all PCIs are some ratios between the difference of the specification limits and the variabilities of the processes subject to the target. For example, from Kane (1986),

$$\begin{aligned} C_p &= \frac{U' - L'}{6\sigma} \\ C_{pk} &= \frac{\min\{U' - \mu,\, \mu - L'\}}{3\sigma} \end{aligned} \tag{6.1}$$

where U' and L' are the upper specification limit and lower specification limit respectively. In practice, both U' and L' are pre-known constants. If there are target T, probably T can replace μ for calculating the most appropriate PCI.

PCI is always a linear comparison between the difference of the specification limits and the actual variation with or without some modifications with respect to the target. However, the comparison between the loss is neglected. PCI should be able to capture the capability of a process with respect to the loss within the limits in comparison with the actual expected loss of the process. When the process is capable, the expected loss will be small when comparing with the expected loss within the limits.

Since we always have a bounded loss and the loss is quantifiable in an infinite support, the Taguchi-loss based Process Capability Index proposed in Abdolshah et al. (2009) will overestimate the loss involved. Therefore, we propose a similar loss-based Process Capability Index with Spiring–Yeung Inverted probability loss based on normal distribution such that it is more accurate to depict the loss with a more reasonable ground.

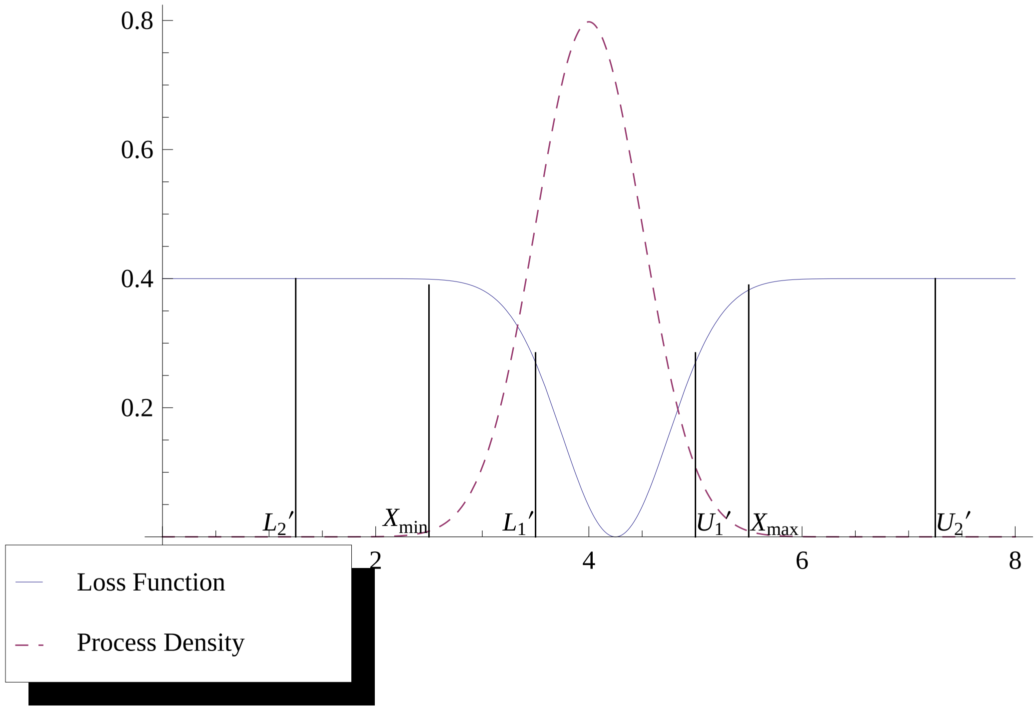


Figure 6.1: Different limits for a process and loss functions with $T = 4.5$

6.1.3 Spiring–Yeung Inverted Probability based PCI

Figure 6.1 shows a particular process with different pairs of lower and upper limits. For any processes, the process mean may not be equivalent to the target value we need and the sample data is often within the range from $X_{\min}$ and $X_{\max}$. Definitely, if the limits are (L_1', U_1'), then the process is incapable by any means of PCIs. Analogously, while the tolerance is larger such that the limits are extended to (L_2', U_2'), the process will be capable. In general, any PCIs are used to compare the interval between the variance with some penalty subject to target and the specification limits, such as C_p.

If we focus on the area of the loss involved under the loss function, it is evident to see that the area covered between $X_{\min}$ and $X_{\max}$ is smaller than the area of the limits if the process is capable. Therefore, in our opinion, it is also satisfactory to compare the area under the loss.

Given that Smith (1987) had already proposed the same loss as the exponential squared error loss before Spiring (1993), we would like to downplay the importance of Spiring and call it exponential squared error loss, but in short, we still call it Spiring loss.

The process mean may not be equal to the target in general. Hence, we need to make the loss function being 0 when it hits exactly on target instead of on the mean. The exponential squared error loss-based PCI is defined to be

$$ESPCI = \frac{\int_{[L',U']} L(x,T)\, \mathbf{d}F_X}{\int_{[X_{\min},X_{\max}]} L(x,T)\, \mathbf{d}F_X} \tag{6.2}$$

where F_X is the probability distribution of X and $L(x,T) = K\left\{1 - \exp\left(-\frac{(x-T)^2}{2\gamma^2}\right)\right\}$ with T being the target and K being the maximum loss.

If $X \sim N(\mu, \sigma^2)$, then both the denominator and the nominator terms are rather complicated. For example, the nominator term for $[L', U']$ is

$$\begin{aligned}
&\int_{[L',U']} L(x,T)\, \mathbf{d}F_X \\
=&\frac{1}{2}\Bigg\{-\text{Erf}\left[\frac{L'-\mu}{\sqrt{2}\sigma}\right] + \text{Erf}\left[\frac{U'-\mu}{\sqrt{2}\sigma}\right] \\
&+\frac{e^{-\frac{(T-\mu)^2}{2(\gamma^2+\sigma^2)}}\gamma\left(\text{Erf}\left[\frac{\gamma^2(L'-\mu)+(L'-T)\sigma^2}{\sqrt{2}\gamma\sigma\sqrt{\gamma^2+\sigma^2}}\right] - \text{Erf}\left[\frac{\gamma^2(U'-\mu)+(-T+U')\sigma^2}{\sqrt{2}\gamma\sigma\sqrt{\gamma^2+\sigma^2}}\right]\right)}{\sqrt{\gamma^2+\sigma^2}}\Bigg\}
\end{aligned} \tag{6.3}$$

where $\text{Erf}[x] = \int_{(-\infty,x)} \frac{2\exp(-t^2)}{\sqrt{\pi}}\, \mathbf{d}t$ and it is approximated by its Taylor–Maclaurin function below.

$$\int_{(-\infty,x)} \frac{2\exp(-t^2)}{\sqrt{\pi}}\, \mathbf{d}t \approx \frac{2}{\sqrt{\pi}}\left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216} - \frac{x^{11}}{1320} + \cdots\right) \tag{6.4}$$

The denominator term is also similar to the nominator term but the integration range changes from $[L', U']$ to $[X_{\min}, X_{\max}]$.

Except the exponential squared loss, $L(x, T)$ can be replaced by any losses within the class of Spiring–Yeung framework (Spiring and Yeung, 1998). The above PCI with exponential squared loss can be seemed as an example or illustration on how to create a IPLF-based PCI.

6.1.4 Numerical Example

Referring back to Section 1.2, we have a data set for pull strength and it follows $N(49.40, 4.69^2)$. Some information is summarised in Table 6.1.

Number of Data	56
Mean, $\bar{X}$	49.40
SD, s	4.79
Min, $X_{\min}$	40.6
Max, $X_{\max}$	58.6

Table 6.1: Summary of the pull strength for PCI

With the loss specified as exponential squared loss with $\gamma = 2$, the specification for the loss is also summarised in Table 6.2.

Upper limit, U'	60
Lower limit, L'	40
Target, T	55

Table 6.2: Specifications of PCI

Since U' and L' are located similar to the location of $X_{\max}$ and $X_{\min}$, the process is expected to be capable. The ESPCI is also calculated as follows:

$$\begin{aligned} ESPCI =& \frac{\int_{[40,60]} L(x,55)\,\mathbf{d}\widehat{F_X}}{\int_{[40.6,58.6]} L(x,55)\,\mathbf{d}\widehat{F_X}} \\ =& 1.026 \end{aligned} \tag{6.5}$$

which matches our expectation that this process is capable.

6.1.5 Conclusion

In this part, a certain extension and a new series of PCI based on Spiring–Yeung IPLF in Spiring and Yeung (1998) was designed. Besides, it is also an extension of TPCI proposed in Abdolshah et al. (2009) in which the mean of the process may be different from the target of the loss. This new PCI is also from a bounded loss and hence it strikes a balance between being realistic and sensitive to the loss function.

6.2 On an Admissibility Problem

6.2.1 Introduction

Except quality control, one of the widest uses of loss function is the statistical estimation of some parameters. This can be seemed as a decision problem and it is also well studied in many monographs and books such as Smith (1987); Berger (1985); Leonard and Hsu (2001); Robert (2001).

Suppose there exists a random variable X depending on the parameter θ, where $\theta \in \Theta$, a well-behaved parameter space. To choose the best parameter, one of the most common decision rule is a rule having the smallest expected loss in the whole Θ with a particular loss function. In general, the parameter Θ is too large and so a certain principles further shrink the space. One of the basic principles, the admissibility principle of the parameter, reduces the choice of the best parameter.

In this part, we limit to study the statistical models and decision problems having special invariance properties. Under this situation, by adopting the invariance principle, both the family of distributions and the search for best decisions are restricted to be in the class of the same invariance properties (Liese and Miescke, 2008). The invariance property we discuss is the location invariance in particular. t. That is, the group of measurable transformation is a homemorphism and we use the notation in Liese and Miescke (2008) to represent the location invariance. Therefore, the group of transformation is

$$\mathcal{U}_l = \{\mathbb{R}_{\oplus n} \mid x + c\mathbf{1},\ x \in \mathbb{R}^n, c \in \mathbb{R}\} \tag{6.6}$$

where $\mathbf{1} = (1, \cdots, 1)^\intercal$ and $\mathbb{R}_{\oplus n}$ as an additive group.

Let $X_1, \cdots, X_n$ be some random variables from a density $f_\theta(x) = f(x - \theta)$. Hence, if the loss is of the form $L(\theta, d) = L(d - \theta)$, then the whole class is location invariant.

It is also possible to use the IPLF in the parameter space, but in general IPLF which is bounded in the domain makes the problem too difficult and complex. Therefore, we have to introduce some transformations on IPLF with the unbounded support such that it is smooth enough in the whole real plane.

6.2.2 Exponentiated Loss from IPLF

For most unbounded loss, such as quadratic losses, one of the most enjoyable properties is the mathematical tractability and easy computability in most cases. Although it has been heavily criticised by many authors, say, Berger (1985); Lehmann and Casella (1998); Robert (2001), this main feature outperforms the distortion or the deviation of the utility function underlying. We believe that the main reason to drive this features is due to the smoothness of quadratic function in the whole real plane.

In Berger (1985), the utility function is derived from some axioms and by definition, the loss is defined to be an affine transformation of the negative of the utility function. Therefore, loss function is well defined. Quadratic losses then implies quadratic utility which is not acceptable or realistic. As a result, some researchers tried to propose some loss functions having the smoothness in the full real plane while having to keep the loss more realistic and less distorted.

Smith (1980, 1987) and Spiring (1993) also understood the inappropriateness of the unboundedness in most cases and wanted to prevent the derivation of the utility function behind. They both suggested using an exponential square loss instead, while Spiring (1993) thought that it was a modification of Gaussian density. It is to consider any exponentiated loss with the initial loss being unbounded and smooth like squared loss. After exponentiating the loss, the new loss is bounded, smooth and absolutely continuous. With the more general Spiring–Yeung framework of losses in Equation 3.2, Inverted Probability Loss Functions (IPLFs), and Theorem 3.2.1, there always exists an associated density such that the loss can fit in IPLF. The general form of the exponentiated loss by setting $D(\theta, d)$ be the unbounded loss is

$$L(\theta, d) = K\left\{1 - \exp\left[-\frac{D(\theta, d)}{C}\right]\right\} \tag{6.7}$$

where $K > 0$ and $C > 0$ and $\exp\left[-C^{-1}D(\theta, d)\right]$ will be a probability density kernel. C is considered to be a constant and d denotes a decision function to estimate θ based on some random variables X_i. Without loss of generality, we also set $K = 1$.

If $D(\theta, d)$ is a quadratic error loss, then

$$D(\theta, d) = (d - \theta)^2. \tag{6.8}$$

It is verified that $\int_{\mathbb{R}} \exp[-C^{-1}(d-\theta)^2]\,\mathbf{d}\theta = \sqrt{\pi C}$ and $D(\theta, d) = 0 \iff L(\theta, d) = 0 \iff \theta = d$. The associated density of this loss function is

$$\Theta \sim N(d, \sqrt{C/2}). \tag{6.9}$$

Analogously, if $D(\theta, d)$ is a absolute loss, then

$$D(\theta, d) = |d - \theta|\,. \tag{6.10}$$

It is also able to verify that $\int_{\mathbb{R}} \exp\left[-C^{-1}\,|d-\theta|\right]\,\mathbf{d}\theta = 2C$ and $D(\theta, d) = 0 \iff L(\theta, d) = 0 \iff \theta = d$. Hence, it is clear to understand that the associated density of this loss function is

$$\Theta \sim La(d, C), \tag{6.11}$$

where $La(\cdot)$ is the Laplace distribution.

Further, if $D(\theta, d)$ is a conventional LINEX loss, then

$$D(\theta, d) = e^{a(d-\theta)} - a(d-\theta) - 1, \qquad a \neq 0. \tag{6.12}$$

It is easy to check that $\int_{\mathbb{R}} \exp\left[-C^{-1}\left\{e^{a(d-\theta)} - a(d-\theta) - 1\right\}\right]\,\mathbf{d}\theta = 1/\,|a|\;e^{1/C}$ $C^{1/C}\Gamma\left(\frac{1}{C}\right)$ where $\Gamma(\cdot)$ is a Gamma function. $D(\theta, d) = 0 \iff L(\theta, d) = 0 \iff$

$\theta = d$. The associated density of this loss function is

$$\begin{aligned} f(\theta) =& \frac{\exp\left[-C^{-1}\left\{e^{a(d-\theta)} - a(d-\theta) - 1\right\}\right]}{1/\left|a\right| e^{1/C} C^{1/C} \Gamma\left(\frac{1}{C}\right)} \\ =& \frac{\left|a\right| C^{-\frac{1}{C}}}{\Gamma\left(\frac{1}{C}\right)} \exp\left\{\frac{a(\theta-d) - e^{a(\theta-d)}}{C}\right\} \\ =& \frac{\left|a\right| C^{-\frac{1}{C}}}{\Gamma\left(\frac{1}{C}\right)} \exp\left\{\frac{-z - e^{-z}}{C}\right\}, \qquad z = -a(d-\theta) \end{aligned} \tag{6.13}$$

which seems that it is a transformation of a Gumbel distribution where the parameter $a \neq 0$ instead of simply $a > 0$. If $a > 0$, it is clearly a conventional Gumbel distribution as $C = 1$.

Hence, all the exponentiated losses under IPLF are bounded and the associated density may be found. These exponentiated losses does not only preserve the properties of the original loss, such as symmetric and location invariant, but they also make the tractability possible, which is not possible in those pretending the loss in a bounded support.

6.2.3 Estimation

Suppose X have a random variable following Gaussian distribution with a real unknown mean θ and known variance σ^2. Given some observed data $X_1, \cdots, X_n$, the posterior distribution is also Gaussian distribution with mean μ and τ^2, where μ and τ are functions of $\sum x_i$ and σ^2. In addition, Gaussian distribution belongs to the group $\mathcal{U}_l$.

Our aim is to estimate θ by a certain kind of exponentiated loss from IPLF in Equation 6.7. Since the exponentiated loss is bounded by 0 and 1, it is smooth enough and the support is in the full plane, it is always able to find the minimum θ to find the best estimator $\hat{\theta}$ by using the first order condition. We then use the same approach as Zellner (1986) in comparing LINEX and quadratic loss.

Let $\hat{\theta}$ be the estimator of θ. Under the exponentiated loss from IPLF, there exists a parameter $\hat{\theta}^*$ minimizing the posterior expected loss which is the form as

$$\begin{aligned}\mathbb{E}[L(\theta,\hat{\theta})] &= \int_{\mathbb{R}} L(\theta,\hat{\theta})\,\mathbf{d}F_{\Theta|X} \\ &= 1 - \frac{1}{\sqrt{2\pi\tau^2}}\int_{\mathbb{R}} \exp\left\{-\frac{D(\theta,\hat{\theta})}{C} - \frac{1}{2}\left(\frac{\theta-\mu}{\tau}\right)^2\right\}\mathbf{d}\theta \\ &= 1 - \frac{1}{\sqrt{2\pi\tau^2}}\int_{\mathbb{R}} \exp\left\{-\frac{D(\hat{\theta}-\theta,0)}{C} - \frac{1}{2}\left(\frac{\theta-\mu}{\tau}\right)^2\right\}\mathbf{d}\theta \\ &= 1 - \frac{1}{\sqrt{2\pi\tau^2}}\int_{\mathbb{R}} \exp\left\{-\frac{D(x,0)}{C} - \frac{1}{2}\left(\frac{x-\hat{\theta}+\mu}{\tau}\right)^2\right\}\mathbf{d}x\end{aligned} \tag{6.14}$$

where $C > 0$ and $J = -\frac{D(x,0)}{C} - \frac{1}{2}\left(\frac{x-\hat{\theta}+\mu}{\tau}\right)^2$.

As a result, we set the Bayes estimate under the exponentiated loss from IPLF as θ^* which is to minimise the expected loss with respect to $\hat{\theta}$. We also denote $\hat{\theta}^* = \arg\min_{\hat{\theta}} \mathbb{E}[L(\hat{\theta},\theta)]$. Certainly, since $D(x,0)$ is smooth enough, we can solve this by taking partial differentiation with respect to $\hat{\theta}$ and setting it to 0 to solve the following equation.

$$\begin{aligned}0 = \frac{\partial}{\partial\hat{\theta}}\mathbb{E}[L(\theta,\hat{\theta})] &= -\frac{1}{\sqrt{2\pi\tau^2}}\int_{\mathbb{R}} \exp J\left(\frac{\partial J}{\partial\hat{\theta}}\right)\mathbf{d}x \\ &= -\frac{1}{\sqrt{2\pi\tau^2}}\int_{\mathbb{R}} \exp J\left(\frac{x-\hat{\theta}+\mu}{\tau^2}\right)\mathbf{d}x\end{aligned} \tag{6.15}$$

Suppose we set $\kappa(\hat{\theta}) = \mathbb{E}[L(\hat{\theta},\theta)]$. Evidently, $\kappa(\cdot)$ is bounded and if $D(\cdot)$ is chosen to be differentiable, κ is also differentiable. Owing to the fact that the derivative of $\kappa(\cdot)$ changes from negative to positive, and $\kappa(\cdot)$ tends to 1 in either extremes, there exists only one $\hat{\theta}^*$ solving the equation as the minimum. In summary, under exponentiated loss of IPLF, the Bayes estimate is unique and solvable when dealing with the Gaussian likelihood with Gaussian prior problem.

6.2.4 Admissibility Under Noninformative Prior

Under noninformative prior, it is normally to choose $\hat{\theta} = \mu$, the frequentist least-square estimator. However, under the above exponentiated loss from IPLF, $\hat{\theta}^*$ will be chosen as the parameter, which is the Bayes estimator.

We assume the true loss is the frequentist quadratic loss. Therefore, the expected loss for $\hat{\theta} = \mu$ is

$$\begin{aligned} R_1(\theta, \hat{\theta}) =& \mathbb{E}[L_1(\theta, \hat{\theta})] \\ =& \frac{1}{\sqrt{2\pi\tau^2}} \int_{\mathbb{R}} [\hat{\theta} - \theta]^2 \exp\left\{ -\frac{1}{2} \left(\frac{\theta - \hat{\theta}}{\tau} \right)^2 \right\} \mathbf{d}\theta \\ =& \tau^2 \end{aligned} \tag{6.16}$$

Correspondingly, if $D(x, 0)$ is not in the form of x^2, then

$$\begin{aligned} R_1(\theta, \hat{\theta}^*) =& \mathbb{E}[L_1(\theta, \hat{\theta}^*)] \\ =& \frac{1}{\sqrt{2\pi\tau^2}} \int_{\mathbb{R}} [\hat{\theta}^* - \theta]^2 \exp\left\{ -\frac{1}{2} \left(\frac{\theta - \hat{\theta}}{\tau} \right)^2 \right\} \mathbf{d}\theta \\ =& \tau^2 + \left(\hat{\theta}^* - \hat{\theta} \right)^2 \\ >& \tau^2 \end{aligned} \tag{6.17}$$

Hence, the estimator $\hat{\theta}^*$ from exponentiated loss is not square-error admissible unless the form of $D(x, 0) = x^2$ in general.

In the opposite, we assume that the true loss is the exponentiated loss from IPLF. Therefore, the expected loss or risk function of exponentiated loss from IPLF for $\hat{\theta} = \mu$ is

$$\begin{aligned} R_2(\theta, \hat{\theta}) =& 1 - \frac{1}{\sqrt{2\pi\tau^2}} \int_{\mathbb{R}} \exp\left\{ -\frac{D(\theta, \hat{\theta})}{C} - \frac{1}{2}\left(\frac{\theta - \hat{\theta}}{\tau}\right)^2 \right\} \mathbf{d}\theta \\ =& 1 - \frac{1}{\sqrt{2\pi\tau^2}} \int_{\mathbb{R}} \exp\left\{ -\frac{D(x, 0)}{C} - \frac{1}{2}\left(\frac{x}{\tau}\right)^2 \right\} \mathbf{d}x \end{aligned} \tag{6.18}$$

which is free of θ.

Correspondingly, the risk of the exponentiated loss from IPLF for $\hat{\theta}^*$ is

$$R_2(\theta, \hat{\theta}^*) = 1 - \frac{1}{\sqrt{2\pi\tau^2}} \int_{\mathbb{R}} \exp\left\{ -\frac{D(x, 0)}{C} - \frac{1}{2}\left(\frac{x - \hat{\theta}^* + \mu}{\tau}\right)^2 \right\} \mathbf{d}x \tag{6.19}$$

which is also free of θ. That is, both risks are constant functions over the space $\theta \in \Theta$. However, since $\hat{\theta}^*$ is the optimized and the risk with respect to its loss is the lowest, $R_2(\theta, \hat{\theta}^*) < R_2(\theta, \hat{\theta})$. As a result, the estimator $\hat{\theta}$ is not exponentiated loss admissible.

Since exponentiated loss is a family containing many bounded losses and here the loss is set arbitrarily that it may be an exponentiated absolute loss or an exponentiated LINEX loss, the estimator admissible in one loss is not admissible in other loss obviously.

6.2.5 Conclusion

We develop and focus on a certain exponentiated loss which make some unbounded loss become bounded by using the Spiring–Yeung IPLF. This special class of bounded loss functions has some useful properties of the corresponding unbounded loss. Further, this class of losses is applied to a Bayesian estimation of normal mean and we study the admissibility of the estimator which is the argument of minimising the particular loss. It further shows that the best estimator of a particular loss is always inadmissible with respect to other losses.

Chapter 7

A General Class of Conjugate Loss

Inverted Probability loss was discovered by Spiring (1993) with the observation that the exponentiated square loss can be written as $1 - f(x)/f(T)$, where f is the Gaussian density and T is the mode. This loss has a property that some expected loss has a closed form when combined with another Gaussian distribution. However, it is from an *ad-hoc* approach without any rigorous grounds. In this chapter, we will explain it clearly that any analogous loss can be formed from the conjugate direction.

7.1 Introduction

In many statistical and decision problems, very little importance has been placed for the loss functions, but the choice of a particular loss function can seriously affect the results, such as estimation of parameters or inference. In making or evaluating a decision, the loss function is used in terms of the utility of the decision maker (Berger, 1985; Robert, 2001; Press, 2002). The loss is generally bounded and not convex, so the conventional quadratic loss is not adequate to reflect the true loss in the given situation.

Similarly, in the field of statistical quality control, Taguchi (1986) proposed the quadratic loss to estimate the actual economic loss. Spiring (1993) was the first one in this field adopting the Inverted Normal loss function instead of the quadratic loss, and therefore it is a more reasonable choice. Actually, Smith (1987) already proposed the same Spiring loss function as the exponential squared error loss before Spiring (1993), because of the symmetry and the boundedness in the properties of such loss functions (Leonard and Hsu, 2001). However, the technicalities sharply increase with the distributions other than the normal distribution when using the Inverted Normal loss function.

In this chapter, we first develop a class of conjugate loss functions which is bounded and similar to the loss function developed in Spiring and Yeung (1998). Other situations relevant to this class are also examined.

7.2 Statistical Decision Theoretical Framework

Beforehand, we have to set up the framework of the statistical decision theory utilised in the field of quality assurance and reliability settings. This statistical decision theory framework can be found in many literature, such as Liese and Miescke (2008); Robert (2001); Blackwell and Girshick (1979); Lehmann and Romano (2005); Smith (1987); Pace and Salvan (1997); Leonard and Hsu (2001); Berger (1985); French and Rios (2000). We choose to follow Ferguson (1967) and Blackwell and Girshick (1979) closely with their logical steps, but in a more general form.

Generally, in this field, the expected loss is the evaluation criterion to select the manufacturing processes, but the most controversy is the practical determination of the form of the loss function by the decision-makers. Therefore, the following is to fit the framework such that the systematic approach is available for further discussion.

Since the performance of a process is random and fluctuated, we have to derive and construct from the axiomatic system of the Probability theory proposed in Kolmogorov (1933). Clearly, in mathematical terms, it is a measurable space $(\Omega, \mathcal{F})$, where

- Ω is the set or space of all possible elementary events;
- $\mathcal{F}$ is the Borel σ-algebra of Ω, that is, the collection of the subsets of events. Evidently, Ω and $\emptyset = \Omega^{\complement}$ both belong to $\mathcal{F}$, that is, the situation of allowing all events occurs almost surely. For convenience, we denote $\mathcal{F} = \mathcal{B}_{\Omega}$.

For every set $A = \biguplus_{i=1}^{n} A_i \in \mathcal{F}$ with disjoint $A_1, \cdots, A_n \in \mathcal{F}$, there exists a σ-additive nonnegative function $\mathbb{P}$ such that $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A_i)$ and $\mathbb{P}(\Omega) = 1$. In other words, $\mathbb{P}$ is a probability on the measurable space $(\Omega, \mathcal{F})$.

Mostly the data is collected from the process under study through the observation of a real-valued random variable X, and the exact information of the probability space induced by X is unknown. Hence, statistics, especially parametric, is based on the statistical model $\mathcal{M} = \left(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}, \mathbb{P}_\theta = \mathbb{P}_\theta \circ X^{-1}\right)$, where $\theta \in \Theta$ for X to estimate the best unknown θ from the possible value in the set Θ. In other words, the major problem in statistics is to determine the best value $\hat{\theta}$ from the collected data of X and then to build up the best probability model $\left(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}, \mathbb{P}_{\hat{\theta}}\right)$ accordingly. The set $\mathfrak{P} = \{\mathbb{P}_\theta \mid \theta \in \Theta\}$ is the family of probability distributions with the parameter θ belonging to Θ as a condition. In general, $\mathcal{P}$ can be detached from the statistical model $\mathcal{M}$ and it is possible to verify between two sets of family of probability distributions $\mathfrak{P}$ and $\mathfrak{Q}$ without assuming a particular Θ on the common measurable space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

For decision-maker facing the statistical problem, he has his own thought and the decision space is $\mathscr{D}$, where the $\mathscr{D}$ is the set of all possible decisions. In particular, $\mathscr{D} = \Theta$ or $\mathscr{D} = \mathbb{R}^k$ is generally the most common case in some fields, such as quality assurance and reliability settings. Here, we assume that his action is made in one-to-one correspondence to his decision, and therefore the action space is equivalent to the decision space in the topological and measurable sense.

Definition 7.2.1 (Loss Functions)**.** *A loss function is any function $L(d, \theta)$ from $\mathscr{D} \times \Theta$ to $\mathbb{R}$. If $L(d, \theta)$ is bounded, it is proper; otherwise, it is improper.*

Since utility always exists with some axioms in DeGroot (2004); von Neumann and Morgenstern (2004), loss function can be simply interpreted as the negation of the utility and also exists in all situations. Therefore, the objective of maximizing the utility is the same as the objective of minimizing the loss. We further assume that loss functions are also bounded from below, that is,

$$\inf_{d \in \mathscr{D}} \inf_{\theta \in \Theta} L^*(d, \theta) \geq -C > -\infty \qquad C \in \mathbb{R}_+,$$

Clearly, with Definition 7.2.1, a proper loss function also satisfies the condition that $\sup_{D \in \mathscr{D}} L^*(d, \theta) \leq C$, where $C \in \mathbb{R}_+$. In many situations, it is more convenient to talk in terms of nonnegative losses, and hence it is also assumed in this dissertation that we focus on the following loss

$$L(d, \theta) = L^*(d, \theta) - C$$

instead. Note that the decision is unaffected when the loss is under affine transformation. Though the simple loss function being the negation of the utility, other types also appear in the literature. A *regret loss* function can also be adopted and defined as $L' = \sup_{d \in \mathscr{D}} U(d, \theta) - U(d, \theta)$, where $U(\cdot)$ is the utility function. This loss is seemed to be more realistic, as it measures the loss by not choosing the optimal choice without control of the occurrence of θ. Nevertheless, Berger (1985) claimed that $L,\ L^*,\ L'$ are equivalent in Bayesian analysis of the Statistical Decision Framework.

After constructing the loss function, it is to desire to find the optimal decision with respect to θ. In general, θ is too complex and it is impossible or very difficult task without a manipulation of θ at the time of decision making. Under a particular decision d, the set of observations $\{X = x\}$ influences the relative correctness of the estimation of θ and the efficiency. Further, X follows a set of distribution $\mathfrak{P}$. In other words, the decision d depends on the outcome of $\{X = x\}$ and is maintained as a function $d = d(X)$, where $d \colon \mathbb{R}^k \to \mathscr{D}$. Therefore, the loss function is constructed to make $(x, \theta) \mapsto L(d(x), \theta) \in \mathbb{R}$. From this viewpoint, we can describe the loss function as a random vector (d, X) on the statistical model. A natural method of determining the best decision is to select according to the expected loss or risk function over all possible X when the decision is made and to choose the one with the minimum expected loss.

Definition 7.2.2 (Frequentist Risk)**.**

Given a statistical model $\mathcal{M} = (\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}, \mathbb{P}_\theta)$ and a loss function $L : \mathscr{D} \times \Theta \to \mathbb{R}$, the expected loss or frequentist risk function (in short, frequentist risk) of a decision

$d \in \mathscr{D}$ is given by

$$R^*(d,\theta) = \mathbb{E}_\theta\, L(d,\theta), \qquad \theta \in \Theta,$$
$$= \int L(d(x),\theta)\,\mathbb{P}_\theta\,(\mathbf{d}x) \tag{7.1}$$
$$= \int L(d(x),\theta)\mathbf{d}F_\theta(x),$$

where $F_\theta(x) = \mathbb{P}_\theta\big(X \in (-\infty,x)\big)$, the cumulative distribution function of $X \mid \Theta = \theta$. In simpler words, the frequentist risk is taking an average of the loss function over $\mathbb{R}$ on the condition of a particular $\Theta = \theta$.

In Bayesian analysis, Θ also follows another prior distribution at the time of decision making. Hence, there is another probability space for θ, $(\Theta, \mathcal{B}_\Theta, \mathbb{T})$. However, in practice, it is very embarrassing to allow the improper prior that $\mathbb{T}(\Theta) = \infty$, such as noninformative prior and Jeffreys prior, while the improper prior appears frequently. Similar to Definition 7.2.2 and with the condition that Θ being σ-finite, we have the joint distribution of (Θ, X) as follows. Suppose there are two suitable dominating measures μ and ν for $\mathbb{P}_\theta$ and $\mathbb{T}$, the joint distribution of (X,Θ) is given by

$$\begin{aligned} \big[\mathbb{P}_\theta \otimes \mathbb{T}\big](A) &= \iint \mathbb{1}_A(x,\theta)\,\mathbb{P}_\theta(\mathbf{d}x)\mathbb{T}(\mathbf{d}\theta), \quad A \in \mathcal{B}_{\mathbb{R}^k} \otimes \mathcal{B}_\Theta \\ &= \iint \mathbb{1}_A(x,\theta) f(x \mid \theta)\pi(\theta)\mathbf{d}\big(\mu \otimes \nu\big) \end{aligned} \tag{7.2}$$

where $f(x \mid \theta)\pi(\theta)$ is the Radon-Nikodym derivative of $\mathbb{P}_\theta \otimes \mathbb{T}$ with respect to

$\mu \otimes \nu$. Therefore, the marginal distribution of $\Theta|X$ can be found as

$$
\begin{aligned}
\mathbb{T}^X(A) &= \int \mathbb{1}_A(\theta) \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)\mathbf{d}\nu} \nu(\mathbf{d}\theta), \qquad A \in \mathcal{B}(\Theta) \\
&= \int \mathbb{1}_A(\theta) \left(\int f(x \mid \theta)\pi(\theta)\mathbf{d}\nu \right)^{-1} \left(\frac{\mathbf{d}\left(\mathbb{P}_\theta \otimes \mathbb{T}\right)}{\mathbf{d}\left(\mu \otimes \nu\right)}(x,\theta) \right) \\
&= \int \mathbb{1}_A(\theta) \left(\frac{\mathbf{d}\left[(\mathbb{P}_\theta \circ X^{-1}) \otimes \mathbb{T}\right]}{\mathbf{d}\mu}(x) \right)^{-1} \left(\frac{\mathbf{d}\left(\mathbb{P}_\theta \otimes \mathbb{T}\right)}{\mathbf{d}\left(\mu \otimes \nu\right)}(x,\theta) \right)
\end{aligned}
\tag{7.3}
$$

and we can obtain the general Bayesian expected loss as the following definition:

Definition 7.2.3 (Bayesian Expected Loss)**.**

Given a statistical model $\mathcal{M} = (\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}, \mathbb{P}_\theta)$*, a parameter measurable space* $(\Theta, \mathcal{B}(\Theta), \mathbb{T})$*, a loss function* $L : \mathscr{D} \times \Theta \to \mathbb{R}$*, the Bayesian expected loss (in short, Bayes loss) of a decision* $d \in \mathscr{D}$ *with a observation* $X = x$ *is given by*

$$
\begin{aligned}
R(d,\theta) &= \int_\Theta L(d,\theta)\mathbb{T}^X(\mathbf{d}\theta), \qquad d \in \mathscr{D} \\
&= \mathbb{E}_X\, L(d(x),\theta).
\end{aligned}
\tag{7.4}
$$

In simpler words, Bayes loss is taking an average of the loss function over Θ *on the condition of* $X = x$*.*

Now, we have introduced the three ingredients rigourously enough to form the basis of the Bayesian Statistical Decision Theory. Summarising that the three major ingredients are:

- the set $\mathfrak{P}$ for the observations;
- the loss function $L(d(x), \theta)$ associated with the decisions; and
- the prior probability measure $\mathbb{T}_X$ given the observation $X = x$.

With the three ingredients, a statistical decision problem can be described in the following definition.

Definition 7.2.4 (Statistical Decision Problem)**.**

A statistical decision problem can always be described by a triple $(\mathcal{M}, \mathscr{D}, L)$*, where* $\mathcal{M}$ *is a statistical model,* $\mathscr{D}$ *is a decision space and* L *is a loss function.*

7.3 Bounded Loss and Advantages

A proper loss is a loss function bounded from below and bounded by above for all values of θ and d. The research target of this dissertation is mainly on the proper loss. Some problems may occur when the loss function is not proper. If the loss function is not bounded, the frequentist risk or the Bayes loss may be infinite or undefined for some values of or even all values of d. For instance, if the distribution of X is Cauchy(x_0, γ) given by

$$F_X(x|\theta, \gamma) = \frac{1}{\pi} \arctan\left(\frac{x-\theta}{\gamma}\right) + \frac{1}{2}, \qquad x \in \mathbb{R}$$

then it is easily checked that all moments are undefined and that the frequentist loss with respect to either conventional absolute loss or a quadratic loss function is always undefined regardless of the decision. Another possible example is when the distribution of X is Uniform$(0, \theta)$, the natural conjugate measure for Θ will be chosen as Pareto(x_m, α), whereas the concept of conjugate families is defined later. The distribution is given by

$$F_\Theta(\theta \mid x_m, \alpha) = \left[1 - \left(\frac{x_m}{\theta}\right)^\alpha\right] \mathbb{1}_{\theta \geq x_m}(\theta).$$

If $\alpha = 1$, it is easily shown that all moments higher than the first order are infinite and that the marginal distribution for $\Theta \mid X = x$ is Pareto$(\max\{x, x_m\}, 2)$. Therefore, any moments higher than the second order for this posterior distribution are infinite. The Bayes loss with respect to the conventional quadratic loss function is always infinite regardless of the decision, but that with respect to the absolute loss function works. In both cases, we cannot figure out the optimal decision or compare different decisions in any values of d.

Sometimes, the optimal decisions arising from the improper loss functions are not robust when the distributions of Θ and X are not absolutely correct and may change slightly. The more detailed example can be found in Robert (2001).

7.4 Conjugate Loss

Since the most common parametric distributions are also from the exponential family, we assume that the prior distribution is a member of exponential family for convenience. Examples of the exponential family are Beta, Dirichlet, Wishart, Gamma, Gaussian, Exponential and Poisson. This family can be characterised by its density with a suitable dominating measure.

As the exponential family is well-researched, many literature have discussed it fully. The main references on this topic are Lehmann and Casella (1998); Lehmann and Romano (2005); Liese and Miescke (2008). Some monographic treatments on exponential families can also be found in Barndorff-Nielsen (1978) and Brown (1986). We mainly follow the procedures of Brown (1986).

Similar in Section 7.2, we have to firmly base on a measurable space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ (because the random variable is real-valued) and a dominating measure μ such that we have a sounded mathematical support. For any $\eta \in \mathcal{H}$, $\mathcal{H}$ being the parameter space,

$$\mathbb{Q}_\eta(A) = \int \mathbb{1}_A(x) \exp\Big[\langle a(\eta), S(x)\rangle - \psi'(\eta)\Big] G(x)\mu'(\mathbf{d}x), \quad A \in \mathcal{B}(\mathbb{R}), \tag{7.5}$$

where $S(x)$ is called the statistic of X, $a(\cdot)$ and $G(x)$ are some well-behaved continuous functions. $\mathbb{Q}_\eta$ is the probability measure induced by the real-valued X and the whole set $\mathfrak{P} = \{\mathbb{Q}_\eta \mid \eta \in H\}$ is the exponential family. Then the density $f(x \mid \eta)$ can be defined as the Radon-Nikodym derivative with respect to μ', where

$$\begin{aligned} f(x \mid \eta) &= \frac{\mathbf{d}\mathbb{Q}_\eta}{\mathbf{d}\mu'}(x) \\ &= \exp\Big[\langle a(\eta), S(x)\rangle - \psi'(\eta)\Big] G(x), \quad x \in \mathbb{R},\ \eta \in \mathcal{H}. \end{aligned} \tag{7.6}$$

Brown (1986) noticed and proved that the aforementioned settings can be further reduced by focusing on $X = S(X)$ and reparametrising η with $\theta = a(\eta)$ and using another suitable dominating measure, say μ, to a minimal form or reduced form. Analogously, in practice it is often more convenient to study the family of induced distributions in Euclidean space, $\mathbb{R}^k$. Hence, the statistical model for the minimal form of the member of standard exponential family can then be rewritten as $\mathcal{M}' = (\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k}, \mathbb{P}_\theta = \mathbb{Q}_\eta \circ S^{-1})$. That is, for every $A \in \mathbb{R}^k$,

$$\mathbb{P}_\theta(A) = \left(\mathbb{Q}_\eta \circ S^{-1}\right)(A) = \int \mathbb{1}_A(s) \exp\Big[\langle \theta, s\rangle - \psi(\theta)\Big]\mu(\mathbf{d}s), \tag{7.7}$$

and

$$f(s \mid \theta) = \frac{\mathbf{d}\,\mathbb{P}_\theta}{\mathbf{d}\mu}(s) = \exp\big[\langle \theta, s\rangle - \psi(\theta)\big], \qquad s \in \mathbb{R}^k. \tag{7.8}$$

If there exists a $\mathbb{P}_\theta$ for each $\theta \in \Theta$, the family $\mathfrak{P} = \{\mathbb{P}_\theta \mid \theta \in \Theta\}$ is *full*. With a further condition of the openness of Θ in $\mathbb{R}^k$, *id est*, $\forall \eta \in \mathcal{H} : \exists \varepsilon > 0 \ni B_\varepsilon(\eta) \subseteq \mathcal{H}$, then it is called as the *regular* exponential family. Under the regularity and fullness conditions, the distribution is in minimal form if and only if it belongs to the *natural* exponential family. Here we further assume that the parameter space $\mathcal{H}$ or Θ is open and the statistic set fulfills $span\,\{S_i\} = \mathbb{R}^k$ to ensure the validity of the regularity condition. Any member of exponential family in this dissertation satisfies the said assumptions and is natural for convenience. To further discuss the conjugate concepts, we need a new definition as below.

Definition 7.4.1 (Conjugate Families)**.** *A family of probability measure* $\mathbb{T}$ *of* Θ*,* $\mathfrak{T} = \{\mathbb{T}_\eta \mid \eta \in \mathcal{H}\}$*, is said to be conjugate for the probability measure* $\mathbb{P}_\theta$ *if for every* $\mathbb{T}_X$ *is formed in Equation 7.3, the posterior measure or marginal measure* $\Theta \mid X$ *is also a member of the same set* $\mathfrak{T}$*.*

For exponential family member, there is a simple way to obtain the natural conjugate family of distribution, such as prior distribution in Bayesian analysis. The natural conjugate is important because the marginal distribution of $\Theta \mid X$ or the posterior distribution in Bayesian analysis can be easily identified with a slight change in parameters when comparing to the prior or marginal. We also assume that the space $(\Theta, \mathcal{B}_\Theta)$ is also measurable and well-behaved.

Then, we introduce the family of natural conjugate for $\mathbb{P}_\theta$, $\mathfrak{T}$. Suppose the hyperparameter in $\mathfrak{T}$ is $\eta = (\omega, \lambda) \in \mathcal{H}$. On the space $(\Theta, \mathcal{B}_\Theta)$ we have,

$$\mathbb{T}_{\omega,\lambda}(\theta) = \int \mathbb{1}_A \exp\Big[\langle \omega, \theta \rangle - \lambda\psi(\theta) + E(\omega, \lambda)\Big]\nu(\mathbf{d}\theta), \qquad A \in \mathcal{B}(\Theta) \quad (7.9)$$

where ν is the suitable dominating measure and $E(\cdot)$ is the normalising factor with $\sup E(\cdot) < \infty$ and

$$E(\omega, \lambda) = \log\left\{-\int \exp\Big[\langle \omega, \theta\rangle - \lambda\psi(\theta)\Big]\nu(\mathbf{d}\theta)\right\} \tag{7.10}$$

It is straightforward to observe that $\mathbb{T}_X$ also belongs to the exponential family. With $\mathbb{T}_{\omega,\lambda}$ as prior and the observation follows $\mathbb{P}_\theta$, a member of exponential family, we further have

$$\begin{aligned}(\mathbb{P}_\theta \otimes \mathbb{T}_{\omega,\lambda})(A) &= \iint \mathbb{1}_A(x,\theta)\,\mathbb{P}_\theta(\mathbf{d}x)\mathbb{T}_{\omega,\lambda}(\mathbf{d}\theta), \qquad A \in \mathcal{B}(\mathbb{R}^k) \times \mathcal{B}(\Theta) \\ &= \iint_A \exp\Big[\langle \theta, x+\omega\rangle - (\lambda+1)\psi(\theta) + E(\omega,\lambda)\Big]\mathbf{d}\big(\mu\otimes\nu\big) \\ &= \iint_A \exp\Big[\langle \theta, x+\omega\rangle - \psi_1(\theta)\Big]\mathbf{d}\upsilon\end{aligned} \tag{7.11}$$

where $\psi_1(\cdot) = (\lambda+1)\psi(\cdot)$ and $\upsilon = \exp E(\omega,\lambda)(\mu\otimes\nu)$. It follows that the joint probability measure also belongs to the exponential familty and the joint density of (Θ, X) trivially is the Radon-Nikodym derivative with respect to $\mu\otimes\nu$, the suitable dominating measure with

$$f(x,\theta) = \frac{\mathbf{d}\big[\mathbb{P}_\theta \otimes \mathbb{T}_{\omega,\lambda}\big]}{\mathbf{d}\big[\mu\otimes\nu\big]}(x,\theta) = \exp\big[\langle\theta, x+\omega\rangle - (\lambda+1)\psi(\theta) + E(\omega,\lambda)\big]. \tag{7.12}$$

From Equation 7.3 and Definition 7.4.1, we can find the marginal distribution $\mathbb{T}^X$

as follows:

$$
\begin{aligned}
\mathbb{T}^X(A) &= \int \mathbb{1}_A(\theta) \frac{(\mathbb{P}_\theta \otimes \mathbb{T}_{\omega,\lambda})(x, \mathbf{d}\theta)}{\int (\mathbb{P}_\theta \otimes \mathbb{T}_{\omega,\lambda})(x, \mathbf{d}\theta)}, \qquad A \in \mathcal{B}(\Theta) \\
&= \int \mathbb{1}_A(\theta) \frac{\exp\left[\langle \theta, x+\omega \rangle - (\lambda+1)\psi(\theta) + E(\omega, \lambda)\right]}{\int \exp\left[\langle \theta, x+\omega \rangle - \psi_1(\theta) + E(\omega,\lambda)\right] \mathbf{d}\nu} \nu(\mathbf{d}\theta) \\
&= \int \mathbb{1}_A(\theta) \frac{\exp\left[\langle \theta, x+\omega \rangle - (\lambda+1)\psi(\theta) + E(\omega, \lambda)\right]}{\exp\left[E(\omega,\lambda) - E(x+\omega, \lambda+1)\right] \int \mathbb{T}_{x+\omega,\lambda+1}(\mathbf{d}\theta)} \nu(\mathbf{d}\theta) \\
&= \int \mathbb{1}_A(\theta) \exp\left[\langle \theta, x+\omega \rangle - (\lambda+1)\psi(\theta) + E(\omega + x, \lambda+1)\right] \nu(\mathbf{d}\theta) \\
&= \int \mathbb{1}_A(\theta) \exp\left[\langle \theta, x+\omega \rangle - \psi_1(\theta)\right] \nu_1(\mathbf{d}\theta)
\end{aligned}
\tag{7.13}
$$

where $\psi_1(\cdot) = (\lambda+1)\psi(\cdot)$ and $\nu_1(\cdot) = \exp E(x+\omega, \lambda+1)\nu(\cdot)$ is the dominating measure for $\mathbb{T}^X$. The conditional density of Θ, given $X = x$, denoted as $\pi(\theta \mid x)$ conventionally, with respect to ν_1 is

$$
\pi(\theta \mid x) = \frac{\mathbf{d}\mathbb{T}^X}{\mathbf{d}\nu_1}(\theta) = \exp\left[\langle \theta, x+\omega \rangle - (\lambda+1)\psi(\theta)\right] \tag{7.14}
$$

In decision problems, we have to go further to analyse the Bayesian expected loss or the Frequentist risk, hence we need to introduce the loss function $L(d, \theta)$ for the loss from decision d when the θ is chosen. The best decision is that having the lowest expected loss, either the expectation in Definition 7.2.2 or Definition 7.2.3, where both Θ and $X \mid \Theta$ follows natural exponential family. Therefore, if the loss function is in a form convenient in calculating the expected loss, the analysis will be as nice as the posterior analysis in Bayesian decision theory. As in Section 7.3, unbounded loss does not always make the expected loss exist and finally it is no way to solve the decision theorem in general. Further, some losses from unbounded loss functions can change seriously while there is a slight change in the probability measure. Therefore, the bounded loss with the properties that the loss is bounded below from 0 and bounded above by K are examined here. We propose to use a conjugate loss function in the following form:

$$\begin{aligned} L(d,\theta) &= K\left\{1 - U(d,\theta)\right\}, \qquad K \neq 0 \\ U(d,\theta) &= G(d)\exp\Big[\langle\theta, \tilde{x}(d)\rangle - \tilde{\lambda}(d)\psi(\theta)\Big] \\ G(d)^{-1} &= \int \exp\Big[\langle\theta, \tilde{x}(d)\rangle - \tilde{\lambda}(d)\psi(\theta)\Big]\,\mathbf{d}\theta \end{aligned} \tag{7.15}$$

where $\tilde{x}(\cdot)$, $\tilde{\lambda}(\cdot)$ and $G(\cdot)$ are some appropriate functions of d and θ and $g(\cdot|\tilde{\theta}(\cdot))$ is the density of a member of exponential family obviously. K is the multiplier; without loss of generality, we can set $K = 1$. $G(d)$ is an important normalising factor such that L is bounded within $(0, K)$. $G(\cdot)$ has a property that $0 < G(\cdot) < \infty$. When $d = \theta$, $L(d,\theta) = 0$. There are many special cases with this loss.

Since U is bounded by 1, $L(d,\theta)$ can be simplified as

$$L(d,\theta) = 1 - g_1(\theta \mid \tilde{x}(d)) \tag{7.16}$$

where g_1 is analogous to g, the density of a member of exponential family. Since this loss is formed from the density, it is also called *Inverted Probability Loss Function* in Spiring and Yeung (1998). For some parameter θ, the frequentist risk of a decision d with $\tilde{\theta}(\cdot) = \theta$ by Definition 7.2.2 is

$$\begin{aligned} R^*(d,\theta) &= \int L(d,\theta)\,\mathbb{P}_\theta(\mathbf{d}x) \\ &= \int \Big\{1 - G(d)\exp\Big[\langle\theta, \tilde{x}(d)\rangle - \tilde{\lambda}(d)\psi(\theta)\Big]\Big\}\exp\big[\langle\theta, x\rangle - \psi(\theta)\big]\,\mathbf{d}\mu \\ &= 1 - G(d)\int \exp\Big\{\langle\theta, x + \tilde{x}(d)\rangle - \Big[\tilde{\lambda}(d) + 1\Big]\psi(\theta)\Big\}\,\mathbf{d}\mu \\ &= 1 - G(d)E_1(x + \tilde{x}(d), \tilde{\lambda}(d) + 1) \end{aligned} \tag{7.17}$$

where $E_1(\omega, \tau) = \int \exp\big[\langle \theta, \omega \rangle - \tau\psi(\theta)\big] \mathbf{d}\mu$. but in general it is unknown until we know the particular form of $\tilde{x}(\cdot)$ and $\tilde{\lambda}(\cdot)$.

Except for the frequentist risk, we can also consider the Bayes loss for this loss. For some hyperparameters $(\omega, \lambda) \in \mathcal{H}$, the Bayes loss of a decision d with $\tilde{\theta}(\cdot) = \theta$ by Definition 7.2.3 given $X = x$ is

$$\begin{aligned}
R(d, \theta) &= \int L(d, \theta) \mathbb{T}^X(\mathbf{d}\theta) \\
&= \int \Big\{1 - G(d) \exp\Big[\langle \theta, \tilde{x}(d) \rangle - \tilde{\lambda}(d)\psi(\theta)\Big]\Big\} e^{\langle \theta, x+\omega \rangle - (\lambda+1)\psi(\theta)} \nu_1(\mathbf{d}\theta) \\
&= 1 - \int G(d) e^{\langle \theta, x+\omega+\tilde{x}(d) \rangle - \left[\tilde{\lambda}(d)+\lambda+1\right]\psi(\theta)} \nu_1(\mathbf{d}\theta) \\
&= 1 - \frac{E(x+\omega, \lambda+1)G(d)}{E(\omega + x + \tilde{x}(d), \tilde{\lambda}(d) + \lambda + 1)}
\end{aligned} \tag{7.18}$$

where $E(\cdot)$ is defined in Equation 7.10. $R(\cdot)$ is in terms of all known functions and variables. Notice that the Bayes loss is unchanged if the random variable follows a probability measure belonging to exponential family. Whatever the observation is seen, the prior and the posterior distributions are of the same form, while this conjugate Bayes loss is still of the same form. Therefore, this loss can always be in closed form and form a closed family.

There are another special case that we should examine. Assume $\tilde{x}(\cdot) = x$, a constant function, $L(d, \theta)$ can be simplified as

$$\begin{aligned}
L(d, \theta) &= 1 - G(d) \exp\Big[\langle \theta, x \rangle - \tilde{\lambda}(d)\psi(\theta)\Big] \\
&= 1 - G(d)H(x)g_2(x \mid \theta)
\end{aligned} \tag{7.19}$$

where $g_2(\cdot)$ is analogous to $g(\cdot)$, a member of exponential family, but with the parameter of x instead of θ. $H(\cdot)$ is the appropriate normalising function for the parameter x. For some parameter θ, the frequentist risk of a decision d is

$$
\begin{aligned}
R^*(d,\theta) &= \int L(d,\theta)\,\mathbb{P}_\theta(\mathbf{d}x) \\
&= \int \Big\{1 - G(d)\exp[\langle\theta,x\rangle - \tilde{\lambda}(d)\psi(\theta)]\Big\}\exp\big[\langle\theta,x\rangle - \psi(\theta)\big]\mu(\mathbf{d}x) \\
&= 1 - G(d)
\end{aligned}
\tag{7.20}
$$

For the Bayes loss of some hyperparameters $(\omega,\lambda)\in\mathcal{H}$, the Bayes loss of a decision d with $\tilde{x}(\cdot) = x$ by Definition 7.2.3 given $X = x$ is

$$
\begin{aligned}
R(d,\theta) &= \int L(d,\theta)\mathbb{T}^X(\mathbf{d}\theta) \\
&= \int \Big\{1 - G(d)\exp\Big[\langle\theta,x\rangle - \tilde{\lambda}(d)\psi(\theta)\Big]\Big\}\, e^{\langle\theta,x+\omega\rangle - (\lambda+1)\psi(\theta)}\nu_1(\mathbf{d}\theta) \\
&= 1 - G(d)\int \exp\Big[\langle\theta, 2x+\omega\rangle - \Big(\tilde{\lambda}(d)+\lambda+1\Big)\psi(\theta)\Big]\ \nu_1(\mathbf{d}\theta) \\
&= 1 - \frac{E(x+\omega,\lambda+1)G(d)}{E(\omega+2x,\tilde{\lambda}(d)+\lambda+1)}
\end{aligned}
\tag{7.21}
$$

In general, a unique minimum for the loss is one of the best features that there exists only one unique optimal choice. Certainly, this is controlled by the shape of g_1. If $\varphi = -\log g_1$ is convex, then the density is logconcave and strongly unimodal (Dharmadhikari and Joag-Dev, 1988). Then, the loss has only one unique minimum if the mode of g_1 is at $\theta = d$.

From another point of view, the first order condition has to be satisfied if it is differentiable. We can differentiate and set the equation to be 0 to solve the unknown. We fix d to let the loss as a function of θ, then in order to meet the minimum at $\theta = d$ and we get

$$
\tilde{x}(d) - \tilde{\lambda}(d)\psi'(\theta) = 0 \tag{7.22}
$$

Hence, we can further simplify the loss by eliminating the whole term $\tilde{x}(d) - \tilde{x}(\theta)$ as follows:

$$
\begin{aligned}
L(d,\theta) &= 1 - G(d)\exp\left\{\langle\theta, \tilde{x}(d)\rangle - \tilde{\lambda}(d)\psi(\theta)\right\} \\
&= 1 - G(d)\exp\left\{\left\langle\theta, \tilde{\lambda}(d)\psi'(\theta)\right\rangle - \tilde{\lambda}(d)\psi(\theta)\right\} \\
&= 1 - G(d)\exp\left\{\left\langle\theta[\psi'(\theta)]^\intercal - \psi(\theta), \tilde{\lambda}(d)\right\rangle\right\}
\end{aligned}
\tag{7.23}
$$

By convention, we set $L(d, \theta = d) = 0$ and thus

$$
G(d) = \exp\left[\left\langle\psi(d) - d[\psi'(d)]^\intercal, \tilde{\lambda}(d)\right\rangle\right] \tag{7.24}
$$

From the above argument, this loss is always bounded by 0 and 1 and in general it is smooth and continuous. The advantage over quadratic loss is that this always lead to a finite expected loss with respect to any probability distributions.

In summary, the conjugate loss is of the form

$$
\begin{aligned}
L(d,\theta) =& 1 - G(d)\exp\left[\left\langle\theta\psi'(\theta) - \psi(\theta), \tilde{\lambda}(d)\right\rangle\right] \\
=& 1 - \exp\left[\left\langle\psi(d) - \psi(\theta) + \theta\psi'(d) - d[\psi'(d)]^\intercal, \tilde{\lambda}(d)\right\rangle\right]
\end{aligned}
\tag{7.25}
$$

and $\lambda(d)$ is the only free variable. If $\psi'(d)$ is a scalar, then we can further simplify to

$$
L(d,\theta) = 1 - \exp\left[\left\langle\psi(d) - \psi(\theta) + \psi'(d)(\theta - d), \hat{\lambda}(d)\right\rangle\right] \tag{7.26}
$$

Expanding $\psi(\cdot)$ of θ in Taylor series when d is close to θ when $\psi(\cdot)$ is analytic, we get

$$
L(d,\theta) \approx 1 - \exp\left[\left\langle -\frac{1}{2}\psi''(d)(d-\theta)^2, \tilde{\lambda}(d)\right\rangle\right] \tag{7.27}
$$

That is, for fixed d, $\hat{\lambda}(d)$ is a constant or can be written as a multiple of $\psi''(d)$. In a local sense, when d is very close to θ, any conjugate losses under exponential family we proposed will be approximately equivalent to the Spiring loss described in Chapter 2.

Therefore, when there is no specific condition and considering around a target θ, the best and safest choice should be the Spiring loss, that is, the more general form of Spiring–Yeung Inverted Probability loss framework based on normal distribution.

7.5 Conclusion and Further Remarks

This chapter analyses some conjugate losses in a more rigorous foundations from the conjugate view. However, this chapter has a lot of generalisations which can be researched. How to minimise the loss with more observed data, how to optimise the loss is also a big topic in this kind of losses.

Chapter 8

Further Discussions and Prospects

In this report, we have done a lot of investigations for Inverted Probability loss functions. However, there are still some problems not being solved.

Over the past decades, loss functions are increasingly important in the quality control and statistical usage. In all 7 chapters, we have developed the Spiring–Yeung framework from a just limited example to a wide range of possibilities in both the traditional quadratic type loss and the bounded loss.

However, the shortcoming of IPLF is also based on a density function, but the existence of a density function is not strong enough such that it can include any unbounded loss in general. With the "exponentiated" technique, it is possible to map an unbounded loss into the framework of Spiring–Yeung framework and enjoy a lot of properties that the original unbounded loss has.

Some applications are included for illustrations in Chapter 6. The family of Spiring–Yeung framework provides a wide range of bounded loss functions that can be used in many situations.

Although we have tried to re-study the Spiring–Yeung framework in a rigorous way, it is not very successful as it is due to the time limit. This framework is based on observation that the exponentiated square loss can be written as an normalised inverted Gaussian density. Actually, from the direction of constructing conjugate loss, we would confirm that it is the dual of viewing whether the loss is a function of the variable itself or a target as a variable itself. If both sides can be written as a density, then two views are dual. The major question is that whether the whole framework can be put in rigor and how to put it back in the statistical field. We will study it in future.

Bibliography

Abdolshah, M., Yusuff, R., Hong, T., and Ismail, M. Y. B. (2009). New process capability index using taguchi loss functions. *Journal of Applied Sciences*, 9(20):3775–3779. 98, 99, 103

Bar-Lev, S., Bshouty, D., and Letac, G. (1992). Natural exponential families and self-decomposability. *Statistics & probability letters*, 13(2):147–152. 71

Barker, T. (1990). *Engineering quality by design: interpreting the Taguchi approach*. CRC. 14, 25

Barndorff-Nielsen, O. (1978). *Information and exponential families: in statistical theory*. John Wiley & Sons. 21, 70, 71, 72, 73, 120

Berger, J. (1985). *Statistical decision theory and Bayesian analysis*. Springer, second edition. 1, 65, 66, 98, 103, 104, 105, 112, 113, 116

Bertin, E., Cuculescu, I., and Theodorescu, R. (1997). *Unimodality of probability measures*. Kluwer Academic Publishers. 21, 72, 73

Blackwell, D. and Girshick, M. (1979). *Theory of games and statistical decisions*. Dover Publications. 113

Box, G. E. and Tiao, G. C. (1992). *Bayesian inference in statistical analysis*. Wiley-Interscience. 98

Brown, L. (1986). *Fundamentals of statistical exponential families: with applications in statistical decision theory*. Institute of Mathematical Statistics. 70, 72, 75, 120, 121

Chen, C.-H. and Chou, C.-Y. (2001). Squared error loss of process capability indices. *Economic Quality Control*, 16(2):205–209. 98

DeGroot, M. (2004). *Optimal statistical decisions*, volume 82. Wiley-interscience, paperback edition. Original published in 1959. 115

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, convexity, and applications*. Academic Press. 21, 72, 127

Fathi, Y. and Poonthanomsook, C. (2007). A quartic quality loss function and its properties. *Journal of Industrial and Systems Engineering*, 1(1):8–22. 31, 61

Ferguson, T. (1967). *Mathematical statistics: A decision theoretic approach*, volume 7. Academic Press New York. 113

Freedman, D. and Diaconis, P. (1981). On the histogram as a density estimator: L 2 theory. *Probability theory and related fields*, 57(4):453–476. 5

Freimer, M., Kollia, G., Mudholkar, G. S., and Lin, C. T. (1988). A study of the generalized tukey lambda family. *Communications in Statistics-Theory and Methods*, 17(10):3547–3567. 79, 81

French, S. and Rios, I. (2000). *Statistical decision theory*, volume 9 of *Kendall's library of statistics*. Arnold Publishers, London. 113

Gilchrist, W. (2002). *Statistical modelling with quantile functions*. Chapman and Hall/CRC. 79, 81, 90, 92

Hastings, C., Mosteller, F., Tukey, J. W., and Winsor, C. P. (1947). Low moments for small samples: a comparative study of order statistics. *The Annals of Mathematical Statistics*, 18(3):413–426. 78

Johnson, T. (1992). The relationship of cpm to squared error loss. *Journal of Quality Technology*, 24(4):211–215. 98

Kane, V. E. (1986). Process capability indices. *Journal of Quality Technology*, 18(1):41–52. 98, 99

Karian, Z. A. and Dudewicz, E. J. (2000). *Fitting statistical distributions: The generalized lambda distribution and generalized bootstrap methods*. Chapman & Hall. 79, 84

Karian, Z. A. and Dudewicz, E. J. (2011). *Handbook of fitting statistical distributions with R*. CRC Press Boca Raton Fl. 79, 84

Kolmogorov, A. N. (1933). Foundations of probability theory. *Chelsea, New York*, 270. 1956 republished. 114

Lehmann, E. and Casella, G. (1998). *Theory of point estimation*. Springer Verlag. 70, 104, 120

Lehmann, E. and Romano, J. (2005). *Testing statistical hypotheses*. Springer Verlag. 113, 120

Leonard, T. and Hsu, J. (2001). *Bayesian methods: an analysis for statisticians and interdisciplinary researchers*, volume 5. Cambridge University Press. 103, 113

Leung, B. and Spiring, F. (2002). The inverted beta loss function: properties and applications. *IIE Transactions*, 34(12):1101–1109. 2, 3, 6, 17, 22, 25, 26, 28, 33, 34, 48, 52, 60, 69, 75

Leung, B. and Spiring, F. (2004). Some properties of the family of inverted probability loss functions. *Quality Technology and Quantitative Management*, 1(1):125–147. 2, 3, 33, 51, 60, 69

Leung, B., Spiring, F., Yau, K., and Chan, C.-k. (2012). The relationship of process capability indices to weighted squared error loss. manuscript, 30 pages, submitted for publication. 98

Liese, F. and Miescke, K. (2008). *Statistical decision theory: estimation, testing, and selection*. Springer. 70, 75, 104, 113, 120

Lukacs, E. (1983). *Developments in characteristic function theory*. Charles Griffin and Company Ltd. 71

Pace, L. and Salvan, A. (1997). *Principles of statistical inference: from a neo-Fisherian perspective*, volume 4 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific Pub Co Inc. 113

Pan, J. (2007). A new loss function-based method for evaluating manufacturing and environmental risks. *International Journal of Quality & Reliability Management*, 24(8):861–887. 24, 33

Pan, J. and Li, Y. (2001). A study of loss functions for unilateral product specifications. *Journal of Quality*, 40(1):39–65. 24, 31

Pan, J. and Pan, J. (2006). A comparative study of various loss functions in the economic tolerance design. In *Management of Innovation and Technology, 2006 IEEE International Conference on*, volume 2, pages 783–787. IEEE. 24, 33

Pan, J. and Pan, J. (2009). Optimization of engineering tolerance design using revised loss functions. *Engineering Optimization*, 41(2):99–118. 24, 33

Pan, J. and Wang, J. (2000). A study of loss functions for product interference analysis. *Industrial Engineering Research*, 2(1):80–100. 23, 31

Press, S. (2002). *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*, volume 590. Wiley-Interscience, second edition. 66, 112

Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R., and Mykytka, E. F. (1979). A probability distribution and its uses in fitting data. *Technometrics*, 21(2):201–214. 79

Ramberg, J. S. and Schmeiser, B. W. (1972). An approximate method for generating symmetric random variables. *Communications of the ACM*, 15(11):987–990. 79, 81

Ramberg, J. S. and Schmeiser, B. W. (1974). An approximate method for generating asymmetric random variables. *Communications of the ACM*, 17(2):78–82. 79, 81

Robert, C. (1996). Intrinsic losses. *Theory and decision*, 40(2):191–214. 66

Robert, C. (2001). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer, second edition. 66, 103, 105, 112, 113, 120

Ryan, T. (2012). *Statistical methods for quality improvement*. John Wiley & Sons, third edition. 13

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66(3):605–610. 5

Scott, D. W. (1992). *Multivariate density estimation: theory, practice, and visualization*, volume 383. Wiley. 5

Smith, J. (1980). Bayes estimates under bounded loss. *Biometrika*, 67(3):629–638. 105

Smith, J. (1987). *Decision analysis: a Bayesian approach*. Chapman & Hall, Ltd. 101, 103, 105, 113

Spiring, F. (1993). The reflected normal loss function. *Canadian Journal of Statistics*, 21(3):321–330. 13, 15, 19, 25, 29, 33, 41, 42, 101, 105, 112, 113

Spiring, F. and Leung, B. (2009). Monetizing process capability. In Gragg, L. and Cassell, J., editors, *Progress in Management Engineering*, pages 87–115. Nova Science Pub Inc. 34

Spiring, F. and Yeung, A. (1998). A general class of loss functions with industrial applications. *Journal of Quality Technology*, 30(2):152–162. iii, 1, 6, 8, 19, 20, 22, 23, 25, 32, 33, 42, 51, 55, 69, 78, 80, 81, 92, 93, 94, 102, 103, 113, 125

Sun, F., Laramée, J., and Ramberg, J. (1996). On Spiring's inverted normal loss function. *Canadian Journal of Statistics*, 2:241–249. 13, 17, 19, 22, 41, 44, 47, 48, 69

Taguchi, G. (1986). *Introduction to quality engineering: designing quality into products and processes*. Asian Productivity Organization Tokyo, Japan. 1, 11, 41, 97, 98, 113

Taguchi, G., Elsayed, E., and Hsiang, T. (1989). *Quality engineering in production systems*. McGraw-Hill College. 11, 97

Tribus, M. and Szonyi, G. (1989). An alternative view of the taguchi approach. *Quality progress*, 22(5):46–52. 98

Tukey, J. W. (1960). The practical relationship between the common transformations of percentages of counts of amounts. Technical Report 36, Princeton University, Princeton, NJ. 79, 81

Varian, H. R. (1975). A bayesian approach to real estate assessment. In Feinberg, S. and Zellner, A., editors, *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, pages 195–208. North Holland: Amsterdam. 66

von Neumann, J. and Morgenstern, O. (2004). *Theory of Games and Economic Behavior*. Princeton university press, sixtieth-anniversary commemorative edition. Original published in 1953. 115

Wen, D. and Levy, M. (2001a). Admissibility of Bayes estimates under BLINEX loss for the normal mean problem. *Communications in Statistics-Theory and Methods*, 30(1):155–163. 68

Wen, D. and Levy, M. (2001b). BLINEX: a bounded asymmetric loss function with application to bayesian estimation. *Communications in Statistics-Theory and Methods*, 30(1):147–153. 68, 69

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81(394):446–451. 66, 108