UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Bayesian estimation of Shannon entropy for bivariate beta priors

by

Liesbeth Joanna Sylvia Bodvin

Submitted in partial fulfillment of the requirements for the degree

MSc (Mathematical Statistics)

In the Faculty of Natural and Agricultural Sciences

University of Pretoria

April 2010

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Declaration of Originality

I, Liesbeth Joanna Sylvia Bodvin, declare that this dissertation, which I hereby submit for the degree MSc(Mathematical Statistics) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: ....LJBodvin..........

DATE: April 2010

Acknowledgements

I would like to thank Prof. A. Bekker and Prof. J.J.J. Roux for their help and guidance. This research was partially funded by a scholarship from the National Research Foundation.

To my parents, for all their love.

Summary

Having just survived what is arguably the worst financial crisis in time, it is expected that the focus on regulatory capital held by financial institutions such as banks will increase significantly over the next few years. The probability of default is an important determinant of the amount of regulatory capital to be held, and the accurate calibration of this measure is vital. The purpose of this study is to propose the use of the Shannon entropy when determining the parameters of the prior bivariate beta distribution as part of a Bayesian calibration methodology. Various bivariate beta distributions will be considered as priors to the multinomial distribution associated with rating categories, and the appropriateness of these bivariate beta distributions will be tested on default data. The formulae derived for the Bayesian estimation of Shannon entropy will be used to measure the certainty obtained when selecting the prior parameters.

Keywords: Bayesian estimation, bivariate beta, calibration, credit risk, probability of default, Shannon entropy

Contents

Chapter 1

Introduction

1.1 Overview

Emerging out of arguably the worst financial crisis known to mankind, the emphasis on the amount of regulatory capital held by banks has increased and is expected to continue increasing significantly. Articles such as the one below (London Evening Standard, 2 March 2010) appear almost on a daily basis in business newspapers. It is therefore imperative that the measures used to calculate the capital required are as good as possible.

Banks 'must cut risks'

Nick Goodway

BANKS will have to face up to a future of lower profits but lower risks if regulators are to prevent another global banking crisis, the head of the Financial Services Authority told MPs today.

"Higher capital requirements by regulators will make it more likely that bits of the banking industry are lower-return but lower-risk than in the past," Lord Turner, chairman of the watchdog, said.

He told MPs that he and Paul Volcker, former US Federal Reserve chairman and now President Obama's special adviser, were "in total agreement", adding that he believed the US administration no longer wanted to ban banks entirely from trading on their own account – so-called proprietary trading.

"We are in full agreement on the means, and that capital requirements for trading activities will be the key. I spent 45 minutes on the phone to him three weeks ago, discussing how we can distinguish between day-by-day trading positions used for market making and proprietary trading.

"We are monitoring this very closely. Where banks go beyond market-making to proprietary trading, we would hit them hard with higher capital requirements," he said.

Turner added that banning trading could solve all banking problems, highlighting the case of HBOS, which had to be bailed out and taken over by Lloyds last year. He said: "HBOS was not involved in fancy trading. It was simply overexuberant in its lending to the commercial real-estate sector."

But he also admitted that banks could always find gaps in any new regulations. He said: "There are ways around both legislation and new capital requirements."

The amount of regulatory capital that is held for each customer on a bank's book depends, amongst others, on the customer's credit quality. The Basel II Accord (2006) is a set of regulatory requirements that banks follow in order to ensure that their capital is adequate. This accord prescribes the rules regarding, amongst others, the calculation of the regulatory capital.

One of the components of the regulatory capital requirement computation for a customer is the probability of default of that customer. A default event is defined in the Basel II captial framework (2006) as the occurence of either the bank considering the customer unlikely to repay its debt obligations in full, or the customer being past due more than 90 days on any material credit obligation.

Estimating the probability of default is relatively easy for high-default portfolios such as retail customers or small business banking. In these cases, the samples are usually quite large and logistic regression can be successfully applied to estimate the probability of default.

In small, low-default portfolios, default does not occur frequently (this is preferred economically!) and the statistical estimation of the probability of default becomes significantly more challenging. In practice, credit rating models are built to provide a ranking of customers, and probabilities of default are assigned to the rating categories, based on historically observed default rates.

The industry of credit rating models is very wide. Companies such as Moody's, Standard and Poor's and Fitch specialise in building rating models. Banks use internal rating models in conjunction with external ratings available to determine the customer's credit rating. Internal rating models are usually determined using a statistical scorecard approach, combining quantitative factors such as financial ratios and qualitative factors such as management strength. A myriad of papers are available on this topic, for a good overview refer to the compilation by Ong (2002).

A typical rating scale consists of 21 grades, with 1 (AAA in practice) being the best quality or least risky, and 21 (C in practice) being the worst quality, or most risky. In order to compute the regulatory capital, probabilities of default have to be assigned to each of the rating categories. This is done by calibrating the model using scarcely available historically observed default rates.

The purpose of this study is to propose a Bayesian calibration methodology, using the Shannon entropy as a measure of certainty in the choice of the prior distribution. For simplicity, this model considers only three events, namely (1) default occurring in the investment grade rating class, (2) default occurring in the speculative grade rating class and (3) no default occurring.

It is assumed that these events follow a multinomial distribution, and various bivariate beta distributions will be considered as priors to the multinomial distribution. The appropriateness of the bivariate beta distributions will be tested on Moody's default rate data, and the formulae derived for the Bayesian estimation of Shannon entropy will be used to select the parameters of the prior distributions by measuring the certainty present in the distribution.

The Bayesian estimator of the Shannon entropy will first be derived in Chapter 2 using the well-known bivariate beta type I distribution as prior, followed by the bivariate beta distribution as by defined Connor and Mosimann (1969) in Chapter 3, and the bivariate beta type III distribution (see Ehlers et al., 2009) in Chapter 4. The extended bivariate beta type I distribution (see Ehlers and Bekker, 2010), which allows for positive and negative correlation, is considered in Chapter 5. In Chapter 6, a practical example illustrates the application of these different bivariate beta priors to default rate distributions, followed by some concluding remarks in Chapter 7. Appendix A contains a summary of the notation used in this study, Appendix B contains important definitions and relations, and Appendix C provides all relevant computer programs used in this study.

To the author's knowledge, the Bayesian estimation of the Shannon entropy has only been studied for the Dirichlet type I distribution, which can be reduced to the bivariate beta type I distribution, see Simion (1999). The derivations of the Shannon entropy for the Connor and Mosimann bivariate beta, bivariate beta type III and extended bivariate beta type I distributions are the first of their kind. Also, not much research has been conducted with regards to Bayesian calibration methods for credit risk models, and this study aims to make another step in that direction.

1.2 Shannon Entropy

Claude E. Shannon was a significant contributor to information science, and introduced his entropy measure in 1948 in a "historically significant article" (see Shannon, 1948), according to the Encyclopedia of Statistical Sciences.

Claude E. Shannon: The Pioneer of Entropy

He broke down information technology to the corner stones of 0's and 1's. Apparently von Neumann recommended that Shannon uses the term "entropy", because "nobody knows what entropy really is, so in any discussion you will always have an advantage", see Bishop (2007).

In this section, an overview of the concept of entropy will be given.

1.2.1 Definition

Entropy aims to measure the amount of information, certainty or homogeneity present in a random variable, and plays an important role in surprisingly many fields. A good overview of the statistical properties of the Shannon entropy can be found in the Encyclopedia of Statistical Sciences. For a discrete variable X that can take one of k possible values, each with probability $p_i = P(X_i = x_i) \geq 0$, for $i = 1, 2, ..., k$ and $\sum_{i=1}^{k} p_i = 1$, the Shannon entropy is defined as

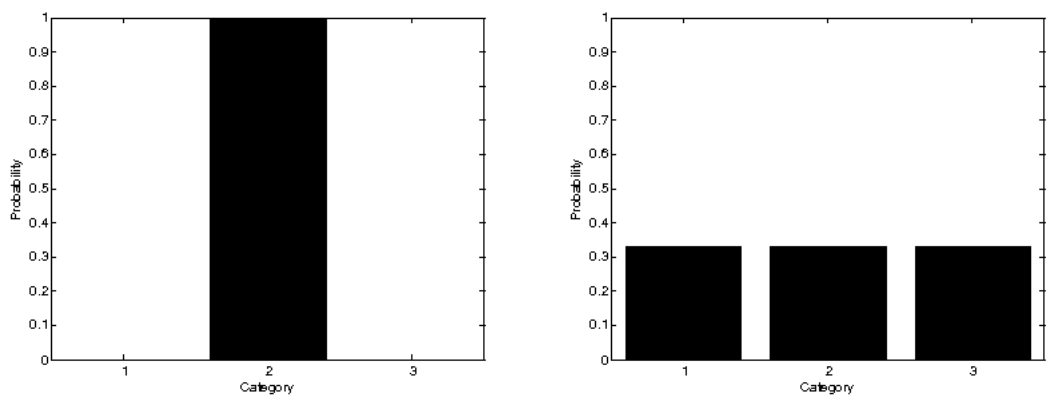
$$\begin{aligned} H_k &= H(p_1, p_2, ..., p_k) \\ &= -\sum_{i=1}^{k} p_i \ln p_i \end{aligned} \qquad (1.1)$$

Entropy measures the extent to which observations are concentrated around a single point, and is a descriptive measure similar to, but not the same as, standard deviation. For the multinomial distribution, entropy is referred to as the heterogeneity present in the variable.

The concept of entropy is illustrated in Figure 1.1 using two extreme cases. Suppose we have a discrete variable X that can take one of three possible values. Figure 1.1 (a) shows the probability of an observation belonging to the second category, and since this probability is 1, there is no uncertainty as to where the next observed value will be. Maximum certainty is attained, which is also referred to as minimum uncertainty or complete homogeneity.

Figure 1.1 (b) shows that the next observed value is equally likely to belong to any of the categories. There is no certainty, also called minimum certainty, maximum uncertainty or complete heterogeneity.

Figure 1.1: Shannon Entropy: A Measure of Heterogeneity



(a) (b)

Generalising, distributions that are concentrated around a few values (i.e. peaked distribution) will have relatively low entropy values (i.e. not much uncertainty), and distributions that are spread around many values will have relatively higher entropy (i.e. more uncertainty), see Bishop (2007).

Since $0 \leq p_i \leq 1$, the entropy is non-negative, and its minimum value of 0 is attained when one of the $p_i = 1$ and all other $p_{j \neq i} = 0$. The maximum entropy is obtained when $p_i = \frac{1}{k}$ for all i, and substituting this into (1.1) shows that the maximum value of Shannon entropy is $\ln k$.

1.2.2 Applications

Entropy is currently used in a variety of fields across theory and practice, with a variety of purposes. Theoretically, Vasicek (1976) used entropy to determine a goodness of fit test for univariate normality, while Ebrahimi et al. (1992) used entropy to determine a goodness of fit test of exponentiality. Zellner (1996) provides an overview of how the maximum entropy method can be used to obtain distributions for random variables.

Essentially, the maximum entropy method is used to determine the distribution of possible values, by optiminsing all the information that is known, but assuming nothing about what is not known. This is similar to a Bayesian estimation of the Shannon entropy, but assuming a non-informative prior (see Section 1.3.2.1). This method is used with large amounts of information, and involves the use of optimisation algorithms, see Berger et al. (1996).

Practically, ecologists measure the diversity of a species of a biological population using entropy, see Pielou (1967). In cryptography, Shannon entropy is used as a cryptographic measure for the key generator module, which forms part of the security of the cipher system, see Simion (2000) and Stephanides (2005). Berger et al. (1996) propose the use of the maximum entropy method to select the correct translated word in natural language processes.

In data mining, the entropy is used to define an error function as part of the learning of weights in multilayer preceptrons in neural networks. The entropy error function is then minimised to determine the optimal weights, see Giudici (2003).

In this study, the Bayesian estimator of the Shannon entropy will be used to measure the level of certainty provided by the parameters of each of the bivariate beta distributions considered as priors (see Section 1.3.2.3). This will be illustrated with a credit risk application.

1.3 Bayesian Estimation

The division between Bayesians and frequentists in statistics is well known, and debates between which approach is best are endless (see O' Hagan and Forster, 2004). Vallverdu (2008) discusses the history and philosophy of Bayesian and frequentist statistical inference in detail; and points out that whilst frequentism is still dominating, more and more Bayesian methods are used for research. Box and Tiao (1992) suggest finding the fine balance between claiming that the Bayesian estimation method can do everything and claiming that it can do nothing, and Berger (2003) proposes to use both methods and compare their possibilities.

Martz and Waller (1982) point out that the frequentist approach might work well for large datasets, but estimates for small datasets may not be reliable or can be heavily influenced by the data set. The quality of parameter estimates can be improved by incorporating historical or expert information using the Bayesian approach.

In this section, the components of Bayesian estimation that are used in this study are briefly discussed.

The fundamental relationship is:

$$f(\mathbf{p}|\mathbf{x}) = \frac{f(\mathbf{x}|\mathbf{p})f(\mathbf{p})}{\int f(\mathbf{x}|\mathbf{p})f(\mathbf{p})d\mathbf{p}} \tag{1.2}$$

where

$\mathbf{p} = (p_1, p_2, ..., p_k)$ is the vector of parameters that has to be estimated,

$\mathbf{x} = (x_1, x_2, ..., x_n)$ is vector of independent observations of the random variable $\mathbf{X}$,

$f(\mathbf{p})$ is the prior distribution of the parameters $\mathbf{P}$, i.e. the information about $\mathbf{P}$ that is available without knowledge of the data,

$f(\mathbf{x}|\mathbf{p})$ is the joint conditional probability distribution of $\mathbf{X}$ given $\mathbf{p}$, regarded in (1.2) as the likelihood function of observing $\mathbf{x}$, given that the parameter values are $\mathbf{p}$, and

$f(\mathbf{p}|\mathbf{x})$ is the posterior distribution of the parameters $\mathbf{P}$, i.e. the information about $\mathbf{P}$ when taking into account the knowledge of the data.

Philosophically, the posterior distribution is interpreted as the distribution of the parameters $\mathbf{P}$ that is obtained after incorporating current as well as previous information.

1.3.1 Likelihood

The likelihood function refers to the distribution of $\mathbf{X}$, and is important since it modifies the prior information about $\mathbf{P}$ when using the data. It is regarded as the function that represents the information about $\mathbf{P}$ contained in the data, see Box and Tiao (1992) and Martz and Waller (1982).

1.3.1.1 This study: multinomial distribution

In this study, the multinomial distribution is considered as the likelihood function, and is defined in Bernardo and Smith (2000). Let:

$\mathbf{X}$ be a discrete random vector variable, whose values can take any of $k+1$ categories,

$x_i = 0, 1, ...$ denote the number of observations of $\mathbf{X}$ in the sample of size $n = 1, 2, ...$ that belong to category i, where $\sum_{i=1}^{k+1} x_i = n$, and

p_i denote the probability of an observation belonging to category i, where $0 < p_i < 1$ and $\sum_{i=1}^{k} p_i < 1$.

Then the mass function of a discrete random variable $\mathbf{X}$ following a multinomial distribution of dimension k with parameters $\mathbf{p} = (p_1, p_2, ..., p_k)$ and n is given by

$$f(\mathbf{x}|\mathbf{p}) = \frac{n!}{\prod_{i=1}^{k} x_i!(n - \sum_{i=1}^{k} x_i)!} \prod_{i=1}^{k} p_i^{x_i} (1 - \sum_{i=1}^{k} p_i)^{n - \sum_{i=1}^{k} x_i}$$

For the bivariate case, this reduces to

$$f(\mathbf{x}|p_1,p_2) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} (1-p_1-p_2)^{n-x_1-x_2} \tag{1.3}$$

where $x_1 + x_2 + x_3 = n$, $0 < p_i < 1$ for $i = 1,2$, and $0 < p_1 + p_2 < 1$.

1.3.2 Prior

The prior is the distribution of the parameters $\mathbf{P}$, without knowledge of the data, and is generally selected subjectively. There are various types of prior distributions, but only the ones necessary for this study will be summarised here.

1.3.2.1 Non-informative prior

If very little or no information is available about a parameter, a non-informative prior can be assigned (see Martz and Waller, 1982, p.223; Bernardo and Smith, 2000, p 357).

1.3.2.2 Expert prior

This type of prior distribution summarises the expert's opinion of the values that a parameter can take, in conjunction with which values are more likely. The expert's estimate of the distribution parameters can be improved as more experience is obtained, or by combining the expert's estimates with estimated parameters as more data becomes available, see Bernardo and Smith (2000), and Martz and Waller (1982).

If there is no data available at all, it is up to the statistician to elicit the required information necessary for the construction of the prior distribution from the expert (who may not necessarily have a statistical background), see O' Hagan and Forster (2004, p159) for some guidelines.

1.3.2.3 Natural conjugate prior

A prior distribution that, when combined with a likelihood function, results in a posterior distribution of the same prior family is called a natural conjugate prior. Raiffa and Schlaifer (1961) suggest to seek a family of priors that have analytic tractability, are flexible and rich and easily interpreted. Conjugate priors are also known as *convenience priors*, due to the mathematical convenience in calculating the posterior distribution. A method to obtain a conjugate prior is described in Martz and Waller (1982).

Similar to the expert prior, the parameters of the natural conjugate prior are obtained based on experience or some initial estimated parameters, or a combination of both, see Press (1989).

In contrast, a non-conjugate prior is obtained when the combination of the likelihood and prior distribution does not result in a member of the same prior family distribution.

1.3.2.4 This study: bivariate beta priors

In this study, the use of various bivariate beta distributions as priors for the multinomial distribution defined in (1.3) is investigated. Both expert and natural conjugate priors will be considered. For illustrative purposes, the parameters of the prior distributions are taken to be the expert opinion; although the parameters can be estimated with the method of moments or maximum likelihood estimation if adequate data is available.

The various bivariate beta priors focused on in the study as well as their relationship with one another are defined in this section. The priors will be discussed in more detail in the following chapters.

Bivariate beta type I

The density function is given by:

$$f(p_1, p_2) = \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1}$$

where $0 < p_i < 1$ for $i = 1, 2$, $0 < p_1 + p_2 < 1$ and $\pi_1, \pi_2, \pi_3 > 0$.

Connor and Mosimann bivariate beta

The density function is given by:

$$f(p_1, p_2) = \frac{\Gamma(\pi_1 + d)\Gamma(\pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1} (1 - p_1)^{d - \pi_2 - \pi_3}$$

where $0 < p_i < 1$ for $i = 1, 2$, $0 < p_1 + p_2 < 1$ and $\pi_1, \pi_2, \pi_3, d > 0$.

Bivariate beta type III

The density function is given by:

$$\begin{aligned} f(p_1, p_2) &= \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{\pi_1 + \pi_2} p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1} \\ &\quad \times [1 + (c - 1)p_1 + (c - 1)p_2]^{-(\pi_1 + \pi_2 + \pi_3)} \end{aligned}$$

where $0 < p_i < 1$ for $i = 1, 2$, $0 < p_1 + p_2 < 1$ and $\pi_1, \pi_2, \pi_3, c > 0$.

Extended bivariate beta type I

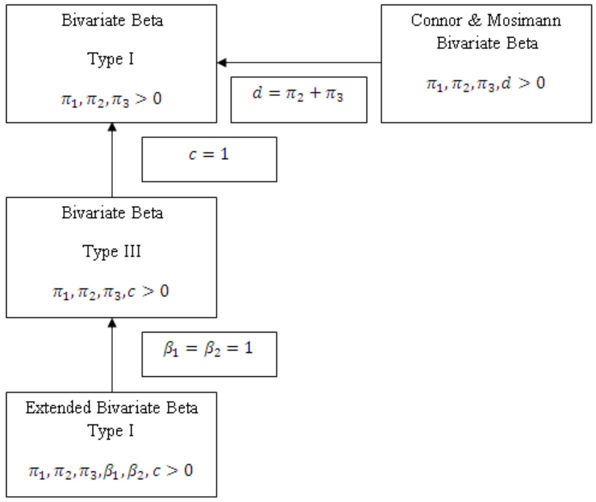
The density function is given by:

$$\begin{aligned} f(p_1, p_2) &= \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} \beta_1^{-\pi_1} \beta_2^{-\pi_2} c^{\pi_1 + \pi_2} p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1} \\ &\quad \times [1 - (1 - \frac{c}{\beta_1})p_1 - (1 - \frac{c}{\beta_2})p_2]^{-(\pi_1 + \pi_2 + \pi_3)} \end{aligned}$$

where $0 < p_i < 1$ for $i = 1, 2$, $0 < p_1 + p_2 < 1$ and $\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c > 0$.

The relationship between these bivariate beta distributions is shown in Figure 1.2.

Figure 1.2: Bivariate Beta Relationships



The bivariate beta distributions considered in this study are an extremely small portion of the universe of bivariate beta distributions. For an overview of other bivariate beta distributions, the reader is referred to Balakrishnan and Lai (2009), Nadarajah and Kotz (2005) and Kotz et al. (2000).

1.3.3 Posterior

The posterior distribution is a summary of the information about $\mathbf{P}$, combining the prior information and the sample data. If the parameters of the posterior distribution are similar to those of the prior distribution, it increases confidence in the subjective selection of the prior distribution.

Bayes' theorem is sequential in nature, therefore making it easy to update the knowledge about $\mathbf{P}$ with new sample or data information as it becomes available, refer to Martz and Waller (1982, p176) for an illustration.

1.3.3.1 This study

This study considers the bivariate case, therefore the posterior distribution in (1.2) simplifies to

$$f(p_1, p_2|\mathbf{x}) = \frac{f(\mathbf{x}|p_1, p_2)f(p_1, p_2)}{\int\int f(\mathbf{x}|p_1, p_2)f(p_1, p_2)dp_1dp_2}$$

where $f(\mathbf{x}|p_1, p_2)$ is the likelihood function of the multinomial distribution, given by (1.3), and $f(p_1, p_2)$ is one of the bivariate beta distributions discussed in Section 1.3.2.3.

It will be shown in the following chapters that if the bivariate beta type I distribution is used as a prior to the multinomial distribution, the posterior distribution is also a bivariate beta type I distribution. Sometimes circumstances require the statistician to select a more complex prior distribution which has the ability to reflect the knowledge of the expert more accurately. This may result in a posterior distribution of a different form than the prior distribution, and whilst this reduces the mathematical convenience it should not be seen as a drawback.

The Connor and Mosimann bivariate beta, bivariate beta type III and extended bivariate beta type I distributions are not natural conjugates to the multinomial distribution.

1.3.4 Loss Function and Estimator

A good overview of loss functions can be found in Berger (1980), Martz and Waller (1982) and Bernardo and Smith (2000). Loss functions find their way from Bayesian decision theory, and aim to measure the loss incurred by estimating a parameter. The smaller the loss, the more accurate the estimator. This study is only concerned with the squared error (or quadratic) loss function.

1.3.4.1 Squared error loss

Suppose we are only estimating a single parameter p. Then the squared error loss function is defined as

$$L(p, \hat{p}) = (p - \hat{p})^2 \tag{1.4}$$

where $\hat{p}$ is the estimate of p.

The squared error loss can be generalised by using weighted squared error loss (see Lee, 2004). Weights are assigned to the loss function, therefore taking into account that the impact of the loss may be more or less severe at certain points. The weighted loss function is given by

$$L(p,\hat{p}) = w(p)(p-\hat{p})^2$$

To extend to the multivariate case, define the squared error loss function as

$$L(\mathbf{p},\hat{\mathbf{p}}) = (\mathbf{p}-\hat{\mathbf{p}})'\mathbf{Q}(\mathbf{p}-\hat{\mathbf{p}}) \tag{1.5}$$

where $\mathbf{p} = (p_1, p_2, ..., p_k)$, $k \geq 2$, is the vector of parameters that has to be estimated,

$\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, ..., \hat{p}_k)$ is the vector of estimates, and

$\mathbf{Q}$ is a $k \times k$ positive definite matrix.

$\mathbf{Q}$ can be seen as the influence that the different loss functions have on one another. If $\mathbf{Q}$ is a diagonal matrix, the loss functions of two parameters are not influenced by each other, and the loss function becomes

$$L(\mathbf{p},\hat{\mathbf{p}}) = \sum_{i=1}^{k} q_i(p_i-\hat{p}_i)^2$$

Weighted and squared error loss functions can be used in the presence of nuisance parameters, which are parameters where some components are known or assumed to have a certain fixed value.

1.3.4.2 Bayes estimator

It is shown in Martz and Waller (1982, p200) and Lee (2004, p205) that under the squared error loss function the Bayes estimator of $\mathbf{P}$ is the expected value with respect to the posterior distribution. That is, when the loss function is specified by (1.4), the Bayes estimator for any specified prior $f(p)$ will be the estimator that minimises the posterior risk given by

$$E[(P-\hat{P})^2|\mathbf{x}] = \int (p-\hat{p})^2 f(p|\mathbf{x})dp$$

and the minimum is attained when

$$\hat{P} = E[P|\mathbf{x}] = \int p f(p|\mathbf{x})dp$$

Under quadratic loss, when the loss function is specified by (1.5), the Bayes estimator is that estimator that minimises the loss, that is, minimises

$$E[(\mathbf{P}-\hat{\mathbf{P}})'\mathbf{Q}(\mathbf{P}-\hat{\mathbf{P}})|\mathbf{x}]$$

and the minimum is attained when

$$\hat{\mathbf{P}} = E[\mathbf{P}|\mathbf{x}] = \int \mathbf{p} f(\mathbf{p}|\mathbf{x})d\mathbf{p}$$

Chapter 2

Bivariate Beta Type I Prior

2.1 The Bivariate Beta Type I prior

In general, the beta type I distribution (be it univariate, bivariate or multivariate) is famous for its ability to model proportions.

2.1.1 Joint Density Function

The conjugate prior for the multinomial distribution is the well-known Dirichlet type I distribution. For the bivariate case, this reduces to the bivariate beta type I distribution, denoted by $BBeta^I(\pi_1, \pi_2, \pi_3)$ and with density function:

$$f(p_1, p_2) = \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1} \tag{2.1}$$

where $0 < p_i < 1$ for $i = 1, 2$, $0 < p_1 + p_2 < 1$ and $\pi_1, \pi_2, \pi_3 > 0$.

2.1.2 Univariate Properties

The marginal density function of P_1 is derived as:

$$\begin{aligned} f(p_1) &= \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} p_1^{\pi_1-1} \int_0^{1-p_1} p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1} dp_2 \\ &= \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} p_1^{\pi_1-1}(1-p_1)^{\pi_2+\pi_3-1} \frac{\Gamma(\pi_2)\Gamma(\pi_3)}{\Gamma(\pi_2+\pi_3)} \end{aligned}$$

using equation (B.1) found in Gradshteyn and Ryzhik (2007). It then follows that:

$$f(p_1) = \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2+\pi_3)} p_1^{\pi_1-1}(1-p_1)^{\pi_2+\pi_3-1}$$

where $0 < p_1 < 1$, and $\pi_1, \pi_2, \pi_3 > 0$, which is the density function of a univariate $Beta(\pi_1, \pi_2 + \pi_3)$ distribution. Similarly, the marginal distribution of P_2 is a univariate $Beta(\pi_2, \pi_1 + \pi_3)$ distribution.

2.1.3 Methods of Derivation

The bivariate beta type I distribution is derived from three independently distributed χ^2 variables. Balakrishnan and Lai (2009, p174) refer to this as the trivariate reduction method.

Let $S_i \sim \chi^2(2\pi_i)$ for $i = 1, 2, 3$ be three independently distributed χ^2 variables. That is,

$$f(s_i) = \frac{1}{2^{\pi_i}\Gamma(\pi_i)} \exp(-\frac{s_i}{2}) s_i^{\pi_i-1}$$

for $s_i > 0$. The joint density function of these variables is given by

$$\begin{aligned} f(s_1, s_2, s_3) &= \prod_{i=1}^{3} f(s_i) \\ &= \frac{1}{2^{\sum_{i=1}^{3}\pi_i} \prod_{i=1}^{3}\Gamma(\pi_i)} \exp(-\frac{1}{2}\sum_{i=1}^{3} s_i) \prod_{i=1}^{3} s_i^{\pi_i-1} \end{aligned}$$

Consider the transformation $P_i = \frac{S_i}{S_1+S_2+S_3}$ for $i = 1, 2$ and $P_3 = S_1 + S_2 + S_3$. Then

$$\begin{aligned} S_1 &= P_1P_3 \\ S_2 &= P_2P_3 \\ S_3 &= P_3(1 - P_1 - P_2) \end{aligned}$$

The Jacobian for this transformation is given by

$$\begin{aligned} J &= J((s_1, s_2, s_3) \rightarrow (p_1, p_2, p_3)) \\ &= \begin{vmatrix} \frac{\partial s_1}{\partial p_1} & \frac{\partial s_1}{\partial p_2} & \frac{\partial s_1}{\partial p_3} \\ \frac{\partial s_2}{\partial p_1} & \frac{\partial s_2}{\partial p_2} & \frac{\partial s_2}{\partial p_3} \\ \frac{\partial s_3}{\partial p_1} & \frac{\partial s_3}{\partial p_2} & \frac{\partial s_3}{\partial p_3} \end{vmatrix} \\ &= \begin{vmatrix} p_3 & 0 & p_1 \\ 0 & p_3 & p_2 \\ -p_3 & -p_3 & 1 - p_1 - p_2 \end{vmatrix} \\ &= p_3[p_3(1 - p_1 - p_2) + p_2p_3] + p_1p_3^2 \\ &= p_3^2 \end{aligned}$$

The joint density function of P_1, P_2 and P_3 is given by

$$\begin{aligned} f(p_1, p_2, p_3) &= f(s_1, s_2, s_3)|J| \\ &= \frac{1}{2^{\sum_{i=1}^3 \pi_i} \prod_{i=1}^3 \Gamma(\pi_i)} (p_1p_3)^{\pi_1 - 1}(p_2p_3)^{\pi_2 - 1}[p_3(1 - p_1 - p_2)]^{\pi_3 - 1} \\ &\quad \times \exp(-\frac{1}{2}[p_1p_3 + p_2p_3 + p_3(1 - p_1 - p_2)]) \\ &= \frac{1}{2^{\sum_{i=1}^3 \pi_i} \prod_{i=1}^3 \Gamma(\pi_i)} p_1^{\pi_1 - 1}p_2^{\pi_2 - 1}(1 - p_1 - p_2)^{\pi_3 - 1}p_3^{\sum_{i=1}^3 \pi_i - 1} \exp(-\frac{p_3}{2}) \end{aligned}$$

The joint density function of P_1 and P_2 is given by

$$\begin{aligned} f(p_1,p_2) &= \int_0^\infty f(p_1,p_2,p_3)dp_3 \\ &= \frac{1}{2^{\sum_{i=1}^3 \pi_i}\prod_{i=1}^3 \Gamma(\pi_i)} p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}\int_0^\infty p_3^{\sum_{i=1}^3 \pi_i-1}\exp(-\frac{p_3}{2})dp_3 \end{aligned}$$

The integral above can be written as

$$\begin{aligned} & \int_0^\infty p_3^{\sum_{i=1}^3 \pi_i-1}\exp(-\frac{p_3}{2})dp_3 \\ =\ & 2^{\sum_{i=1}^3 \pi_i}\Gamma(\sum_{i=1}^3 \pi_i)\int_0^\infty \frac{1}{2^{\sum_{i=1}^3 \pi_i}\Gamma(\sum_{i=1}^3 \pi_i)} p_3^{\sum_{i=1}^3 \pi_i-1}\exp(-\frac{p_3}{2})dp_3 \qquad (2.2) \\ =\ & 2^{\sum_{i=1}^3 \pi_i}\Gamma(\sum_{i=1}^3 \pi_i) \end{aligned}$$

where the integral in (2.2) is 1 since it is the total probability of a $\chi^2(\sum_{i=1}^3 \pi_i)$ distribution. Then

$$f(p_1,p_2) = \frac{\Gamma(\sum_{i=1}^3 \pi_i)}{\prod_{i=1}^3 \Gamma(\pi_i)} p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}$$

where $0 < p_i < 1$, $i = 1, 2$ and $0 < p_1 + p_2 < 1$, and it follows that P_1 and P_2 are jointly distributed as $BBeta^I(\pi_1, \pi_2, \pi_3)$.

2.1.4 Correlation

The product moments are derived as

$$\begin{aligned} E[P_1^i P_2^j] &= \int_0^1\int_0^{1-p_2} p_1^i p_2^j f(p_1,p_2)dp_1dp_2 \\ &= \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)}\int_0^1\int_0^{1-p_2} p_1^{\pi_1+i-1}p_2^{\pi_2+j-1}(1-p_1-p_2)^{\pi_3-1}dp_1dp_2 \end{aligned}$$

Note that the integral above is proportional to a $BBeta^I(\pi_1 + i, \pi_2 + j, \pi_3)$ distribution (see

(2.1)). Then

$$
\begin{aligned}
E[P_1^i P_2^j] &= \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} \frac{\Gamma(\pi_1+i)\Gamma(\pi_2+j)\Gamma(\pi_3)}{\Gamma(\pi_1+i+\pi_2+j+\pi_3)} \\
&= \frac{\Gamma(\pi_1+i)\Gamma(\pi_2+j)}{\Gamma(\pi_1)\Gamma(\pi_2)} \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j)} \qquad (2.3)
\end{aligned}
$$

It follows from (2.3) that

$$E[P_i] = \frac{\pi_i}{\pi_1+\pi_2+\pi_3}$$

and

$$E[P_i^2] = \frac{\pi_i(\pi_i+1)}{(\pi_1+\pi_2+\pi_3)(\pi_1+\pi_2+\pi_3+1)}$$

for $i = 1, 2$. The variance of P_1 and P_2 are then respectively given by

$$var(P_1) = \frac{\pi_1(\pi_2+\pi_3)}{(\pi_1+\pi_2+\pi_3)^2(\pi_1+\pi_2+\pi_3+1)}$$

and

$$var(P_2) = \frac{\pi_2(\pi_1+\pi_3)}{(\pi_1+\pi_2+\pi_3)^2(\pi_1+\pi_2+\pi_3+1)}$$

The covariance and correlation are respectively given by

$$cov(P_1, P_2) = \frac{-\pi_1\pi_2}{(\pi_1+\pi_2+\pi_3)^2(\pi_1+\pi_2+\pi_3+1)}$$

and

$$corr(P_1, P_2) = -\sqrt{\frac{\pi_1\pi_2}{(\pi_1+\pi_3)(\pi_2+\pi_3)}} \qquad (2.4)$$

Since $\pi_1, \pi_2, \pi_3 > 0$, the correlation between P_1 and P_2 will always be negative. Balakrishnan and Lai (2009) comment on the unusualness of the negative correlation, and suggest negating one of the variables in order to attain positive correlation. One advantage of the bivariate beta type I distribution is that for variables that are restricted to negative correlation, good analytical tractability simplifies Bayesian analyses. Table 2.1 lists the correlation for some values of π_1, π_2 and π_3.

Table 2.1: Bivariate Beta Type I Distribution: Correlation

π_1	π_2	π_3	Correlation
2	2	2	-0.5
1	2	2	-0.408
10	2	2	-0.645
2	1	2	-0.408
2	10	2	-0.646
2	2	1	-0.667
2	2	10	-0.167

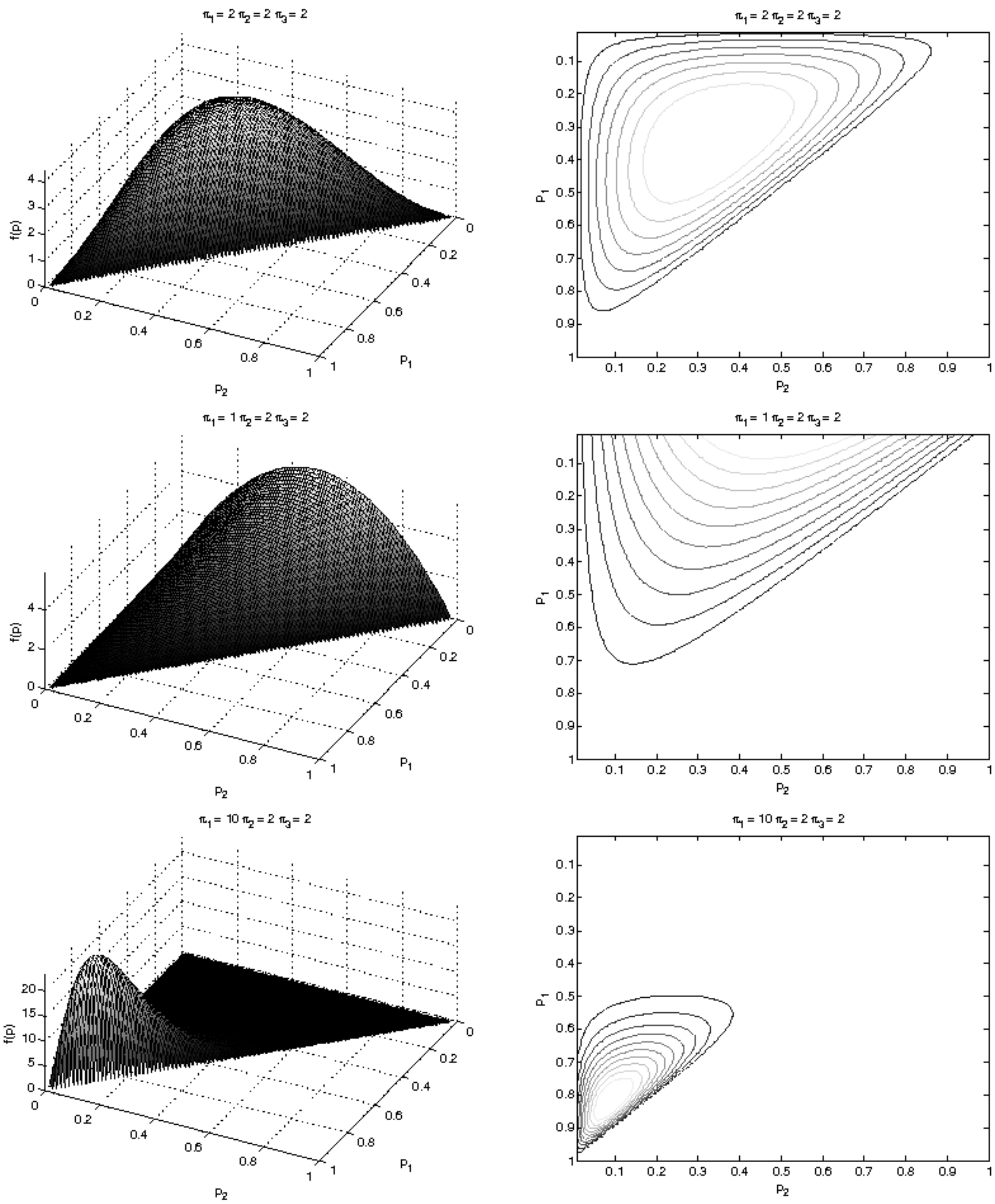
Since the bivariate beta type I distribution is symmetric in P_1 and P_2, the correlation between interchanged parameter values for π_1 and π_2 is also the same.

2.1.5 Shape Analysis

The parameters of the prior distribution can be estimated from historical data using maximum likelihood estimation, or determined from prior knowledge or expert judgement, see O' Hagan and Forster (2004). The examples that follow will show the effect of π_1, π_2 and π_3 on the shape and concentration of the distribution. A reference case will be used, where $\pi_1 = \pi_2 = \pi_3 = 2$.

In Figure 2.1 it can be seen that if π_1 is decreased and π_2 and π_3 remain constant, the distribution shifts towards the marginal distribution of P_2 on the right axis. If π_1 is increased and π_2 and π_3 remain constant, the distribution shifts towards larger values of P_1 and smaller values of P_2 along the line $p_1 + p_2 = 1$.

Figure 2.1: Bivariate Beta Type I Distribution: Changing π_1



Doing the same for π_2, it can be seen in Figure 2.2 that if π_2 is decreased and π_1 and π_3 remain constant, the distribution shifts towards the marginal distribution of P_1 on the left axis. If π_2 is increased, the distribution shifts towards smaller values of P_1 and larger values of P_2 along the line $p_1 + p_2 = 1$.

Figure 2.2: Bivariate Beta Type I Distribution: Changing π_2

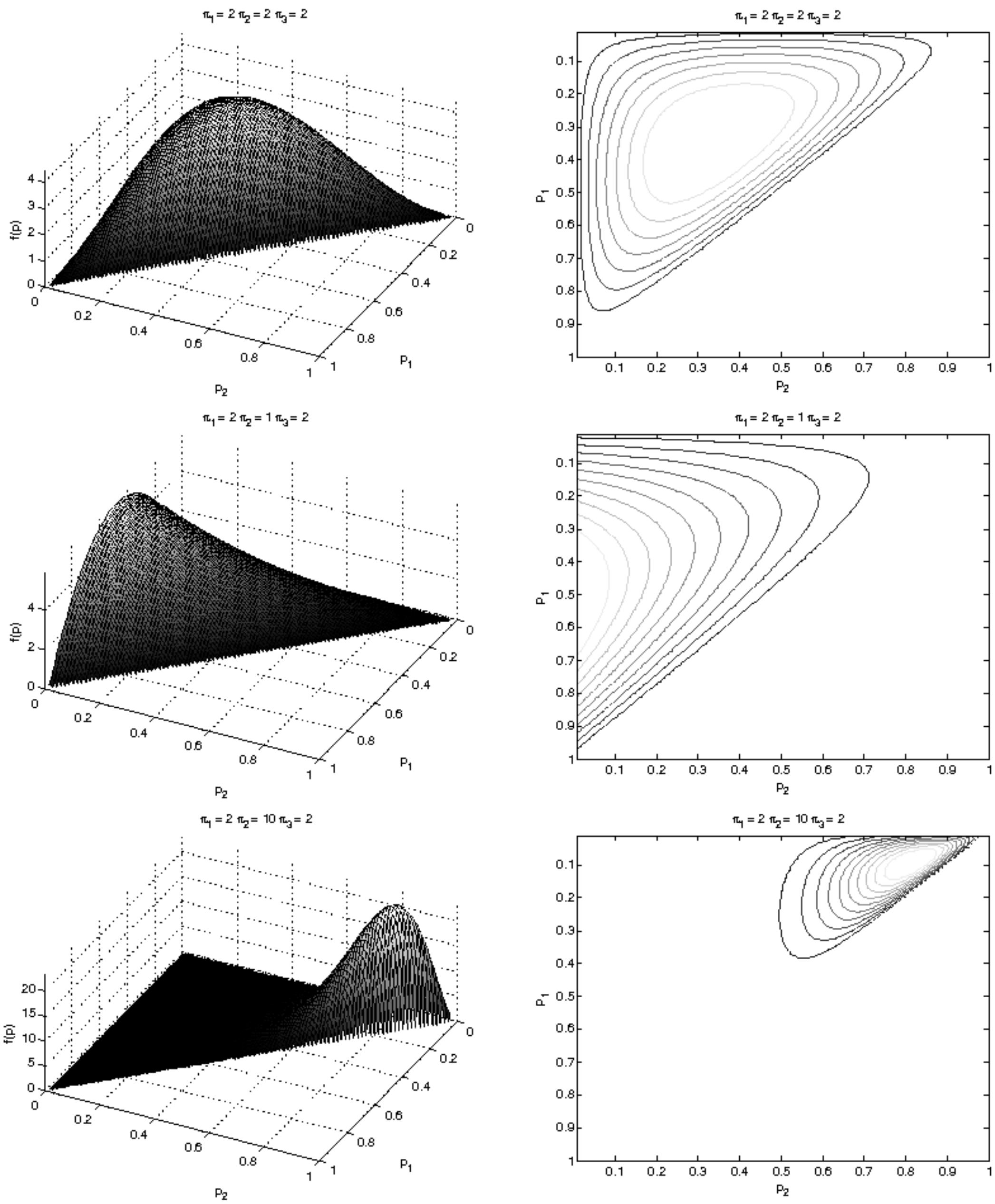
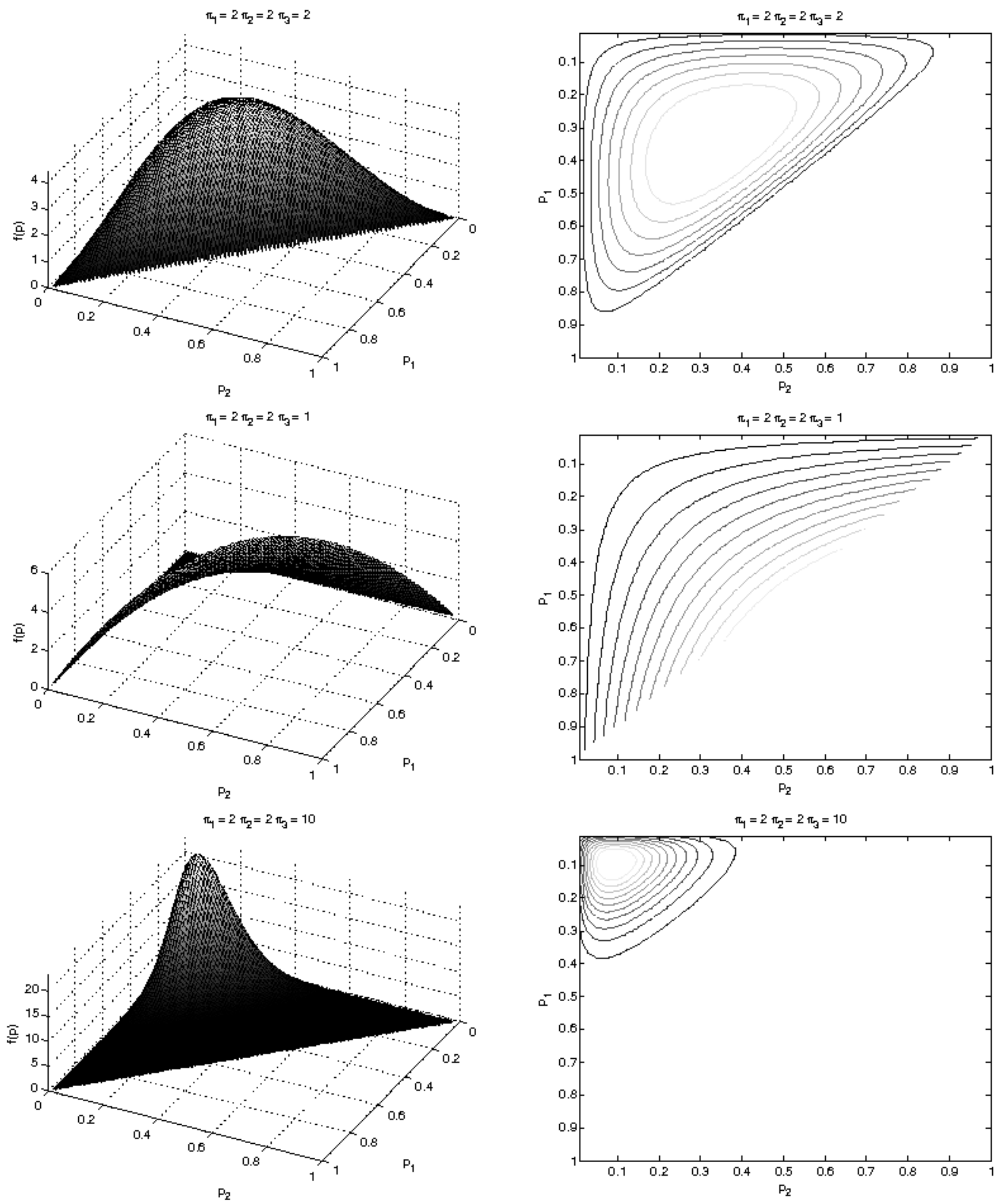


Figure 2.3 shows that if π_3 is decreased, the concentration shifts towards the line $p_1 + p_2 = 1$. If π_3 is increased, the concentration shifts towards small values of P_1 and P_2. This is particularly useful in practice if one deals with two variables P_1 and P_2 which are concentrated towards the lower values of both.

Figure 2.3: Bivariate Beta Type I Distribution: Changing π_3



2.2 Bayesian Estimation of Shannon Entropy

2.2.1 Derivation

Theorem 2.1

The posterior distribution for the multinomial likelihood in (1.3) *and bivariate beta type I prior distribution in* (2.1) *follows a* $BBeta^I(\pi_1 + x_1, \pi_2 + x_2, \pi_3 + x_3)$ *distribution, that is*

$$f(p_1, p_2|\mathbf{x}) = [B(\pi_1 + x_1, \pi_2 + x_2, \pi_3 + x_3)]^{-1} p_1^{\pi_1+x_1-1} p_2^{\pi_2+x_2-1} (1 - p_1 - p_2)^{\pi_3+x_3-1} \quad (2.5)$$

where $0 < p_i < 1$ *for* $i = 1, 2$, $0 < p_1 + p_2 < 1$, $\pi_1, \pi_2, \pi_3 > 0$ *and* $B(\cdot)$ *is the beta function (see Appendix B, Definition 1).*

Proof

Using Bayes' theorem, the posterior distribution is defined as

$$f(p_1, p_2|\mathbf{x}) = \frac{f(\mathbf{x}|p_1, p_2) f(p_1, p_2)}{\int\int f(\mathbf{x}|p_1, p_2) f(p_1, p_2) dp_1 dp_2}$$

where $f(\mathbf{x}|p_1, p_2)$ is the likelihood function of the multinomial distribution, given by (1.3), and $f(p_1, p_2)$ is the bivariate beta type I prior distribution, given by (2.1). The numerator of the posterior distribution is given by

$$f(\mathbf{x}|p_1, p_2) f(p_1, p_2) = [B(\pi_1, \pi_2, \pi_3)]^{-1} \frac{n!}{x_1! x_2! x_3!} p_1^{\pi_1+x_1-1} p_2^{\pi_2+x_2-1} (1 - p_1 - p_2)^{\pi_3+x_3-1} \quad (2.6)$$

The denominator of the posterior distribution is given by

$$\begin{aligned}
& \int_0^1 \int_0^{1-p_2} f(\mathbf{x}|p_1,p_2) f(p_1,p_2) dp_1 dp_2 \\
= \quad & \frac{n!}{x_1!x_2!x_3!} \frac{B(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)}{B(\pi_1,\pi_2,\pi_3)} \\
& \times \int_0^1 \int_0^{1-p_2} [B(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)]^{-1} \\
& \times p_1^{\pi_1+x_1-1} p_2^{\pi_2+x_2-1} (1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \qquad (2.7)
\end{aligned}$$

Note that the integral on the right hand side of (2.7) is equal to 1, since it corresponds to the total probability of a $BBeta^I(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)$ distribution. The denominator of the posterior distribution is therefore a constant, given by

$$\int_0^1 \int_0^{1-p_2} f(\mathbf{x}|p_1,p_2) f(p_1,p_2) dp_1 dp_2 = \frac{n!}{x_1!x_2!x_3!} \frac{B(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)}{B(\pi_1,\pi_2,\pi_3)} \qquad (2.8)$$

Combining the numerator (2.6) and denominator (2.8), the posterior distribution is given by

$$f(p_1,p_2|\mathbf{x}) = [B(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)]^{-1} p_1^{\pi_1+x_1-1} p_2^{\pi_2+x_2-1} (1-p_1-p_2)^{\pi_3+x_3-1}$$

where $0 < p_i < 1$ for $i = 1, 2$, $0 < p_1 + p_2 < 1$ and $\pi_1, \pi_2, \pi_3 > 0$; which is the density function of a $BBeta^I(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)$ distribution. ■

Theorem 2.2

The Bayesian estimator of the Shannon entropy under the squared error loss function using the bivariate beta type I distribution as a prior for the multinomial distribution in (1.3) *is given by:*

$$\hat{H}_3^I = -\sum_{i=1}^{3} \frac{\beta_i}{\sum_{j=1}^{3}\beta_j}(\psi(\beta_i + 1) - \psi(\sum_{j=1}^{3}\beta_j + 1)) \tag{2.9}$$

where

$$\begin{aligned}\beta_1 &= \pi_1 + x_1 \\ \beta_2 &= \pi_2 + x_2 \\ \beta_3 &= \pi_3 + x_3\end{aligned}$$

are the parameters of the posterior distribution in (2.5).

Proof

From (1.1) the Shannon entropy for the bivariate beta type I distribution is denoted by

$$H_3^I = -\sum_{i=1}^{3} p_i \ln p_i$$

The Bayesian estimator of the Shannon entropy under squared error loss is defined as the expected value of the Shannon entropy with respect to the posterior distribution (see Section 1.3.4.2 of

Chapter 1). That is,

$$
\begin{aligned}
\hat{H}_3^I &= E_{f(p_1,p_2|\mathbf{x})}[H_3^I] \\
&= -\int_0^1\int_0^{1-p_2}\sum_{i=1}^{3} p_i(\ln p_i)[B(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)]^{-1} \\
&\quad \times p_1^{\pi_1+x_1-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \\
&= -[B(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)]^{-1} \\
&\quad \times \sum_{i=1}^{3}\int_0^1\int_0^{1-p_2} p_i(\ln p_i)p_1^{\pi_1+x_1-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \\
&= K\sum_{i=1}^{3} I_i
\end{aligned}
$$

where

$K = -[B(\pi_1+x_1,\pi_2+x_2,\pi_3+x_3)]^{-1}$, and

$I_i = \int_0^1\int_0^{1-p_2} p_i(\ln p_i)p_1^{\pi_1+x_1-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2.$

The simplification of I_i will only be shown for I_1 , but follows similarly for I_2 and I_3.

$$
\begin{aligned}
I_1 &= \int_0^1\int_0^{1-p_2} (\ln p_1)p_1^{\pi_1+x_1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \\
&= \int_0^1\int_0^{1-p_2} (\frac{\partial}{\partial\pi_1}p_1^{\pi_1+x_1})p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2
\end{aligned}
$$

since $\frac{d}{dx}\ln a^x = a^x \ln a$. Changing the order of integration and differentiation:

$$
\begin{aligned}
I_1 &= \frac{\partial}{\partial\pi_1}\int_0^1\int_0^{1-p_2} p_1^{\pi_1+x_1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \\
&= \frac{\partial}{\partial\pi_1}B(\pi_1+x_1+1,\pi_2+x_2,\pi_3+x_3) \\
&\quad \times\int_0^1\int_0^{1-p_1} [B(\pi_1+x_1+1,\pi_2+x_2,\pi_3+x_3)]^{-1}p_1^{\pi_1+x_1}p_2^{\pi_2+x_2-1} \\
&\quad \times(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \qquad (2.10) \\
&= \frac{\partial}{\partial\pi_1}B(\pi_1+x_1+1,\pi_2+x_2,\pi_3+x_3)
\end{aligned}
$$

The integral in (2.10) is 1 since it is the density function of a $BBeta^I(\pi_1+x_1+1, \pi_2+x_2, \pi_3+x_3)$ distribution. Using the product and chain rules for differentiation:

$$\begin{aligned} & I_1 \\ = \quad & \frac{\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{[\Gamma(\pi_1+x_1+1+\pi_2+x_2+\pi_3+x_3)]^2}[\Gamma'(\pi_1+x_1+1)\Gamma(\pi_1+x_1+1+\pi_2+x_2+\pi_3+x_3) \\ & -\Gamma(\pi_1+x_1+1)\Gamma'(\pi_1+x_1+1+\pi_2+x_2+\pi_3+x_3)] \end{aligned}$$

from which it follows that

$$\begin{aligned} KI_1 &= -\frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)[\Gamma(\pi_1+x_1+1+\pi_2+x_2+\pi_3+x_3)]^2} \\ &\quad \times[\Gamma'(\pi_1+x_1+1)\Gamma(\pi_1+x_1+1+\pi_2+x_2+\pi_3+x_3) \\ &\quad -\Gamma(\pi_1+x_1)\Gamma'(\pi_1+x_1+1+\pi_2+x_2+\pi_3+x_3] \\ &= -\frac{\pi_1+x_1}{\pi_1+x_1+\pi_2+x_2+\pi_3+x_3}[\psi(\pi_1+x_1+1)-\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+1)] \end{aligned}$$

where $\psi(x)$ denotes the polygamma function (see Appendix B, Definition 2). Similarly,

$$KI_2 = -\frac{\pi_2+x_2}{\pi_1+x_1+\pi_2+x_2+\pi_3+x_3}[\psi(\pi_2+x_2+1)-\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+1)]$$

and

$$KI_3 = -\frac{\pi_3+x_3}{\pi_1+x_1+\pi_2+x_2+\pi_3+x_3}[\psi(\pi_3+x_3+1)-\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+1)]$$

For simplification, let

$$\begin{aligned} \beta_1 &= \pi_1+x_1 \\ \beta_2 &= \pi_2+x_2 \\ \beta_3 &= \pi_3+x_3 \end{aligned}$$

The Bayesian estimator of Shannon entropy is then given by

$$\hat{H}_3^I = -\sum_{i=1}^{3} \frac{\beta_i}{\sum_{j=1}^{3} \beta_j} (\psi(\beta_i + 1) - \psi(\sum_{j=1}^{3} \beta_j + 1))$$

which proves this theorem. ■

A generalization of this result can be found in Simion (1999).

2.2.2 Numerical Analysis

Table 2.2 below summarises the Bayesian estimates of Shannon entropy obtained when using (2.9), for each of the parameter combinations used in the shape analysis. The multinomial frequencies were assumed to be $x_1 = 1$, $x_2 = 2$, $x_3 = 10$.

Table 2.2: Bayesian Estimates of Shannon Entropy: Bivariate Beta Type I Prior

π_1	π_2	π_3	$\hat{H}_3^I$
2	2	2	0.860
1	2	2	0.797
10	2	2	0.973
2	1	2	0.815
2	10	2	0.929
2	2	1	0.881
2	2	10	0.714

If π_1 is decreased, a lower $\hat{H}_3^I$ is obtained, indicating less uncertainty. If π_1 is increased, a higher $\hat{H}_3^I$ is obtained, indicating more uncertainty. This is somewhat counterintuitive with the results observed in the shape analysis in Figure 2.1, where a lower value of π_1 is associated with less concentration, indicating more uncertainty; and a larger value of π_1 is associated with more concentration, indicating less uncertainty.

Similarly, if π_2 is decreased, a lower $\hat{H}_3^I$ is obtained, indicating less uncertainty; and if π_2 is increased, a higher $\hat{H}_3^I$ is obtained, indicating more uncertainty. This is also counterintuitive with the results from the shape analysis in Figure 2.2, where a lower value of π_2 is associated with less concentration, indicating more uncertainty; and a larger value of π_2 is associated with more concentration, indicating less uncertainty.

Now, if π_3 is decreased, a higher $\hat{H}_3^I$ is obtained, indicating more uncertainty; and if π_3 is increased, a lower $\hat{H}_3^I$ is obtained, indicating less uncertainty. These results are consistent with the results from the shape analysis in Figure 2.3, were a lower value of π_3 is associated with less concentration, indicating more uncertainty; and a larger value of π_3 is associated with more concentration, indicating less uncertainty.

The difference between these intuitive and counterintuitive results for parameters π_1 and π_2 may be disturbing at first, but can be explained by the location of the distribution. In the first two cases, changing π_1 or π_2 leads to distributions that are either tending to the marginal distribution of P_2 or P_1 respectively, or are tending towards a specific point along the line $p_1 + p_2 = 1$. As the concentration in the distribution remains closer to small values of P_1 and P_2, $\hat{H}_3^I$ stays lower, but as soon as the concentration moves away from these small values to some point along the line $p_1 + p_2 = 1$ the uncertainty increases. This suggests that the Bayesian estimate of Shannon entropy may contain information of not only concentration, but of location as well.

It should also be kept in mind that since $\hat{H}_3^I$ is the *Bayesian* estimate of the Shannon entropy, it also contains information about the likelihood function. Therefore, the multinomial frequencies will also have an effect on the behaviour of $\hat{H}_3^I$. In this example, the multinomial frequencies were assumed to be $x_1 = 1, x_2 = 2$ and $x_3 = 10$. If the multinomial model was symmetric, for example $x_1 = x_2 = 2$, the values for $\hat{H}_3^I$ would have been the same for the case where π_1 or π_2 were changed, since the posterior parameters would have been the same.

Chapter 3

Connor and Mosimann Bivariate Beta Prior

3.1 The Connor and Mosimann Bivariate Beta Prior

The Connor and Mosimann distribution stems from the concept of "neutrality", where for some practical reason a single proportion out of a set of proportions is eliminated. The concept of neutrality is often used in biological data, for example to test the effect of fluoridation in the chemical compostion of bones in rats, see Connor and Mosimann (1969) and Lochner (1975).

3.1.1 Joint Density Function

Consider as a prior to the multinomial distribution in (1.3) the bivariate beta distribution as defined by Connor and Mosimann (1969), denoted by $BBeta^{CM}(\pi_1, \pi_2, \pi_3, d)$ and with density function:

$$f(p_1, p_2) = \frac{\Gamma(\pi_1 + d)\Gamma(\pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1} (1 - p_1)^{d - \pi_2 - \pi_3} \tag{3.1}$$

where $0 < p_1 + p_2 < 1$ and $\pi_1, \pi_2, \pi_3, d > 0$. This prior distribution is not a natural conjugate for the multinomial distribution in (1.3). Note that if $d = \pi_2 + \pi_3$, (3.1) reduces to the bivariate beta type I distribution in (2.1).

3.1.2 Univariate Properties

Due to the additional factor $(1-p_1)^{d-\pi_2-\pi_3}$ in the kernel of (3.1) the two marginal distributions will not be of the same form. The marginal distribution of P_1 is derived as

$$\begin{aligned} f(p_1) &= \frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} p_1^{\pi_1-1}(1-p_1)^{d-\pi_2-\pi_3} \int_0^{1-p_1} p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}dp_2 \\ &= \frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} p_1^{\pi_1-1}(1-p_1)^{d-\pi_2-\pi_3}(1-p_1)^{\pi_2+\pi_3-1}\frac{\Gamma(\pi_2)\Gamma(\pi_3)}{\Gamma(\pi_2+\pi_3)} \end{aligned}$$

using equation (B.1). This simplifies to

$$f(p_1) = \frac{\Gamma(\pi_1+d)}{\Gamma(\pi_1)\Gamma(d)} p_1^{\pi_1-1}(1-p_1)^{d-1}$$

for $0 < p_1 < 1$ and $\pi_1, d > 0$, which is the density of an univariate $Beta(\pi_1, d)$ distribution.

The marginal distribution of P_2 is derived as

$$\begin{aligned} f(p_2) &= \frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} p_2^{\pi_2-1} \int_0^{1-p_2} p_1^{\pi_1-1}(1-p_1-p_2)^{\pi_3-1}(1-p_1)^{d-\pi_2-\pi_3}dp_1 \\ &= \frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} p_2^{\pi_2-1}(1-p_2)^{\pi_1+\pi_3-1}\frac{\Gamma(\pi_1)\Gamma(\pi_3)}{\Gamma(\pi_1+\pi_3)} \\ &\quad \times {}_2F_1(\pi_1, \pi_2+\pi_3-d-1; \pi_1+\pi_3; 1-p_2) \end{aligned}$$

using equation (B.2). This simplifies to

$$f(p_2) = \frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_2)\Gamma(\pi_1+\pi_3)\Gamma(d)} {}_2F_1(\pi_1, \pi_2+\pi_3-d-1; \pi_1+\pi_3; 1-p_2) p_2^{\pi_2-1}(1-p_2)^{\pi_1+\pi_3-1}$$

for $0 < p_2 < 1$ and $\pi_1, \pi_2, \pi_3, d > 0$ and where ${}_2F_1(a, b; c; z)$ is the Gauss hypergeometric function (see Appendix B, Definition 4). Whilst this is not an exact univariate beta type I distribution, the kernel suggests that this marginal distribution is proportional to an univariate $Beta(\pi_2, \pi_1 + \pi_3)$

distribution.

3.1.3 Methods of Derivation

The Connor and Mosimann distribution is a generalization of the Dirichlet type I distribution, taking into account the concept of neutrality. For the bivariate case a vector of proportions (P_1, P_2, P_3) is said to be neutral if P_1 is independent of $(\frac{P_2}{1-P_1}, \frac{P_3}{1-P_1})$. This means that if P_1 is removed from the sample, the proportional division of P_2 and P_3 over the remaining interval $(P_1, 1)$ is not influenced by P_1. It follows that P_1 is independent of $\frac{P_2}{1-P_1}$ and $\frac{P_3}{1-P_1}$.

A simple derivation is presented for the bivariate case, but a comprehensive extension to the multivariate case can be found in Connor and Mosimann (1969).

Consider three variables P_1, P_2, and P_3 and assume that (P_1, P_2, P_3) is neutral. Then the transformation

$$\begin{aligned} Z_1 &= P_1 \\ Z_2 &= \frac{P_2}{1-P_1} \\ Z_3 &= \frac{1-P_1-P_2}{1-P_1-P_2} = 1 \end{aligned}$$

results in two independent random variables, since Z_3 is constant. The Jacobian of the transformation is given by

$$\begin{aligned} J((z_1, z_2) \rightarrow (p_1, p_2)) &= \begin{vmatrix} \frac{\partial z_1}{\partial p_1} & \frac{\partial z_1}{\partial p_2} \\ \frac{\partial z_2}{\partial p_1} & \frac{\partial z_2}{\partial p_2} \end{vmatrix} \\ &= \begin{vmatrix} 1 & 0 \\ \frac{p_2}{(1-p_1)^2} & \frac{1}{1-p_1} \end{vmatrix} \\ &= \frac{1}{1-p_1} \end{aligned}$$

Let the density functions of Z_1 and Z_2 be univariate beta distributions, i.e. $Z_i \sim Beta(a_i, b_i)$, for $i = 1, 2$. Then

$$f(z_i) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} z_i^{a_i - 1}(1 - z_i)^{b_i - 1}$$

for $i = 1, 2$. The joint density function of P_1 and P_2 is given by

$$\begin{aligned} f(p_1, p_2) &= f(z_1, z_2)J((z_1, z_2) \to (p_1, p_2)) \\ &= f(z_1)f(z_2)J((z_1, z_2) \to (p_1, p_2)) \end{aligned}$$

since Z_1 and Z_2 are independent. Then

$$\begin{aligned} f(p_1, p_2) &= \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(b_1)} \frac{\Gamma(a_2 + b_2)}{\Gamma(a_2)\Gamma(b_2)} p_1^{a_1 - 1}(1 - p_1)^{b_1 - 1}(\frac{p_2}{1 - p_1})^{a_2 - 1}(1 - \frac{p_2}{1 - p_1})^{b_2 - 1}(1 - p_1)^{-1} \\ &= \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\Gamma(b_1)} \frac{\Gamma(a_2 + b_2)}{\Gamma(a_2)\Gamma(b_2)} p_1^{a_1 - 1} p_2^{a_2 - 1}(1 - p_1 - p_2)^{b_2 - 1}(1 - p_1)^{b_1 - a_2 - b_2} \end{aligned}$$

Letting $a_1 = \pi_1$, $a_2 = \pi_2$, $b_1 = d$ and $b_2 = \pi_3$ the expression simplifies to (3.1).

3.1.4 Correlation

A comprehensive generalisation of the calculation of moments for the multivariate case can be found in Connor and Mosimann (1969), which involves neutrality. The simplified derivation of the correlation in this section only relies on the mathematical properties of the bivariate distribution.

Using the binomial expansion in (B.3), it follows that:

$$(1 - p_1)^{d - \pi_2 - \pi_3} = \sum_{r=0}^{\infty} \binom{d - \pi_2 - \pi_3}{r} (-1)^r p_1^r \tag{3.2}$$

which will always converge since $0 < p_1 < 1$. The density function in (3.1) can thus be written as

$$f(p_1, p_2) = \frac{\Gamma(\pi_1 + d)\Gamma(\pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} \sum_{r=0}^{\infty} \binom{d - \pi_2 - \pi_3}{r} (-1)^r p_1^{\pi_1 + r - 1} p_2^{\pi_2 - 1}(1 - p_1 - p_2)^{\pi_3 - 1} \tag{3.3}$$

The product moments can be derived as

$$\begin{aligned}E[P_1^iP_2^j] &= \frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} \\ &\quad\times\int_0^1\int_0^{1-p_2} p_1^{\pi_1+i-1}p_2^{\pi_2+j-1}(1-p_1-p_2)^{\pi_3-1}(1-p_1)^{d-\pi_2-\pi_3}dp_1dp_2 \end{aligned} \tag{3.4}$$

Considering the integral in (3.4) and using (3.2),

$$\begin{aligned}&\int_0^1\int_0^{1-p_2} p_1^{\pi_1+i-1}p_2^{\pi_2+j-1}(1-p_1-p_2)^{\pi_3-1}(1-p_1)^{d-\pi_2-\pi_3}dp_1dp_2 \\ =\ &\int_0^1\int_0^{1-p_2} p_1^{\pi_1+i-1}p_2^{\pi_2+j-1}(1-p_1-p_2)^{\pi_3-1}\sum_{r=0}^{\infty}\binom{d-\pi_2-\pi_3}{r}(-1)^r p_1^r dp_1dp_2 \\ =\ &\sum_{r=0}^{\infty}\binom{d-\pi_2-\pi_3}{r}(-1)^r\int_0^1\int_0^{1-p_2} p_1^{\pi_1+i+r-1}p_2^{\pi_2+j-1}(1-p_1-p_2)^{\pi_3-1}dp_1dp_2 \\ =\ &\sum_{r=0}^{\infty}\binom{d-\pi_2-\pi_3}{r}(-1)^r B(\pi_1+i+r,\pi_2+j,\pi_3) \\ &\times\int_0^1\int_0^{1-p_2}[B(\pi_1+i+r,\pi_2+j,\pi_3)]^{-1}p_1^{\pi_1+i+r-1}p_2^{\pi_2+j-1} \\ &\times(1-p_1-p_2)^{\pi_3-1}dp_1dp_2 \end{aligned} \tag{3.5}$$

The integral in (3.5) is equal to 1, since it corresponds to the total probability of a $BBeta^I(\pi_1+i+r,\pi_2+j,\pi_3)$ distribution. Combining (3.4) and (3.5),

$$\begin{aligned}&E[P_1^iP_2^j] \\ =\ &\frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)}\sum_{r=0}^{\infty}\binom{d-\pi_2-\pi_3}{r}(-1)^r[B(\pi_1+i+r,\pi_2+j,\pi_3)]^{-1}\end{aligned} \tag{3.6}$$

The summation in (3.6) can be simplified as

$$\begin{aligned}
&\sum_{r=0}^{\infty}\binom{d-\pi_2-\pi_3}{r}(-1)^r[B(\pi_1+i+r,\pi_2+j,\pi_3)]^{-1}\\
=&\sum_{r=0}^{\infty}\binom{d-\pi_2-\pi_3}{r}(-1)^r\frac{\Gamma(\pi_1+i+r)\Gamma(\pi_2+j)\Gamma(\pi_3)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j+r)}\\
=&\,\Gamma(\pi_2+j)\Gamma(\pi_3)\sum_{r=0}^{\infty}\frac{(-1)^{2r}}{r!}(\pi_2+\pi_3-d)_r\frac{\Gamma(\pi_1+i+r)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j+r)}
\end{aligned}$$

where the last step is obtained by using relation (B.7). Then

$$\begin{aligned}
&\Gamma(\pi_2+j)\Gamma(\pi_3)\sum_{r=0}^{\infty}\frac{(-1)^{2r}}{r!}(\pi_2+\pi_3-d)_r\frac{\Gamma(\pi_1+i+r)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j+r)}\\
=&\frac{\Gamma(\pi_2+j)\Gamma(\pi_3)\Gamma(\pi_1+i)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j)}\sum_{r=0}^{\infty}\frac{1}{r!}(\pi_2+\pi_3-d)_r\frac{\Gamma(\pi_1+i+r)\Gamma(\pi_1+\pi_2+\pi_3+i+j)}{\Gamma(\pi_1+i)\Gamma(\pi_1+\pi_2+\pi_3+i+j+r)}\\
=&\frac{\Gamma(\pi_2+j)\Gamma(\pi_3)\Gamma(\pi_1+i)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j)}\sum_{r=0}^{\infty}\frac{1}{r!}\frac{(\pi_2+\pi_3-d)_r(\pi_1+i)_r}{(\pi_1+\pi_2+\pi_3+i+j)_r}\\
=&\frac{\Gamma(\pi_2+j)\Gamma(\pi_3)\Gamma(\pi_1+i)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j)}\,{}_2F_1(\pi_2+\pi_3-d,\pi_1+i;\pi_1+\pi_2+\pi_3+i+j;1) \qquad (3.7)
\end{aligned}$$

where (3.7) is obtained by using the definitions of the Pochhammer coefficient (Appendix B, Definition 3) and the Gauss hypergeometric function (Appendix B, Definition 4) respectively. By using relation (B.8), (3.7) becomes

$$\begin{aligned}
&\frac{\Gamma(\pi_2+j)\Gamma(\pi_3)\Gamma(\pi_1+i)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j)}\frac{\Gamma(\pi_1+\pi_2+\pi_3+i+j)\Gamma(j+d)}{\Gamma(\pi_1+i+j+d)\Gamma(\pi_2+\pi_3+j)}\\
=&\frac{\Gamma(\pi_2+j)\Gamma(\pi_3)\Gamma(\pi_1+i)\Gamma(j+d)}{\Gamma(\pi_2+\pi_3+j)\Gamma(\pi_1+i+j+d)} \qquad (3.8)
\end{aligned}$$

and combining (3.6) and (3.8) gives

$$E[P_1^iP_2^j]=\frac{\Gamma(\pi_1+i)\Gamma(\pi_2+j)\Gamma(d+j)\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_1+d+i+j)\Gamma(\pi_2+\pi_3+j)} \qquad (3.9)$$

Using (3.9) the variance of P_1 can be calculated as:

$$E[P_1] = \frac{\pi_1}{\pi_1 + d}$$

$$E[P_1^2] = \frac{\pi_1(\pi_1+1)}{(\pi_1+d)(\pi_1+d+1)}$$

$$var(P_1) = \frac{\pi_1 d}{(\pi_1+d)^2(\pi_1+d+1)}$$

Similarly,

$$E[P_2] = \frac{\pi_2 d}{(\pi_1+d)(\pi_2+\pi_3)}$$

$$E[P_2^2] = \frac{\pi_2 d(\pi_2+1)(d+1)}{(\pi_1+d)(\pi_1+d+1)(\pi_2+\pi_3)(\pi_2+\pi_3+1)}$$

$$var(P_2) = \frac{\pi_2 d\{\pi_1\pi_2(\pi_2+1)+\pi_3[(d+1)(d+\pi_1)+\pi_1\pi_2]\}}{(\pi_1+d)^2(\pi_1+d+1)(\pi_2+\pi_3)^2(\pi_2+\pi_3+1)}$$

The covariance can be calculated as:

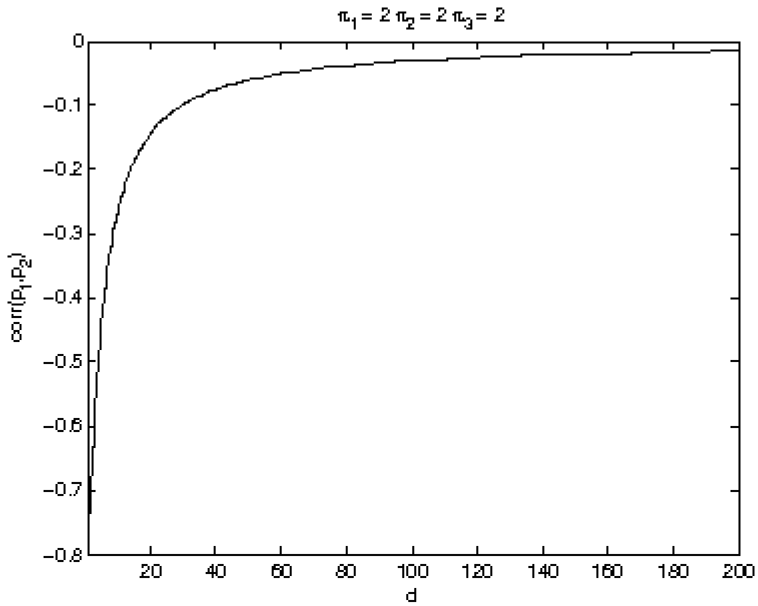
$$cov(P_1,P_2) \;=\; \frac{\pi_1\pi_2 d}{(\pi_1+d)(\pi_1+d+1)(\pi_2+\pi_3)} - \frac{\pi_1\pi_2 d}{(\pi_1+d)^2(\pi_2+\pi_3)}$$

The covariance will be negative if

$$\begin{aligned}
\frac{\pi_1\pi_2 d}{(\pi_1+d)(\pi_1+d+1)(\pi_2+\pi_3)} &< \frac{\pi_1\pi_2 d}{(\pi_1+d)^2(\pi_2+\pi_3)} \\
\Longleftrightarrow (\pi_1+d)(\pi_1+d+1)(\pi_2+\pi_3) &> (\pi_1+d)^2(\pi_2+\pi_3) \\
\Longleftrightarrow \pi_1+d+1 &> \pi_1+d \\
\Longleftrightarrow 1 &> 0
\end{aligned}$$

and since this always holds, the correlation between P_1 and P_2 will always be negative. Figure 3.1 plots the correlation for various values of d where $\pi_1 = \pi_2 = \pi_3 = 2$. The correlation tends to 0 for larger values of d, but remains negative, as per the derivation above.

Figure 3.1: Connor and Mosimann Bivariate Beta Distribution: Correlation



Note that for the bivariate case, the Connor and Mosimann bivariate beta distribution only allows for negative correlation. For more than 2 variables, the generalised covariance structure does allow for positive correlation, see Connor and Mosimann (1969), and Wong (1998).

3.1.5 Shape Analysis

The parameters of the prior distribution can be estimated from historical data using maximum likelihood estimation, or determined from prior knowledge or expert judgement, see O' Hagan and Forster (2004). The examples that follow will study the effect of d on the shape and concentration of the distribution. The reference case considered is where $\pi_1 = \pi_2 = \pi_3 = 2$ and $d = 4$.

Figure 3.2 shows that if π_1 is decreased and π_2, π_3 and d remain constant, the distribution shifts towards the marginal distribution of P_2. If π_1 is increased and π_2, π_3 and d remain constant, the distribution shifts towards smaller values of P_2 along the line $p_1 + p_2 = 1$. This behaviour is exactly the same observed in Figure 2.1. This is not surprising since $d = \pi_2 + \pi_3$, which means that for this specific set of parameters the Connor and Mosimann bivariate beta distribution is actually the bivariate beta type I distribution.

Figure 3.2: Connor and Mosimann Bivariate Beta Distribution: Changing π_1

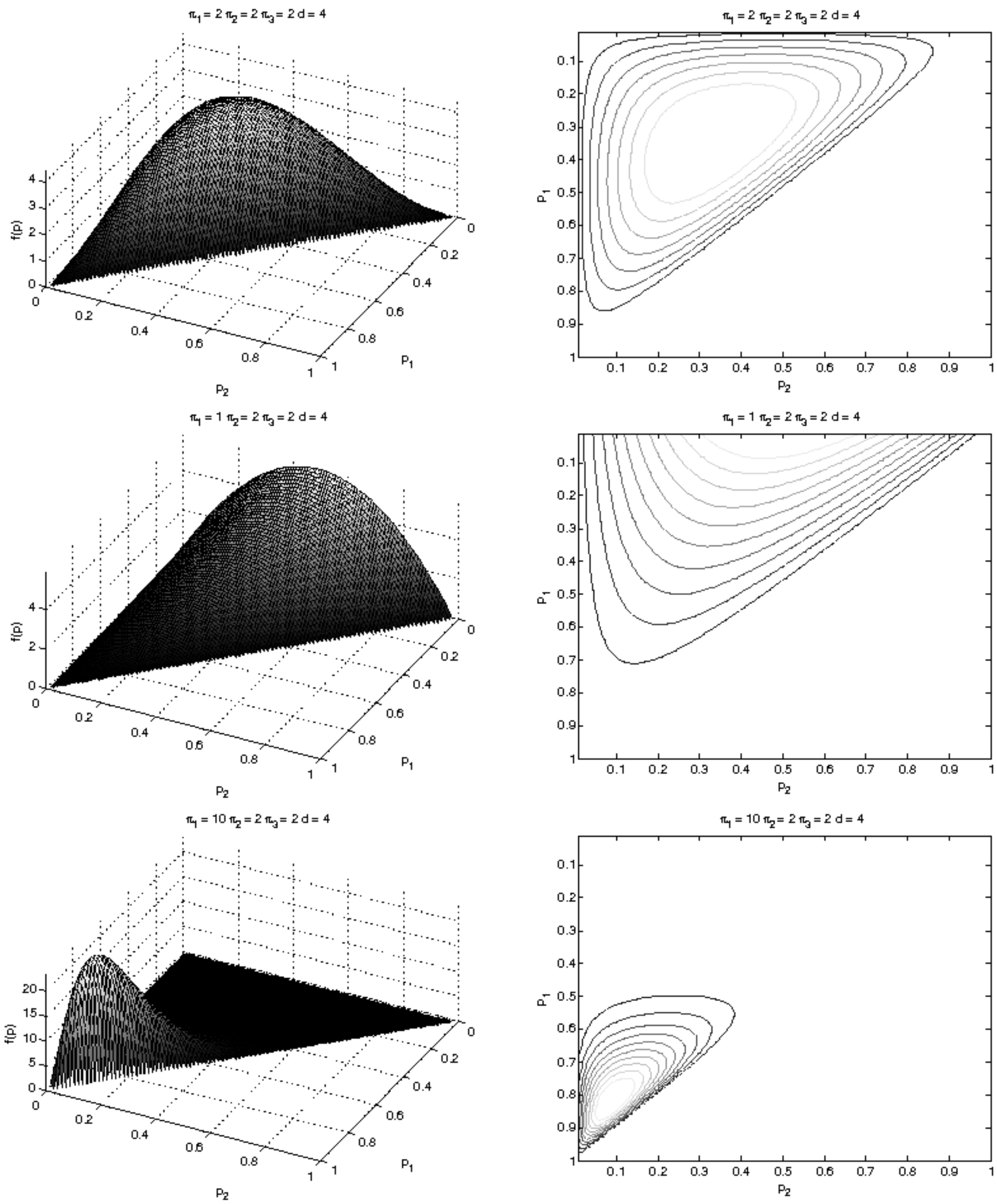


Figure 3.3 shows that if π_2 is decreased and π_1, π_3 and d remain constant, the distribution shifts towards the marginal distribution of P_1 on the left axis. If π_2 is increased and π_1, π_2 and d remain constant, the distribution shifts towards larger values of P_2 along the line $p_1 + p_2 = 1$. Whilst the direction of the shift is the same as observed in Figure 2.2 for the bivariate beta type I distribution,

the shift is not as pronounced. It appears as if the inclusion of the parameter d dampens the severity of the shift.

Figure 3.3: Connor and Mosimann Bivariate Beta Distribution: Changing π_2

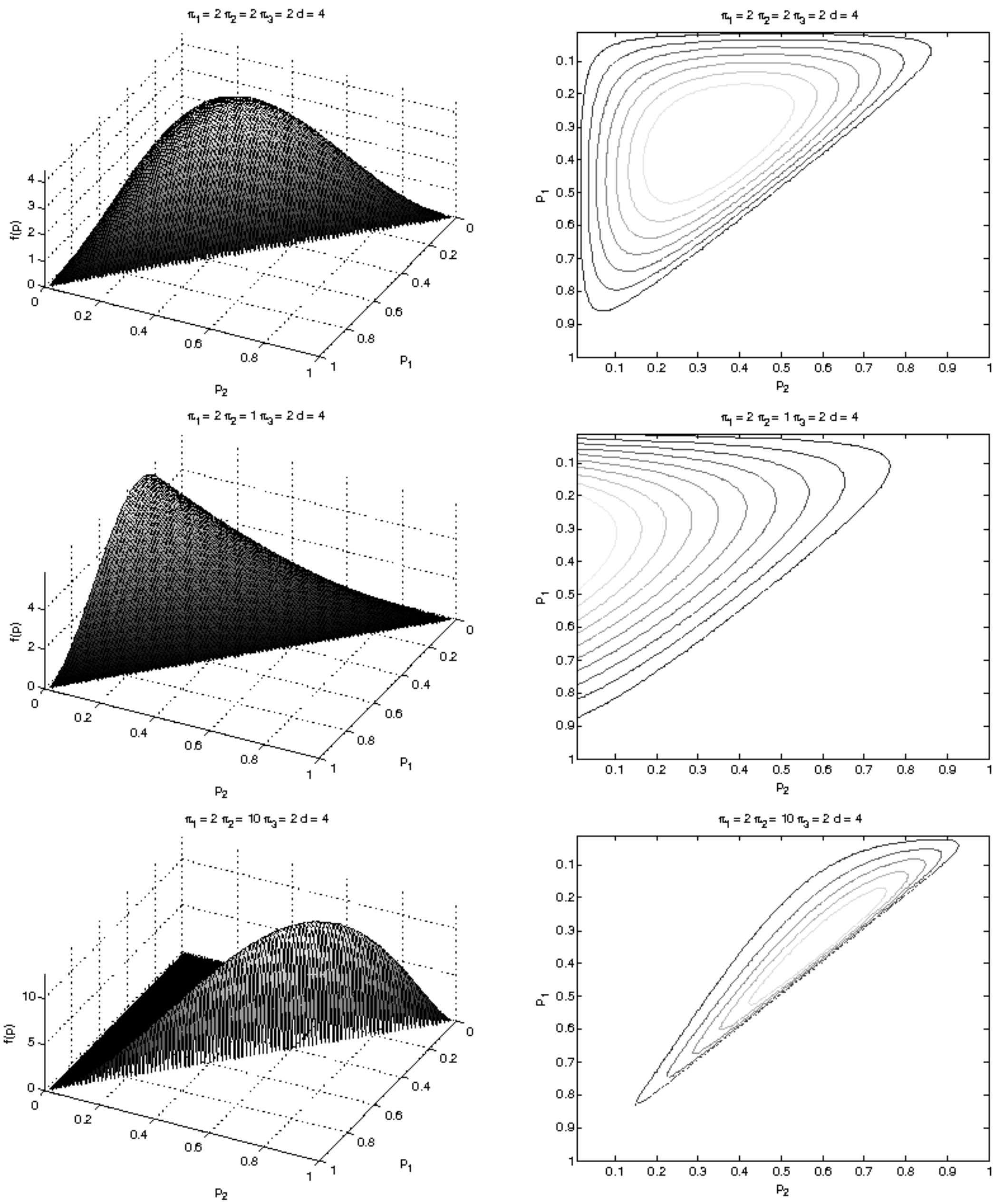


Figure 3.4 shows that if π_3 is decreased and π_1, π_2 and d remain constant, the concentration shifts towards the line $p_1 + p_2 = 1$. If π_3 is increased and π_1, π_2 and d remain constant, the distribution shifts towards small values of P_1 and P_2. Again, the direction of the shifts is the same as observed

in Figure 2.3 for the bivariate beta type I distribution, but the magnitude of this shift is much smaller. This confirms that the inclusion of the parameter d dampens the severity of the shift.

Figure 3.4: Connor and Mosimann Bivariate Beta Distribution: Changing π_3

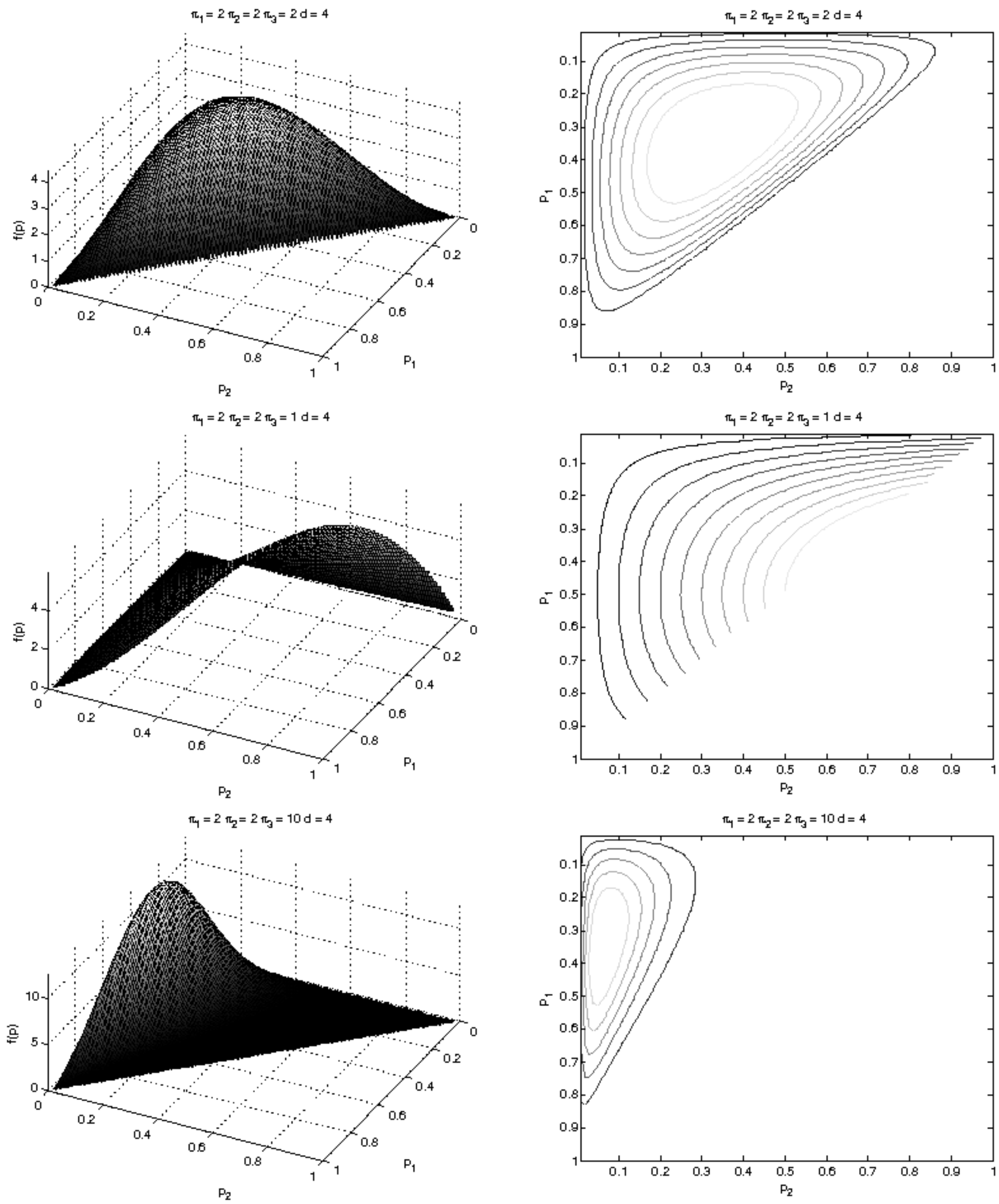
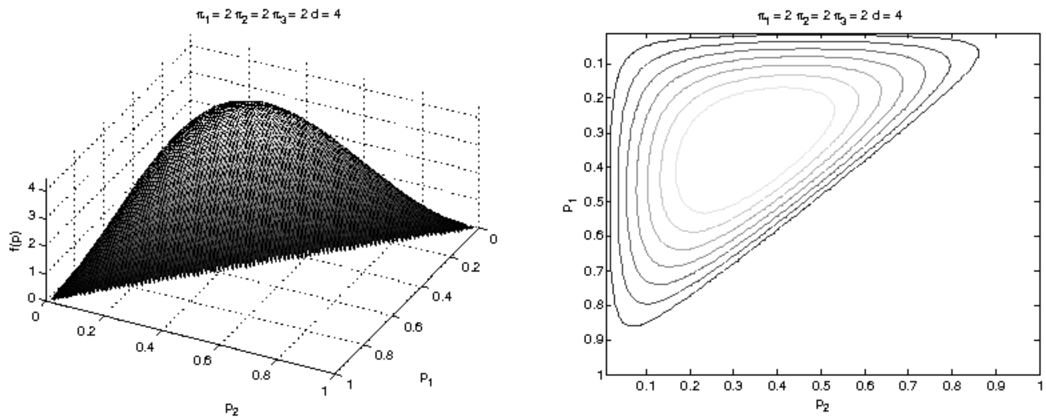
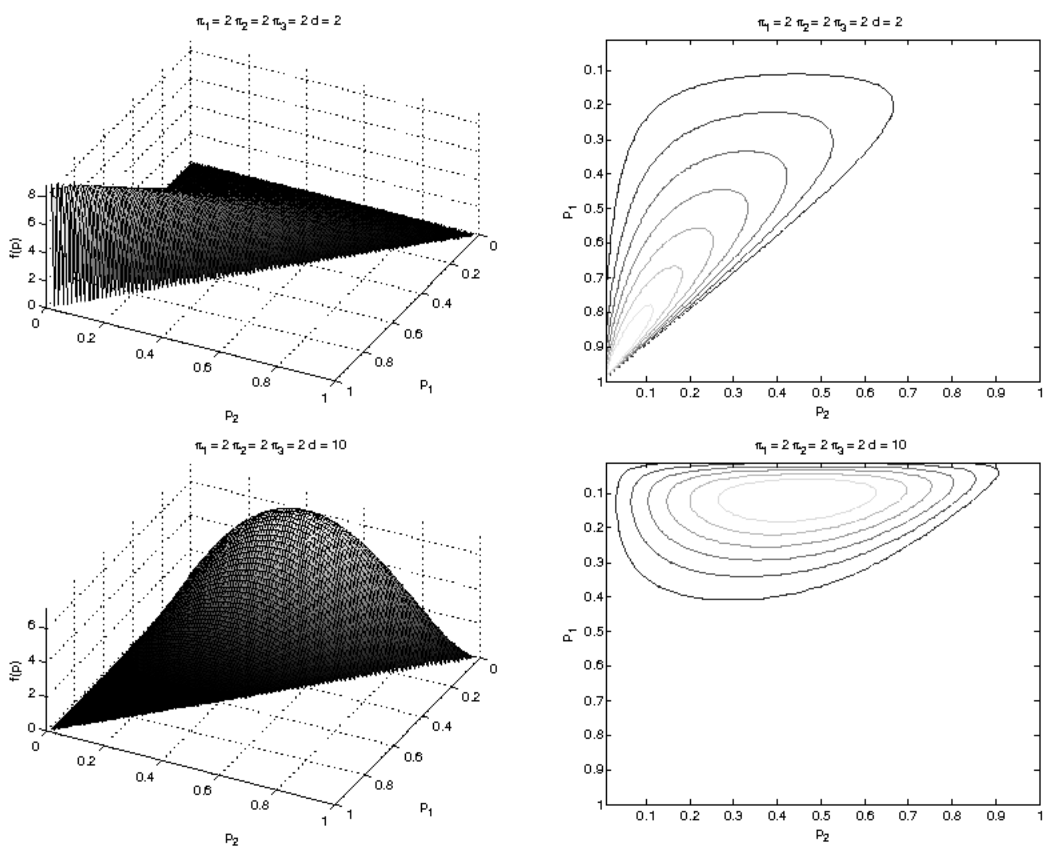


Figure 3.5 shows that if d is decreased, and π_1, π_2 and π_3 remain constant, the distribution shifts towards smaller values of P_2 along the line $p_1 + p_2 = 1$. If d is increased and π_1, π_2 and π_3 remain constant, the distribution shifts towards the marginal distribution of P_2.

Figure 3.5: Connor and Mosimann Bivariate Beta Distribution: Changing d



(a)



3.2 Bayesian Estimation of Shannon Entropy

3.2.1 Derivation

Theorem 3.1

The posterior distribution for the multinomial likelihood in (1.3) *and the Connor and Mosimann bivariate beta distribution in* (3.1) *is given by*

$$f(p_1, p_2|\mathbf{x}) = K\sum_{r=0}^{\infty}\begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix}(-1)^r p_1^{\pi_1+x_1+r-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1} \quad (3.10)$$

where $K = \frac{\Gamma(\pi_2+x_2+\pi_3+x_3)\Gamma(\pi_1+x_1+x_2+x_3+d)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(x_2+x_3+d)}$ *is the normalising coefficient,* $0 < p_i < 1$ *for* $i = 1, 2$, $0 < p_1 + p_2 < 1$ *and* $\pi_1, \pi_2, \pi_3, d > 0$.

Proof

Using the binomially expanded form of the Connor and Mosimann bivariate beta distribution given in (3.3), the numerator of the posterior distribution is given by

$$\begin{aligned} & f(\mathbf{x}|p_1, p_2)f(p_1, p_2) \\ = \; & \frac{n!}{x_1!x_2!x_3!}\frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} \\ & \times\sum_{r=0}^{\infty}\begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix}(-1)^r p_1^{\pi_1+x_1+r-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1} \end{aligned} \quad (3.11)$$

The denominator of the posterior distribution is given by

$$
\begin{aligned}
& \int\int f(\mathbf{x}|p_1,p_2)f(p_1,p_2)dp_1dp_2 \\
= \; & \frac{n!}{x_1!x_2!x_3!}\frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)}\sum_{r=0}^{\infty}\left(\begin{array}{c} d-\pi_2-\pi_3 \\ r \end{array}\right)(-1)^r \\
& \times \int_0^1\int_0^{1-p_2} p_1^{\pi_1+x_1+r-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \qquad (3.12) \\
= \; & \frac{n!}{x_1!x_2!x_3!}\frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} \\
& \times \sum_{r=0}^{\infty}\left(\begin{array}{c} d-\pi_2-\pi_3 \\ r \end{array}\right)(-1)^r B(\pi_1+x_1+r,\pi_2+x_2,\pi_3+x_3) \qquad (3.13)
\end{aligned}
$$

since the integral in (3.12) is proportional to a $BBeta^I(\pi_1+x_1+r,\pi_2+x_2,\pi_3+x_3)$ distribution. The summation in (3.13) can be simplified to

$$
\begin{aligned}
& \sum_{r=0}^{\infty}\left(\begin{array}{c} d-\pi_2-\pi_3 \\ r \end{array}\right)(-1)^r B(\pi_1+x_1+r,\pi_2+x_2,\pi_3+x_3) \\
= \; & \sum_{r=0}^{\infty}\left(\begin{array}{c} d-\pi_2-\pi_3 \\ r \end{array}\right)(-1)^r \frac{\Gamma(\pi_1+x_1+r)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)} \\
= \; & \Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\sum_{r=0}^{\infty}\frac{(-1)^{2r}}{r!}(\pi_2+\pi_3-d)_r\frac{\Gamma(\pi_1+x_1+r)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)}
\end{aligned}
$$

by using relation (B.7). Then

$$\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\sum_{r=0}^{\infty}\frac{(-1)^{2r}}{r!}(\pi_2+\pi_3-d)_r\frac{\Gamma(\pi_1+x_1+r)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)}$$

$$\begin{aligned}
&= \frac{\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(\pi_1+x_1)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)} \\
&\quad \times \sum_{r=0}^{\infty}\frac{1}{r!}(\pi_2+\pi_3-d)_r\frac{\Gamma(\pi_1+x_1+r)\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)} \\
&= \frac{\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(\pi_1+x_1)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}\sum_{r=0}^{\infty}\frac{1}{r!}\frac{(\pi_2+\pi_3-d)_r(\pi_1+x_1)_r}{(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)_r} \\
&= \frac{\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(\pi_1+x_1)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)} \\
&\quad \times {}_2F_1(\pi_2+\pi_3-d,\pi_1+x_1;\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1) \qquad (3.14)
\end{aligned}$$

where (3.14) is obtained by using the definitions of the Pochhammer coefficient (Appendix B, Definition 3) and Gauss hypergeometric function (Appendix B, Definition 4) respectively. By using relation (B.8), (3.14) simplifies to

$$\frac{\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(\pi_1+x_1)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}\frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)\Gamma(n_2+x_3+d)}{\Gamma(\pi_1+x_1+x_2+x_3+d)\Gamma(\pi_2+x_2+\pi_3+x_3+d)}$$

$$= \frac{\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(\pi_1+x_1)\Gamma(x_2+x_3+d)}{\Gamma(\pi_2+x_2+\pi_3+x_3)\Gamma(\pi_1+x_1+x_2+x_3+d)} \qquad (3.15)$$

and combining (3.13) and (3.15) gives

$$\begin{aligned}
&\int\int f(\mathbf{x}|p_1,p_2)f(p_1,p_2)dp_1dp_2 \\
&= \frac{n!}{x_1!x_2!x_3!}\frac{\Gamma(\pi_1+d)\Gamma(\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)\Gamma(d)} \\
&\quad \times\frac{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(x_2+x_3+d)}{\Gamma(\pi_2+x_2+\pi_3+x_3)\Gamma(\pi_1+x_1+x_2+x_3+d)} \qquad (3.16)
\end{aligned}$$

Combining the expressions for the numerator, (3.11), and denominator, (3.16), the posterior distribution is given by:

$$f(p_1, p_2|\mathbf{x}) = K \sum_{r=0}^{\infty} \begin{pmatrix} d - \pi_2 - \pi_3 \\ r \end{pmatrix} (-1)^r p_1^{\pi_1+x_1+r-1} p_2^{\pi_2+x_2-1} (1 - p_1 - p_2)^{\pi_3+x_3-1}$$

where $K = \frac{\Gamma(\pi_2+x_2+\pi_3+x_3)\Gamma(\pi_1+x_1+x_2+x_3+d)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(x_2+x_3+d)}$. ■

Since the form of the posterior distribution is not the same as that of the prior distribution, it follows that the Connor and Mosimann bivariate beta distribution is not a natural conjugate for the multinomial distribution.

Theorem 3.2

The Bayesian estimator of the Shannon entropy under the squared error loss function using the Connor and Mosimann bivariate beta distribution as a prior for the multinomial distribution in (1.3) *is given by:*

$$\begin{aligned}\hat{H}_3^{CM} &= -K\sum_{r=0}^{\infty}\binom{d-\pi_2-\pi_3}{r}(-1)^r\frac{\Gamma(\delta_1)\Gamma(\delta_2)\Gamma(\delta_3)}{\Gamma(\delta_1+\delta_2+\delta_3+1)}\\ &\quad\times\sum_{i=1}^{3}\delta_i[\psi(\delta_i+1)-\psi(\sum_{j=1}^{3}\delta_j+1)]\end{aligned} \tag{3.17}$$

where

$$K=\frac{\Gamma(\pi_2+x_2+\pi_3+x_3)\Gamma(\pi_1+x_1+x_2+x_3+d)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(x_2+x_3+d)}$$

and

$$\begin{aligned}\delta_1 &= \pi_1+x_1+r\\ \delta_2 &= \pi_2+x_2\\ \delta_3 &= \pi_3+x_3\end{aligned}$$

are the normalising coefficient and parameters of the posterior distribution in (3.10) *respectively.*

Proof

From (1.1) the Shannon entropy for the Connor and Mosimann bivariate beta distribution is denoted by

$$H_3^{CM}=-\sum_{i=1}^{3}p_i\ln p_i$$

The Bayesian estimator of the Shannon entropy under squared error loss is defined as the expected value of the Shannon entropy with respect to the posterior distribution obtained in (3.10). That is,

$$\begin{aligned}
\hat{H}_3^{CM} &= E_{f(p_1,p_2|\mathbf{x})}[H_3^{CM}] \\
&= -K\int_0^1\int_0^{1-p_2}\sum_{i=1}^{3} p_i \ln p_i \sum_{r=0}^{\infty}\begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix}(-1)^r p_1^{\pi_1+x_1+r-1}p_2^{\pi_2+x_2-1} \\
&\quad \times(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \\
&= -K\sum_{i=1}^{3} I_i
\end{aligned}$$

where

$$K = \frac{\Gamma(\pi_2+x_2+\pi_3+x_3)\Gamma(\pi_1+x_1+x_2+x_3+d)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)\Gamma(x_2+x_3+d)}$$

$$\begin{aligned}
I_i &= \int_0^1\int_0^{1-p_2} p_i \ln p_i \\
&\quad \times\sum_{r=0}^{\infty}\begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix}(-1)^r p_1^{\pi_1+x_1+r-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2
\end{aligned}$$

for $i = 1, 2$, and

$$\begin{aligned}
I_3 &= \int_0^1\int_0^{1-p_2}(1-p_1-p_2)\ln(1-p_1-p_2) \\
&\quad \times\sum_{r=0}^{\infty}\begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix}(-1)^r p_1^{\pi_1+x_1+r-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2
\end{aligned}$$

The simplification of I_i will only be shown for I_1, but follows similarly for I_2 and I_3.

$$\begin{aligned}
I_1 &= \int_0^1 \int_0^{1-p_2} p_1 \ln p_1 \\
&\quad \times \sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r p_1^{\pi_1+x_1+r-1} p_2^{\pi_2+x_2-1} (1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \\
&= \sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r \int_0^1 \int_0^{1-p_2} (\ln p_1) p_1^{\pi_1+x_1+r} p_2^{\pi_2+x_2-1} (1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \\
&= \sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r \int_0^1 \int_0^{1-p_2} [\frac{\partial}{\partial \pi_1} p_1^{\pi_1+x_1+r}] p_2^{\pi_2+x_2-1} (1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2
\end{aligned}$$

since $\frac{d}{dx} a^x = a^x \ln a$. Changing the order of integration and differentiation:

$$\begin{aligned}
I_1 &= \sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r \\
&\quad \times \frac{\partial}{\partial \pi_1} \int_0^1 \int_0^{1-p_2} p_1^{\pi_1+x_1+r} p_2^{\pi_2+x_2-1} (1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \qquad (3.18) \\
&= \sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r \frac{\partial}{\partial \pi_1} B(\pi_1+x_1+r+1, \pi_2+x_2, \pi_3+x_3)
\end{aligned}$$

since the integral in (3.18) is proportional to a $BBeta^I(\pi_1+x_1+r+1, \pi_2+x_2, \pi_3+x_3)$ distribution. Using the product and chain rules for differentiation:

$$\begin{aligned}
I_1 &= \sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r B(\pi_1+x_1+r+1, \pi_2+x_2, \pi_3+x_3) \\
&\quad \times [\psi(\pi_1+x_1+r+1) - \psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+1)]
\end{aligned}$$

where $\psi(x)$ denotes the polygamma function. Similarly,

$$\begin{aligned} I_2 &= \sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r B(\pi_1+x_1+r, \pi_2+x_2+1, \pi_3+x_3) \\ &\quad \times[\psi(\pi_2+x_2+1)-\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+1)] \end{aligned}$$

and

$$\begin{aligned} I_3 &= \sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r B(\pi_1+x_1+r, \pi_2+x_2, \pi_3+x_3+1) \\ &\quad \times[\psi(\pi_3+x_3+1)-\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+1)] \end{aligned}$$

Adding I_1, I_2 and I_3 together

$$\begin{aligned} &\sum_{i=1}^{3} I_i \\ = &\sum_{r=0}^{\infty} \begin{pmatrix} d-\pi_2-\pi_3 \\ r \end{pmatrix} (-1)^r \frac{\Gamma(\pi_1+x_1+r)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)} \\ &\times[(\pi_1+x_1+r)\psi(\pi_1+x_1+r+1)+(\pi_2+x_2)\psi(\pi_2+x_2+1)+(\pi_3+x_3)\psi(\pi_3+x_3+1) \\ &-(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)] \end{aligned}$$

For simplification, let

$$\begin{aligned} \delta_1 &= \pi_1+x_1+r \\ \delta_2 &= \pi_2+x_2 \\ \delta_3 &= \pi_3+x_3 \end{aligned}$$

The Bayesian estimator of the Shannon entropy under squared error loss using the Connor and

Mosimann bivariate beta distribution as a prior is then written as:

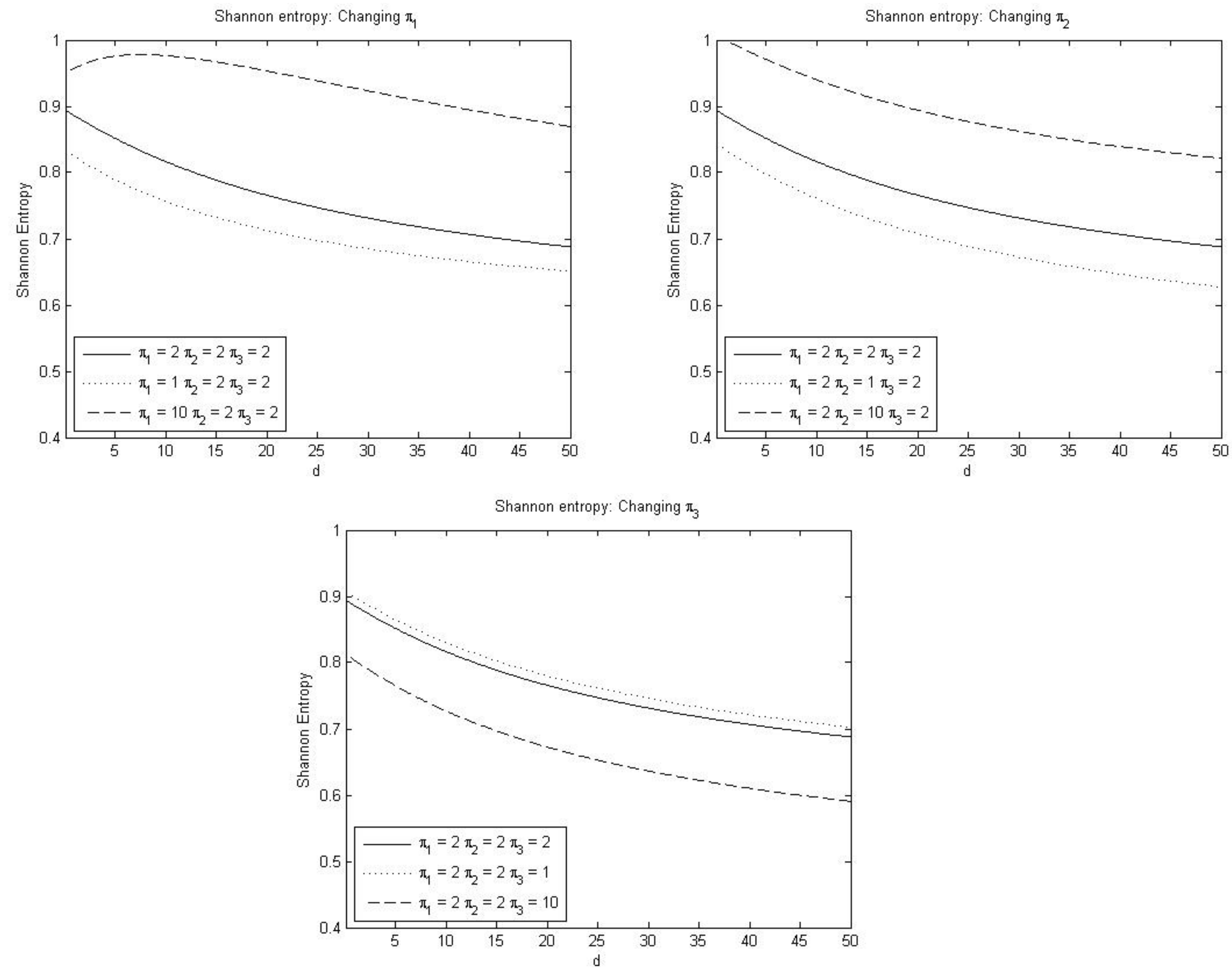
$$\hat{H}_3^{CM} = -K \sum_{r=0}^{\infty} \binom{d - \pi_2 - \pi_3}{r} (-1)^r \frac{\Gamma(\delta_1)\Gamma(\delta_2)\Gamma(\delta_3)}{\Gamma(\delta_1 + \delta_2 + \delta_3 + 1)} \sum_{i=1}^{3} \delta_i [\psi(\delta_i + 1) - \psi(\sum_{j=1}^{3} \delta_j + 1)]$$

where K is the normalising coefficient of the posterior distribution in (3.10). ■

3.2.2 Numerical Analysis

Figure 3.6 compares the Shannon entropy values obtained when using the Bayesian estimator derived in (3.17), for each of the parameter combinations discussed in the shape analysis. The multinomial frequencies were assumed to be $x_1 = 1$, $x_2 = 2$, $x_3 = 10$.

Figure 3.6: Bayesian Estimates of Shannon Entropy: Connor and Mosiman Bivariate Beta Prior



Note that, for the first chart where the effect of changing π_1 is analysed, if $d = 4$ the Connor and Mosimann bivariate beta distribution reduces to the bivariate beta type I distribution and we have that $\hat{H}_3^{CM} = \hat{H}_3^{I}$. That is, if $\pi_1 = 1$,$\pi_2 = \pi_3 = 2$ and $d = 4$, then $\hat{H}_3^{CM} \approx 0.8$ (reading off the chart). Similarly, if $\pi_1 = 10$,$\pi_2 = \pi_3 = 2$ and $d = 4$, then $\hat{H}_3^{CM} \approx 0.97$, see Table 2.2.

In the first two charts it can be seen that if π_1 and π_2 are decreased respectively, the absolute level of $\hat{H}_3^{CM}$ decreases, indicating less uncertainty. The magnitude of the decrease is not the same, and is caused by the asymmetry of the Connor and Mosiman bivariate beta distribution. Similarly, increasing π_1 and π_2 respectively increases the absolute level of $\hat{H}_3^{CM}$, indicating more uncertainty. The magnitude and shape of the change is not the same.

Interestingly, if π_1 is increased, the Bayesian estimate of Shannon entropy is not monotonic in d; but in general larger values of d are associated with lower values of $\hat{H}_3^{CM}$, indicating less uncertainty. It is possible that this phenomenon is linked to the asymmetry of the Connor and Mosiman bivariate beta distribution.

The results of the shape analyses in Figures 3.2 and 3.3 indicate that larger values of π_1 and π_2 respectively are associated with more concentration, indicating less uncertainty. Similarly, lower values of π_1 and π_2 are associated with lower concentration, indicating more uncertainty. These results appear to be counterintuitive, given the results observed above for $\hat{H}_3^{CM}$.

Conversely, if π_3 is decreased this leads to an increase in $\hat{H}_3^{CM}$, indicating more uncertainty; and if π_3 is increased this leads to a decrease in $\hat{H}_3^{CM}$, indicating less uncertainty. This is consistent with the result in Figure 3.4, where a smaller value of π_3 indicated less concentration (i.e. more uncertainty) and a larger value of π_3 indicated more concentration (i.e. less uncertainty).

The difference between these intuitive and counterintuitive results observed for parameters π_1 and π_2 may be disturbing at first, but can be explained by the location of the distribution. In the first two cases, changing π_1 or π_2 leads to distributions that are either tending to the marginal distribution of P_2 or P_1 respectively, or are tending towards a specific point along the line $p_1 + p_2 = 1$. As the concentration in the distribution remains closer to small values of P_1 and P_2, $\hat{H}_3^{CM}$ stays lower, but as soon as the concentration moves away from these small values to some point along the line $p_1 + p_2 = 1$ the uncertainty increases. This suggests that the Bayesian estimate of Shannon entropy may contain information of not only concentration, but of location as well.

Again, it should be kept in mind that since $\hat{H}_3^{CM}$ is the *Bayesian* estimate of the Shannon entropy, it also contains information about the likelihood function. Therefore, the multinomial frequencies will also have an effect on the behaviour of $\hat{H}_3^{CM}$.

Chapter 4

Bivariate Beta Type III Prior

4.1 The Bivariate Beta type III Prior

The bivariate beta type III distribution is a simple extension of the bivariate beta type I distribution, where the extension enables the distribution to allow for both negative and positive correlation between P_1 and P_2. Cardeño et al. (2005) study the univariate properties of this distribution in detail.

4.1.1 Joint Density Function

Consider the bivariate beta type III distribution, defined in Ehlers et al. (2009), denoted by $BBeta^{III}(\pi_1, \pi_2, \pi_3, c)$ and with density function:

$$\begin{aligned} f(p_1, p_2) &= \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{\pi_1+\pi_2} p_1^{\pi_1-1} p_2^{\pi_2-1} (1-p_1-p_2)^{\pi_3-1} \\ &\quad \times [1+(c-1)p_1+(c-1)p_2]^{-(\pi_1+\pi_2+\pi_3)} \end{aligned} \tag{4.1}$$

where $0 < p_i < 1$, for $i = 1, 2$, $0 < p_1 + p_2 < 1$, and $\pi_1, \pi_2, \pi_3, c > 0$.

If $c = 1$, the last factor in the kernel above reduces to 1, therefore reducing the bivariate beta type III distribution to a bivariate beta type I distribution.

It is shown in Ehlers et al. (2009) that the inclusion of the parameter c allows the density of the distribution to shift between the larger and smaller values of P_1 and P_2, making it a more flexible distribution than the bivariate beta type I distribution. In addition, specific choices of c allows for positive correlation between P_1 and P_2, which is not the case with the bivariate beta type I or the Connor and Mosimann bivariate beta distribution.

4.1.2 Univariate Properties

The marginal density functions of P_1 and P_2 can be found by integrating $f(p_1, p_2)$ over P_2 and P_1 respectively. Since $f(p_1, p_2)$ is symmetric, the marginal density functions will be of the same form. The derivation of the marginal distribution is only shown for P_1, but follows similarly for P_2.

$$\begin{aligned} f(p_1) &= \int_0^{1-p_1} \{\frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{\pi_1+\pi_2} p_1^{\pi_1-1} p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1} \\ &\quad \times[1+(c-1)p_1+(c-1)p_2]^{-(\pi_1+\pi_2+\pi_3)}\} dp_2 \\ &= \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{\pi_1+\pi_2} p_1^{\pi_1-1} \\ &\quad \times \int_0^{1-p_1} p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}[1+(c-1)p_1+(c-1)p_2]^{-(\pi_1+\pi_2+\pi_3)} dp_2 \end{aligned} \tag{4.2}$$

Using equation (B.2), the integral in (4.2) simplifies to

$$(1-p_1)^{\pi_2+\pi_3-1}[1+(c-1)p_1]^{-(\pi_1+\pi_2+\pi_3)} \frac{\Gamma(\pi_2)\Gamma(\pi_3)}{\Gamma(\pi_2+\pi_3)} {}_2F_1(\pi_2, \pi_1+\pi_2+\pi_3; \pi_2+\pi_3; -\frac{(c-1)(1-p_1)}{1+(c-1)p_1})$$

Then applying relation (B.6),

$$\begin{aligned} &{}_2F_1(\pi_2, \pi_1+\pi_2+\pi_3; \pi_2+\pi_3; -\frac{(c-1)(1-p_1)}{1+(c-1)p_1}) \\ &= (\frac{c}{1+(c-1)p_1})^{-(\pi_1+\pi_2+\pi_3)} {}_2F_1(\pi_1+\pi_2+\pi_3, \pi_3; \pi_2+\pi_3; \frac{(c-1)(1-p_1)}{c}) \end{aligned}$$

and the marginal density function of P_1 is given by

$$f(p_1) = \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2+\pi_3)} c^{-\pi_3} p_1^{\pi_1-1}(1-p_1)^{\pi_2+\pi_3-1}\,{}_2F_1[\pi_3, \pi_1+\pi_2+\pi_3; \pi_2+\pi_3; \frac{c-1}{c}(1-p_1)]$$

for $0 < p_1 < 1$ and $\pi_1, \pi_2, \pi_3, c > 0$. Similarly,

$$f(p_2) = \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_2)\Gamma(\pi_1+\pi_3)} c^{-\pi_3} p_2^{\pi_2-1}(1-p_2)^{\pi_1+\pi_3-1}\,{}_2F_1[\pi_3, \pi_1+\pi_2+\pi_3; \pi_1+\pi_3; \frac{c-1}{c}(1-p_2)]$$

for $0 < p_2 < 1$ and $\pi_1, \pi_2, \pi_3, c > 0$.

4.1.3 Methods of Derivation

The bivariate beta type III distribution is derived using a transformation of three independently distributed χ^2 variables and the trivariate reduction method, and is briefly discussed in Ehlers et al. (2009).

Let $S_i \sim \chi^2(2\pi_i)$ for $i = 1, 2, 3$ be three independently distributed χ^2 variables. That is,

$$f(s_i) = \frac{1}{2^{\pi_i}\Gamma(\pi_i)} \exp(-\frac{s_i}{2}) s_i^{\pi_i-1}$$

for $s_i > 0$. The joint density function of these variables is given by

$$\begin{aligned} f(s_1, s_2, s_3) &= \prod_{i=1}^{3} f(s_i) \\ &= \frac{1}{2^{\sum_{i=1}^{3}\pi_i}\prod_{i=1}^{3}\Gamma(\pi_i)} \exp(-\frac{1}{2}\sum_{i=1}^{3} s_i) \prod_{i=1}^{3} s_i^{\pi_i-1} \end{aligned}$$

Consider the transformation $P_i = \frac{S_i}{S_1+S_2+cS_3}$ for $i = 1, 2$ and $P_3 = S_1 + S_2 + cS_3$. Then

$$\begin{aligned} S_1 &= P_1 P_3 \\ S_2 &= P_2 P_3 \\ S_3 &= \frac{1}{c} P_3 (1 - P_1 - P_2) \end{aligned}$$

The Jacobian for this transformation is given by

$$\begin{aligned} J &= J((s_1, s_2, s_3) \to (p_1, p_2, p_3)) \\ &= \begin{vmatrix} \frac{\partial s_1}{\partial p_1} & \frac{\partial s_1}{\partial p_2} & \frac{\partial s_1}{\partial p_3} \\ \frac{\partial s_2}{\partial p_1} & \frac{\partial s_2}{\partial p_2} & \frac{\partial s_2}{\partial p_3} \\ \frac{\partial s_3}{\partial p_1} & \frac{\partial s_3}{\partial p_2} & \frac{\partial s_3}{\partial p_3} \end{vmatrix} \\ &= \begin{vmatrix} p_3 & 0 & p_1 \\ 0 & p_3 & p_2 \\ -\frac{p_3}{c} & -\frac{p_3}{c} & \frac{1-p_1-p_2}{c} \end{vmatrix} \\ &= p_3 [p_3 \frac{(1 - p_1 - p_2)}{c} + p_2 \frac{p_3}{c}] + p_1 \frac{p_3^2}{c} \\ &= \frac{p_3^2}{c} \end{aligned}$$

The joint density function of P_1, P_2 and P_3 is given by

$$\begin{aligned} f(p_1, p_2, p_3) &= f(s_1, s_2, s_3)|J| \\ &= \frac{1}{2^{\sum_{i=1}^3 \pi_i} \prod_{i=1}^3 \Gamma(\pi_i)} (p_1 p_3)^{\pi_1 - 1} (p_2 p_3)^{\pi_2 - 1} [\frac{p_3}{c}(1 - p_1 - p_2)]^{\pi_3 - 1} \\ &\quad \times \exp(-\frac{1}{2}[p_1 p_3 + p_2 p_3 + \frac{p_3}{c}(1 - p_1 - p_2)]) \times \frac{p_3^2}{c} \\ &= \frac{1}{2^{\sum_{i=1}^3 \pi_i} \prod_{i=1}^3 \Gamma(\pi_i)} (\frac{1}{c})^{\pi_3} p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1} p_3^{\sum_{i=1}^3 \pi_i - 1} \\ &\quad \times \exp(-\frac{p_3}{2c}[1 + (c-1)p_1 + (c-1)p_2]) \end{aligned}$$

Integrating over P_3 the joint density function of P_1 and P_2 is obtained as

$$
\begin{aligned}
f(p_1,p_2) &= \int_0^\infty f(p_1,p_2,p_3)dp_3 \\
&= \frac{1}{2^{\sum_{i=1}^3\pi_i}\prod_{i=1}^3\Gamma(\pi_i)}(\frac{1}{c})^{\pi_3}p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1} \\
&\quad\times\int_0^\infty p_3^{\sum_{i=1}^3\pi_i-1}\exp(-\frac{p_3}{2c}[1+(c-1)p_1+(c-1)p_2])dp_3
\end{aligned}
$$

The integral above can be written as

$$
\begin{aligned}
&\int_0^\infty p_3^{\sum_{i=1}^3\pi_i-1}\exp(-\frac{p_3}{2c}[1+(c-1)p_1+(c-1)p_2])dp_3 \\
=\;& \frac{\Gamma(\sum_{i=1}^3\pi_i)}{(\frac{1}{2c}[1+(c-1)p_1+(c-1)p_2])^{\sum_{i=1}^3\pi_i}} \\
&\times\int_0^\infty\frac{(\frac{1}{2c}[1+(c-1)p_1+(c-1)p_2])^{\sum_{i=1}^3\pi_i}}{\Gamma(\sum_{i=1}^3\pi_i)}p_3^{\sum_{i=1}^3\pi_i-1} \\
&\times\exp(-\frac{p_3}{2c}[1+(c-1)p_1+(c-1)p_2])dp_3 \qquad (4.3)\\
=\;& \frac{\Gamma(\sum_{i=1}^3\pi_i)}{(\frac{1}{2c}[1+(c-1)p_1+(c-1)p_2])^{\sum_{i=1}^3\pi_i}}
\end{aligned}
$$

where the integral in (4.3) is 1 since it is the total probability of a $Gamma(\sum_{i=1}^3\pi_i, \frac{1}{2c}[1+(c-1)p_1+(c-1)p_2])$ distribution. Then

$$
\begin{aligned}
&f(p_1,p_2) \\
=\;& \frac{1}{2^{\sum_{i=1}^3\pi_i}\prod_{i=1}^3\Gamma(\pi_i)}(\frac{1}{c})^{\pi_3}p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}\frac{\Gamma(\sum_{i=1}^3\pi_i)}{(\frac{1}{2c}[1+(c-1)p_1+(c-1)p_2])^{\sum_{i=1}^3\pi_i}} \\
=\;& \frac{\Gamma(\sum_{i=1}^3\pi_i)}{\prod_{i=1}^3\Gamma(\pi_i)}c^{\pi_1+\pi_2}p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}[1+(c-1)p_1+(c-1)p_2])^{-\sum_{i=1}^3\pi_i}
\end{aligned}
$$

where $0<p_i<1$ for $i=1,2$, $0<p_1+p_2<1$, and $\pi_1,\pi_2,\pi_3,c>0$. It follows that P_1 and P_2 have the joint density function of a $BBeta^{III}(\pi_1,\pi_2,\pi_3,c)$, as defined in (4.1).

4.1.4 Correlation

The product moments of P_1 and P_2 are derived as:

$$
\begin{aligned}
& E[P_1^i P_2^j] \\
= & \int_0^1 \int_0^{1-p_2} p_1^i p_2^j f(p_1, p_2) dp_1 dp_2 \\
= & \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{(\pi_1+\pi_2)} \\
& \times \int_0^1 \int_0^{1-p_2} p_1^{\pi_1+i-1} p_2^{\pi_2+j-1} (1-p_1-p_2)^{\pi_3-1} [1+(c-1)p_1+(c-1)p_2]^{-(\pi_1+\pi_2+\pi_3)} dp_1 dp_2
\end{aligned}
$$

Using equation (B.4), we get:

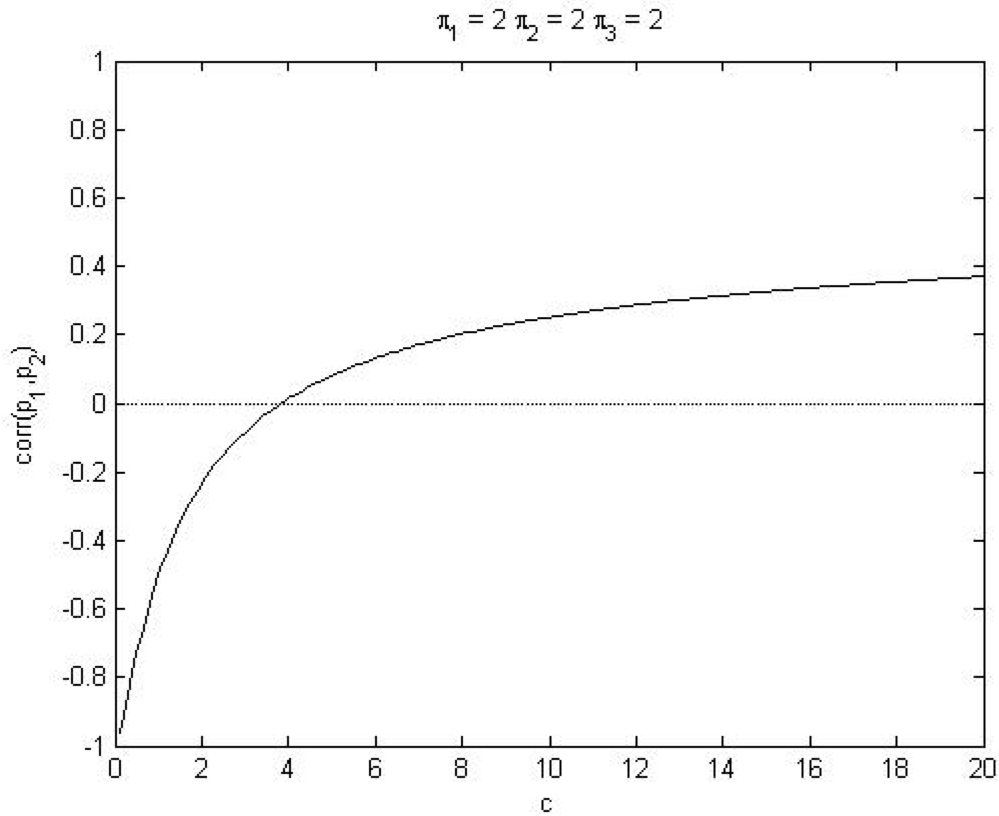
$$
\begin{aligned}
E[P_1^i P_2^j] = & \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{(\pi_1+\pi_2)} \frac{\Gamma(\pi_1+i)\Gamma(\pi_2+j)\Gamma(\pi_3)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j)} \\
& \times F_1(\pi_1+\pi_2+\pi_3, \pi_1+i, \pi_2+j, \pi_1+\pi_2+\pi_3+i+j; 1-c, 1-c)
\end{aligned}
$$

where $F_1(.)$ denotes the hypergeometric function of two variables (see Appendix B, Definition 5). Using Relation 5 in Appendix B allows this to simplify to

$$
\begin{aligned}
E[P_1^i P_2^j] = & \frac{\Gamma(\pi_1+i)\Gamma(\pi_2+j)\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_1+\pi_2+\pi_3+i+j)} c^{(\pi_1+\pi_2)} \\
& \times {}_2F_1(\pi_1+\pi_2+\pi_3, \pi_1+\pi_2+i+j; \pi_1+\pi_2+\pi_3+i+j; 1-c)
\end{aligned}
$$

Figure 4.1 shows that by keeping π_1, π_2 and π_3 constant and letting c vary, the correlation between P_1 and P_2 can be positive and negative for the bivariate beta type III distribution, dependening on the value of c.

Figure 4.1: Bivariate Beta Type III Distribution: Correlation



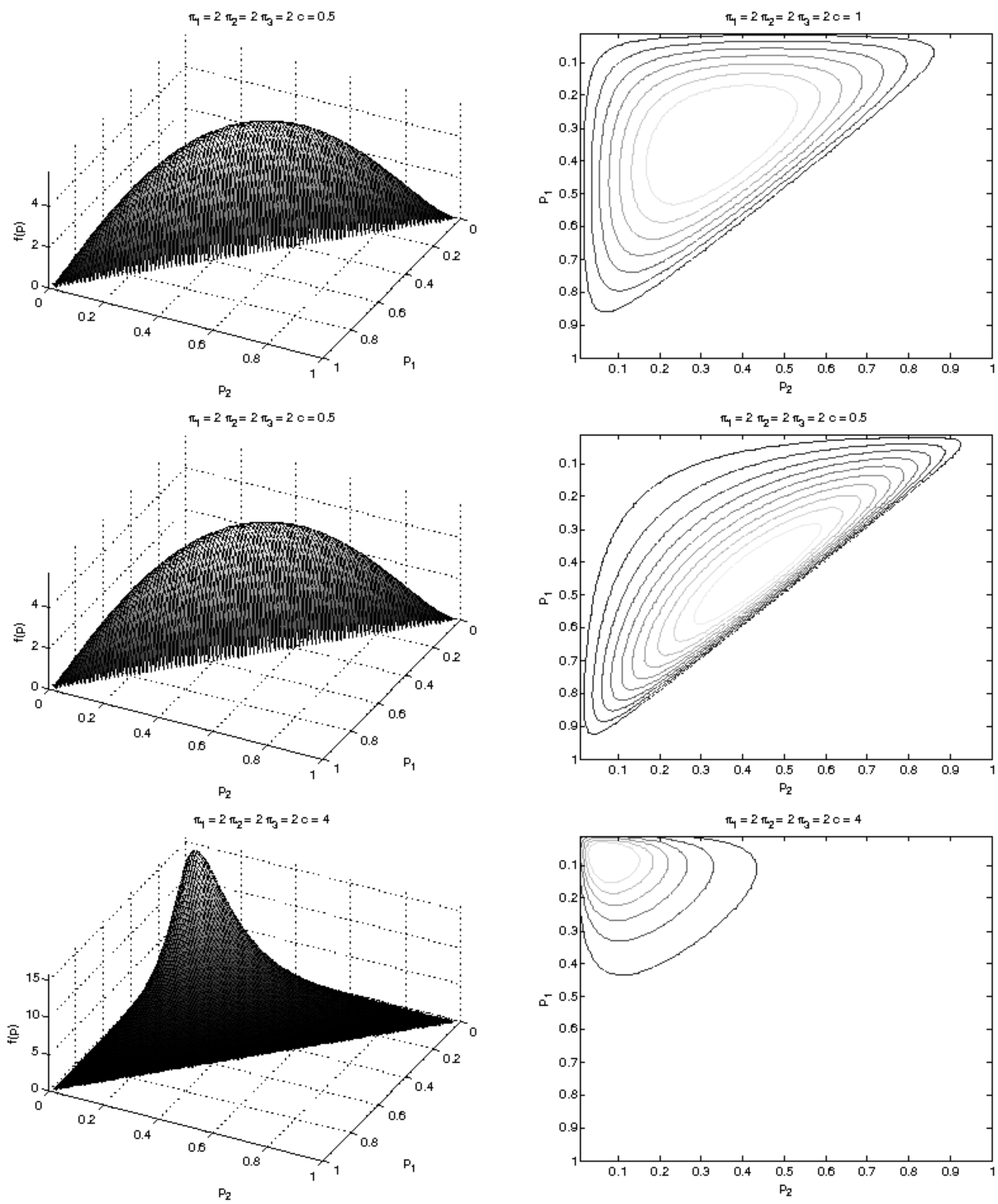
4.1.5 Shape Analysis

The parameters of the prior distribution can be estimated from historical data using maximum likelihood estimation, or determined from prior knowledge or expert judgement, see O' Hagan and Forster (2004).

Since $c = 1$ reduces the bivariate beta type III distribution in (4.1) to the bivariate beta type I distribution in (2.1), the results for the shape analysis of π_1, π_2 and π_3 are as presented in Section 2.1.5 and not repeated here.

Figure 4.2 shows the effect of c on the shape and concentration of the bivariate beta type III distribution. The reference case considered is where $\pi_1 = \pi_2 = \pi_3 = 2$ and $c = 1$; this is in fact the bivariate beta type I distribution. If c is decreased, the distribution shifts towards the line $p_1 + p_2 = 1$. If c is increased, the distribution shifts towards small values of P_1 and P_2 in the corner.

Figure 4.2: Bivariate Beta Type III Distribution: Changing c



4.2 Bayesian Estimation of the Shannon Entropy

4.2.1 Derivation

Theorem 4.1

The posterior distribution for the multinomial likelihood in (1.3) *and bivariate beta type III distribution in* (4.1) *is given by*

$$\begin{aligned} f(p_1, p_2|\mathbf{x}) &= \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}c^{(\pi_1+\pi_2+\pi_3)} \\ &\quad \times[\,{}_2F_1(\pi_1+\pi_2+\pi_3,\pi_3+x_3;\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;\frac{c-1}{c})]^{-1} \\ &\quad \times\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r \\ &\quad \times p_1^{\pi_1+x_1+r-s-1}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1} \end{aligned} \tag{4.4}$$

where $0 < p_i < 1$ *for* $i = 1, 2$, $0 < p_1 + p_2 < 1$, $\pi_1, \pi_2, \pi_3, c > 0$ *and* ${}_2F_1(\cdot)$ *is the Gauss hypergeometric function (see Appendix B, Definition 4).*

Proof

Using the binomial expansion, the last factor in (4.1) is

$$\begin{aligned} & [1+(c-1)p_1+(c-1)p_2]^{-(\pi_1+\pi_2+\pi_3)} \\ =\ & [1-(1-c)p_1-(1-c)p_2]^{-(\pi_1+\pi_2+\pi_3)} \\ =\ & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r p_1^{r-s}p_2^{s} \end{aligned} \tag{4.5}$$

which will converge if $|(c-1)p_1+(c-1)p_2| < 1$.

Using (4.5), the prior distribution can be written as

$$
\begin{aligned}
& f(p_1,p_2) \\
= \; & \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{\pi_1+\pi_2} \\
& \times \sum_{r=0}^{\infty}\sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r p_1^{\pi_1+r-s-1} p_2^{\pi_2+s-1}(1-p_1-p_2)^{\pi_3-1} \quad (4.6)
\end{aligned}
$$

This alternative form of the bivariate beta distribution is sometimes referred to as a mixture of bivariate beta type I distributions, see Cardeño et al. (2005). The numerator of the posterior distribution is given by

$$
\begin{aligned}
& f(\mathbf{x}|p_1,p_2)f(p_1,p_2) \\
= \; & \frac{n!}{x_1!x_2!x_3!}\frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{\pi_1+\pi_2} \\
& \times \sum_{r=0}^{\infty}\sum_{s=0}^{r} \left(\begin{array}{c} -(\pi_1+\pi_2+\pi_3) \\ \\ r \end{array}\right)\left(\begin{array}{c} r \\ \\ s \end{array}\right)(c-1)^r \\
& \times p_1^{\pi_1+x_1+r-s-1} p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1} \quad (4.7)
\end{aligned}
$$

The denominator of the posterior distribution is given by

$$
\begin{aligned}
&\int_0^1 \int_0^{1-p_2} f(\mathbf{x}|p_1,p_2) f(p_1,p_2) dp_1 dp_2 \\
= \quad & \frac{n!}{x_1!x_2!x_3!} \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{\pi_1+\pi_2} \sum_{r=0}^{\infty}\sum_{s=0}^{r} \begin{pmatrix} -(\pi_1+\pi_2+\pi_3) \\ r \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} (c-1)^r \\
& \times \int_0^1 \int_0^{1-p_2} p_1^{\pi_1+x_1+r-s-1} p_2^{\pi_2+x_2+s-1} (1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \qquad (4.8) \\
= \quad & \frac{n!}{x_1!x_2!x_3!} \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} c^{\pi_1+\pi_2} \\
& \times \sum_{r=0}^{\infty}\sum_{s=0}^{r} \begin{pmatrix} -(\pi_1+\pi_2+\pi_3) \\ r \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} (c-1)^r \\
& \times B(\pi_1+x_1+r-s, \pi_2+x_2+s, \pi_3+x_3) \qquad (4.9)
\end{aligned}
$$

since the integral in (4.8) corresponds to the total probability of a $BBeta^I(\pi_1+x_1+r-s, \pi_2+x_2+s, \pi_3+x_3)$ distribution.

The summation in (4.9) can be simplified as

$$
\begin{aligned}
&\sum_{r=0}^{\infty}\sum_{s=0}^{r}\left(\begin{array}{c}-(\pi_1+\pi_2+\pi_3)\\ r\end{array}\right)\left(\begin{array}{c}r\\ s\end{array}\right)(c-1)^r B(\pi_1+x_1+r-s,\pi_2+x_2+s,\pi_3+x_3)\\
=\;& \sum_{r=0}^{\infty}\left(\begin{array}{c}-(\pi_1+\pi_2+\pi_3)\\ r\end{array}\right)(c-1)^r\frac{\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+r+\pi_2+x_2+\pi_3+x_3)}\\
&\times\sum_{s=0}^{r}\left(\begin{array}{c}r\\ s\end{array}\right)\Gamma(\pi_1+x_1+r-s)\Gamma(\pi_2+x_2+s)\\
=\;& \sum_{r=0}^{\infty}\left(\begin{array}{c}-(\pi_1+\pi_2+\pi_3)\\ r\end{array}\right)(c-1)^r\frac{\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+r+\pi_2+x_2+\pi_3+x_3)}\\
&\times\frac{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_1+x_1+\pi_2+x_2+r)}{\Gamma(\pi_1+x_1+\pi_2+x_2)}\\
=\;& \frac{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2)}\\
&\times\sum_{r=0}^{\infty}\left(\begin{array}{c}-(\pi_1+\pi_2+\pi_3)\\ r\end{array}\right)(c-1)^r\frac{\Gamma(\pi_1+x_1+\pi_2+x_2+r)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)}\\
=\;& \frac{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2)}\frac{\Gamma(\pi_1+x_1+\pi_2+x_2)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}\\
&\times {}_2F_1(\pi_1+x_1+\pi_2+x_2,\pi_1+\pi_2+\pi_3;\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1-c) \qquad (4.10)
\end{aligned}
$$

Using equation (B.6), (4.10) reduces to

$$
\frac{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}c^{-(\pi_1+\pi_2+\pi_3)}\,{}_2F_1(\pi_1+\pi_2+\pi_3,\pi_3+x_3;\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;\frac{c-1}{c}) \qquad (4.11)
$$

Combining (4.9) and (4.11), the denominator is given by

$$\begin{aligned}
&\int\int f(\mathbf{x}|p_1,p_2)f(p_1,p_2)dp_1dp_2 \\
= \;& \frac{n!}{x_1!x_2!x_3!}\frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)}c^{\pi_1+\pi_2}\frac{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}c^{-(\pi_1+\pi_2+\pi_3)} \\
&\times {}_2F_1(\pi_1+\pi_2+\pi_3,\pi_3+x_3;\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;\frac{c-1}{c}) \qquad (4.12)
\end{aligned}$$

Using the numerator in (4.7) and the denominator in (4.12), the posterior distribution is given by

$$\begin{aligned}
&f(p_1,p_2|\mathbf{x}) \\
= \;& \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}c^{(\pi_1+\pi_2+\pi_3)} \\
&\times[\,{}_2F_1(\pi_1+\pi_2+\pi_3,\pi_3+x_3;\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;\frac{c-1}{c})]^{-1} \\
&\times\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r p_1^{\pi_1+x_1+r-s-1}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1},
\end{aligned}$$

as given in (4.4). ■

This posterior distribution is not the same form as the prior distribution, indicating that the bivariate beta type III distribution is not a natural conjugate for the multinomial distribution given in (1.3).

Theorem 4.2

The Bayesian estimator of the Shannon entropy under the squared error loss function using the bivariate beta type III distribution as prior for the multinomial model in (1.3) *is given by:*

$$\begin{aligned}\hat{H}_3^{III} &= -K\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r\frac{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)}{\Gamma(\gamma_1+\gamma_2+\gamma_3+1)}\\ &\quad\times\sum_{i=1}^{3}\gamma_i(\psi(\gamma_i+1)-\psi(\sum_{j=1}^{3}\gamma_j+1)) \qquad (4.13)\end{aligned}$$

where

$$\begin{aligned}K &= \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}c^{(\pi_1+\pi_2+\pi_3)}\\ &\quad\times[\,{}_2F_1(\pi_1+\pi_2+\pi_3,\pi_3+x_3;\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;\frac{c-1}{c})]^{-1}\end{aligned}$$

and

$$\begin{aligned}\gamma_1 &= \pi_1+x_1+r-s\\ \gamma_2 &= \pi_2+x_2+s\\ \gamma_3 &= \pi_3+x_3\end{aligned}$$

are the normalising coefficient and parameters of the posterior distribution in (4.4) *respectively.*

Proof

From (1.1) the Shannon entropy for the bivariate beta type III distribution is denoted by

$$H_3^{III} = -\sum_{i=1}^{3}p_i\ln p_i$$

The Bayesian estimator of the Shannon entropy under squared error loss is given by the expected value with respect to the posterior distribution obtained in (4.4). That is,

$$
\begin{aligned}
\hat{H}_3^{III} &= E_{f(p_1,p_2|\mathbf{x})}[H_3^{III}] \\
&= -\frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)} c^{(\pi_1+\pi_2+\pi_3)} \\
&\quad \times [\,{}_2F_1(\pi_1+\pi_2+\pi_3, \pi_3+x_3; \pi_1+x_1+\pi_2+x_2+\pi_3+x_3; \frac{c-1}{c})]^{-1} \\
&\quad \times \int_0^1 \int_0^{1-p_2} [p_1 \ln p_1 + p_2 \ln p_2 + (1-p_1-p_2)\ln(1-p_1-p_2)] \\
&\quad \times \sum_{r=0}^{\infty} \sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r} \binom{r}{s} (c-1)^r \\
&\quad \times p_1^{\pi_1+x_1+r-s-1} p_2^{\pi_2+x_2+s-1} (1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \\
&= K \sum_{i=1}^{3} I_i
\end{aligned}
$$

where

$$
\begin{aligned}
K &= \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)} c^{(\pi_1+\pi_2+\pi_3)} \\
&\quad \times [\,{}_2F_1(\pi_1+\pi_2+\pi_3, \pi_3+x_3; \pi_1+x_1+\pi_2+x_2+\pi_3+x_3; \frac{c-1}{c})]^{-1} \qquad (4.14)
\end{aligned}
$$

$$
\begin{aligned}
I_i &= \int_0^1 \int_0^{1-p_2} p_i \ln p_i \\
&\quad \times \sum_{r=0}^{\infty} \sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r} \binom{r}{s} (c-1)^r p_1^{\pi_1+x_1+r-s-1} p_2^{\pi_2+x_2+s-1} (1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2
\end{aligned}
$$

for $i=1,2$, and

$$I_3 = \int_0^1\int_0^{1-p_2}(1-p_1-p_2)\ln(1-p_1-p_2) \times \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r p_1^{\pi_1+x_1+r-s-1}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2$$

The simplification of I_i will only be shown for I_1, but follows similarly for I_2 and I_3.

$$\begin{aligned} & I_1 \\ = & \int_0^1\int_0^{1-p_2} p_1\ln p_1 \\ & \times\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r p_1^{\pi_1+x_1+r-s-1}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \\ = & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r \\ & \times\int_0^1\int_0^{1-p_2}(\ln p_1)p_1^{\pi_1+x_1+r-s}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \\ = & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r \\ & \times\int_0^1\int_0^{1-p_2}[\frac{\partial}{\partial\pi_1}p_1^{\pi_1+x_1+r-s}]p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \end{aligned}$$

since $\frac{d}{dx}a^x = a^x\ln a$. Changing the order of integration and differentiation:

$$\begin{aligned} I_1 = & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r \\ & \times\frac{\partial}{\partial\pi_1}\int_0^1\int_0^{1-p_2}p_1^{\pi_1+x_1+r-s}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \qquad (4.15) \\ = & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r\frac{\partial}{\partial\pi_1}\frac{\Gamma(\pi_1+x_1+r-s+1)\Gamma(\pi_2+x_2+s)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)} \end{aligned}$$

The integral in (4.15) is 1 because it corresponds to the total probability of a $BBeta^I(\pi_1+x_1+$

$r - s + 1, \pi_2 + x_2 + s, \pi_3 + x_3)$ distribution. Using the product and chain rules for differentiation:

$$\begin{aligned} I_1 &= \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r B(\pi_1+x_1+r-s+1, \pi_2+x_2+s, \pi_3+x_3) \\ &\quad \times[\psi(\pi_1+x_1+r-s+1) - \psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)] \end{aligned}$$

where $\psi(x)$ denotes the polygamma function. Similarly,

$$\begin{aligned} I_2 &= \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r B(\pi_1+x_1+r-s, \pi_2+x_2+s+1, \pi_3+x_3) \\ &\quad \times[\psi(\pi_2+x_2+s+1) - \psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)] \end{aligned}$$

and

$$\begin{aligned} I_3 &= \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r B(\pi_1+x_1+r-s, \pi_2+x_2+s, \pi_3+x_3+1) \\ &\quad \times[\psi(\pi_3+x_3+1) - \psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)] \end{aligned}$$

Adding I_1, I_2 and I_3 together

$$\begin{aligned} & \sum_{i=1}^{3} I_i \\ = & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(c-1)^r \frac{\Gamma(\pi_1+x_1+r-s)\Gamma(\pi_2+x_2+s)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)} \\ & \times[(\pi_1+x_1+r-s)\psi(\pi_1+x_1+r-s+1) + (\pi_2+x_2+s)\psi(\pi_2+x_2+s+1) \\ & +(\pi_3+x_3)\psi(\pi_3+x_3+1) \\ & -(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)] \end{aligned}$$

For simplification, let

$$\begin{aligned}
\gamma_1 &= \pi_1 + x_1 + r - s \\
\gamma_2 &= \pi_2 + x_2 + s \\
\gamma_3 &= \pi_3 + x_3
\end{aligned}$$

The Bayesian estimator of the Shannon entropy under squared error loss using the bivariate beta type III distribution as prior is then written as:

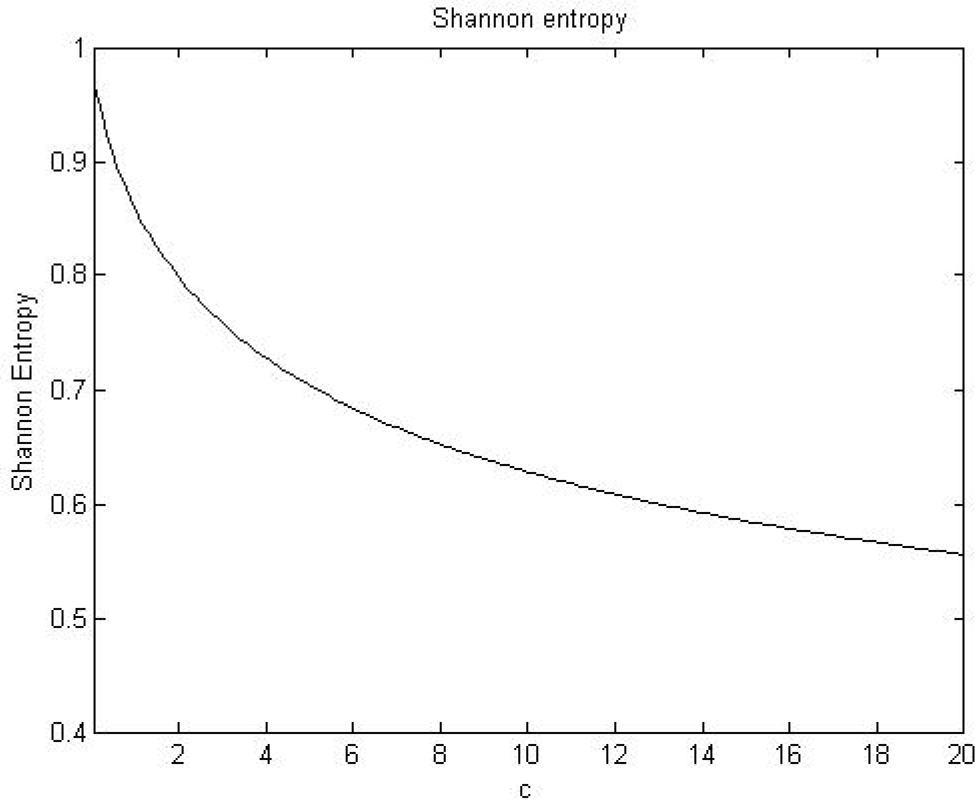
$$\begin{aligned}
\hat{H}_3^{III} &= -K \sum_{r=0}^{\infty} \sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r} \binom{r}{s} (c-1)^r \frac{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)}{\Gamma(\gamma_1+\gamma_2+\gamma_3+1)} \\
&\quad \times \sum_{i=1}^{3} \gamma_i (\psi(\gamma_i+1) - \psi(\sum_{j=1}^{3} \gamma_j + 1))
\end{aligned}$$

where K is the normalising coefficient in (4.14). ■

4.2.2 Numerical Analysis

Figure 4.3 plots the Bayesian estimates of Shannon entropy values for various values of c, with $\pi_1 = \pi_2 = \pi_3 = 2$ and multinomial frequencies $x_1 = 1$, $x_2 = 2$, $x_3 = 10$.

Figure 4.3: Bayesian Estimates of Shannon Entropy: Bivariate Beta Type III Prior



For larger values of c the Bayesian estimate of Shannon entropy decreases, indicating less uncertainty. This is consistent with the result from the shape analysis, where the concentration of the distribution around small values of P_1 and P_2 increases when c is increased, see Figure 4.2.

Recall from Section 4.1.4 that larger values of c also allowed for positive correlation. This suggests that in order to obtain a bivariate beta type III distribution that has positive correlation between P_1 and P_2, and not much uncertainty, c should be chosen to be large.

Chapter 5

Extended Bivariate Beta Type I Prior

5.1 The Extended Bivariate Beta Bype I Prior

5.1.1 Joint Density

Consider as a prior for the multinomial model in (1.3) the extended bivariate beta type I distribution studied by Ehlers and Bekker (2010), denoted as $BBeta^{E}(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c)$ and with density function:

$$\begin{aligned} f(p_1, p_2) &= \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} \beta_1^{-\pi_1} \beta_2^{-\pi_2} c^{\pi_1 + \pi_2} \\ &\quad \times p_1^{\pi_1 - 1} p_2^{\pi_2 - 1} (1 - p_1 - p_2)^{\pi_3 - 1} [1 - (1 - \frac{c}{\beta_1}) p_1 - (1 - \frac{c}{\beta_2}) p_2]^{-(\pi_1 + \pi_2 + \pi_3)} \end{aligned} \quad (5.1)$$

where $0 < p_1 + p_2 < 1$, $0 < p_i < 1$, for $i = 1, 2$ and $\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c > 0$. The inclusion of the additional parameters β_1 , β_2 and c add to the flexibility of the distribution, and also increases the range over which correlation between P_1 and P_2 can be positive.

If $\beta_1 = \beta_2 = 1$, then (5.1) reduces to the $BBeta^{III}(\pi_1, \pi_2, \pi_3, c)$ distribution in (4.1). In addition, if $c = 1$, (5.1) reduces to the $BBeta^{I}(\pi_1, \pi_2, \pi_3)$ distribution in (2.1).

5.1.2 Univariate Density

The univariate density functions of P_1 and P_2 can be found by integrating $f(p_1, p_2)$ over P_2 and P_1 respectively. Since $f(p_1, p_2)$ is symmetric, the univariate density functions will be of the same form. Only the derivation of the marginal distribution of P_1 is shown here, but it follows similarly P_2.

$$\begin{aligned} & f(p_1) \\ = \; & \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)}\beta_1^{-\pi_1}\beta_2^{-\pi_2}c^{\pi_1+\pi_2}p_1^{\pi_1-1} \\ & \times\int_0^{1-p_1} p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_2-1}[1-(1-\frac{c}{\beta_1})p_1-(1-\frac{c}{\beta_2})p_2]^{-(\pi_1+\pi_2+\pi_3)}dp_2 \end{aligned} \tag{5.2}$$

Using equation (B.2), the integral in (5.2) simplifies to

$$(1-p_1)^{\pi_2+\pi_3-1}[1-(1-\frac{c}{\beta_1})p_1]^{-(\pi_1+\pi_2+\pi_3)}\frac{\Gamma(\pi_2)\Gamma(\pi_3)}{\Gamma(\pi_2+\pi_3)}\,{}_2F_1(\pi_2,\pi_1+\pi_2+\pi_3;\pi_2+\pi_3,\frac{(1-\frac{c}{\beta_2})(1-p_1)}{1-(1-\frac{c}{\beta_1})p_1})$$

Let $z=\frac{(1-\frac{c}{\beta_2})(1-p_1)}{1-(1-\frac{c}{\beta_1})p_1}$, then it follows from relation (B.6) that

$$\begin{aligned} & {}_2F_1(\pi_2,\pi_1+\pi_2+\pi_3;\pi_2+\pi_3,\frac{(1-\frac{c}{\beta_2})(1-p_1)}{1-(1-\frac{c}{\beta_1})p_1}) \\ = \; & \frac{c}{\beta_2}\frac{1-(1-\frac{\beta_2}{\beta_1})p_1}{1-(1-\frac{c}{\beta_1})p_1}\,{}_2F_1(\pi_1+\pi_2+\pi_3,\pi_3;\pi_2+\pi_3;\frac{\beta_2}{c}\frac{(1-\frac{c}{\beta_2})(1-p_1)}{1-(1-\frac{\beta_2}{\beta_1})p_1}) \end{aligned}$$

Combining these results, the marginal density function of P_1 is given by

$$\begin{aligned} f(p_1) \; = \; & \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2+\pi_3)}\beta_1^{-\pi_1}\beta_2^{\pi_1+\pi_3}c^{-\pi_3}p_1^{\pi_1-1}(1-p_1)^{\pi_2+\pi_3-1}[1-(1-\frac{\beta_2}{\beta_1})p_1]^{-(\pi_1+\pi_2+\pi_3)} \\ & \times\,{}_2F_1(\pi_1+\pi_2+\pi_3,\pi_3;\pi_2+\pi_3;\frac{(1-\frac{\beta_2}{c})(1-p_1)}{1-(1-\frac{\beta_2}{\beta_1})p_1}) \end{aligned}$$

where $0 < p_1 < 1$ and $\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c > 0$. Similarly,

$$\begin{aligned} f(p_2) &= \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_2)\Gamma(\pi_1+\pi_3)} \beta_1^{\pi_2+\pi_3}\beta_2^{-\pi_2}c^{-\pi_2}p_2^{\pi_2-1}(1-p_2)^{\pi_1+\pi_3-1}[1-(1-\frac{\beta_1}{\beta_2})p_2]^{-(\pi_1+\pi_2+\pi_3)} \\ &\quad \times {}_2F_1(\pi_1+\pi_2+\pi_3, \pi_3; \pi_1+\pi_3; \frac{(1-\frac{\beta_1}{c})(1-p_2)}{1-(1-\frac{\beta_1}{\beta_2})p_2}) \end{aligned}$$

where $0 < p_2 < 1$ and $\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c > 0$.

5.1.3 Methods of Derivation

The extended bivariate beta type I distribution is derived using a transformation of three independently distributed χ^2 variables and the trivariate reduction method.

Let $S_i \sim \chi^2(2\pi_i)$ for $i = 1, 2, 3$ be three independently distributed χ^2 variables. That is,

$$f(s_i) = \frac{1}{2^{\pi_i}\Gamma(\pi_i)} \exp(-\frac{s_i}{2}) s_i^{\pi_i - 1}$$

for $s_i > 0$. The joint density function of these variables is given by

$$\begin{aligned} f(s_1, s_2, s_3) &= \prod_{i=1}^{3} f(s_i) \\ &= \frac{1}{2^{\sum_{i=1}^{3}\pi_i}\prod_{i=1}^{3}\Gamma(\pi_i)} \exp(-\frac{1}{2}\sum_{i=1}^{3} s_i) \prod_{i=1}^{3} s_i^{\pi_i - 1} \end{aligned}$$

Consider the transformation $P_i = \frac{\beta_i S_i}{\beta_1 S_1 + \beta_2 S_2 + c S_3}$ for $i = 1, 2$ and $P_3 = \beta_1 S_1 + \beta_2 S_2 + c S_3$. Then the inverse transformation is

$$\begin{aligned} S_1 &= \frac{1}{\beta_1} P_1 P_3 \\ S_2 &= \frac{1}{\beta_2} P_2 P_3 \\ S_3 &= \frac{1}{c} P_3 (1 - P_1 - P_2) \end{aligned}$$

The Jacobian for this transformation is given by

$$
\begin{aligned}
J &= J((s_1,s_2,s_3)\rightarrow(p_1,p_2,p_3)) \\
&= \begin{vmatrix} \frac{\partial s_1}{\partial p_1} & \frac{\partial s_1}{\partial p_2} & \frac{\partial s_1}{\partial p_3} \\ \frac{\partial s_2}{\partial p_1} & \frac{\partial s_2}{\partial p_2} & \frac{\partial s_2}{\partial p_3} \\ \frac{\partial s_3}{\partial p_1} & \frac{\partial s_3}{\partial p_2} & \frac{\partial s_3}{\partial p_3} \end{vmatrix} \\
&= \begin{vmatrix} \frac{p_3}{\beta_1} & 0 & \frac{p_1}{\beta_1} \\ 0 & \frac{p_3}{\beta_2} & \frac{p_2}{\beta_2} \\ -\frac{p_3}{c} & -\frac{p_3}{c} & \frac{1-p_1-p_2}{c} \end{vmatrix} \\
&= \frac{p_3}{\beta_1}[\frac{p_3}{\beta_2}\frac{(1-p_1-p_2)}{c}+\frac{p_2}{\beta_2}\frac{p_3}{c}]+\frac{p_1}{\beta_1}\frac{p_3^2}{\beta_2 c} \\
&= \frac{p_3^2}{\beta_1\beta_2 c}
\end{aligned}
$$

The joint density function of P_1, P_2 and P_3 is given by

$$
\begin{aligned}
f(p_1,p_2,p_3) &= f(s_1,s_2,s_3)|J| \\
&= \frac{1}{2^{\sum_{i=1}^3\pi_i}\prod_{i=1}^3\Gamma(\pi_i)}(\frac{p_1p_3}{\beta_1})^{\pi_1-1}(\frac{p_2p_3}{\beta_2})^{\pi_2-1}(\frac{p_3(1-p_1-p_2)}{c})^{\pi_3-1} \\
&\quad \times\exp[-\frac{1}{2}(\frac{p_1p_3}{\beta_1}+\frac{p_2p_3}{\beta_2}+\frac{p_3(1-p_1-p_2)}{c})]\times\frac{p_3^2}{\beta_1\beta_2 c} \\
&= \frac{1}{2^{\sum_{i=1}^3\pi_i}\prod_{i=1}^3\Gamma(\pi_i)}\beta_1^{-\pi_1}\beta_2^{-\pi_2}c^{-\pi_3}p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}p_3^{\sum_{i=1}^3\pi_i-1} \\
&\quad \times\exp[-\frac{1}{2\beta_1\beta_2 c}(\beta_2cp_1p_3+\beta_1cp_2p_3+\beta_1\beta_2p_3(1-p_1-p_2))] \\
&= \frac{1}{2^{\sum_{i=1}^3\pi_i}\prod_{i=1}^3\Gamma(\pi_i)}\beta_1^{-\pi_1}\beta_2^{-\pi_2}c^{-\pi_3}p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}p_3^{\sum_{i=1}^3\pi_i-1} \\
&\quad \times\exp[-\frac{p_3}{2c}(1+(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2)]
\end{aligned}
$$

The joint density function of P_1 and P_2 is obtained when integrating $f(p_1, p_2, p_3)$ over P_3.

$$\begin{aligned} f(p_1,p_2) &= \int_0^\infty f(p_1,p_2,p_3)dp_3 \\ &= \frac{1}{2^{\sum_{i=1}^3 \pi_i} \prod_{i=1}^3 \Gamma(\pi_i)} \beta_1^{-\pi_1}\beta_2^{-\pi_2}c^{-\pi_3}p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1} \\ &\quad \times \int_0^\infty p_3^{\sum_{i=1}^3 \pi_i -1} \exp[-\frac{p_3}{2c}(1+(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2)]dp_3 \end{aligned} \quad (5.3)$$

The integral in (5.3) can be written as

$$\int_0^\infty p_3^{\sum_{i=1}^3 \pi_i -1} \exp[-\frac{1}{2c}p_3(1+(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2)]dp_3$$

$$\begin{aligned} &= \frac{\Gamma(\sum_{i=1}^3 \pi_i)}{\{\frac{1}{2c}[1+(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2]\}^{\sum_{i=1}^3 \pi_i}} \\ &\quad \times \int_0^\infty \frac{\{\frac{1}{2c}[1+(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2]\}^{\sum_{i=1}^3 \pi_i}}{\Gamma(\sum_{i=1}^3 \pi_i)} p_3^{\sum_{i=1}^3 \pi_i -1} \\ &\quad \times \exp[-\frac{p_3}{2c}(1+(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2)]dp_3 \end{aligned} \quad (5.4)$$

$$= \frac{\Gamma(\sum_{i=1}^3 \pi_i)}{\{\frac{1}{2c}[1+(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2]\}^{\sum_{i=1}^3 \pi_i}}$$

where the integral in (5.4) is 1 since it corresponds to the total probability of a $Gamma(\sum_{i=1}^3 \pi_i, \frac{1}{2c}[1+(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2])$ distribution. Combining the results,

$$\begin{aligned} & f(p_1,p_2) \\ &= \frac{1}{2^{\sum_{i=1}^3 \pi_i} \prod_{i=1}^3 \Gamma(\pi_i)} \beta_1^{-\pi_1}\beta_2^{-\pi_2}c^{-\pi_3}p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1} \\ &\quad \times \frac{\Gamma(\sum_{i=1}^3 \pi_i)}{\{\frac{1}{2c}[1-(1-\frac{c}{\beta_1})p_1-(1-\frac{c}{\beta_2})p_2]\}^{\sum_{i=1}^3 \pi_i}} \\ &= \frac{\Gamma(\sum_{i=1}^3 \pi_i)}{\prod_{i=1}^3 \Gamma(\pi_i)} \beta_1^{-\pi_1}\beta_2^{-\pi_2}c^{\pi_1+\pi_2}p_1^{\pi_1-1}p_2^{\pi_2-1}(1-p_1-p_2)^{\pi_3-1}[1-(1-\frac{c}{\beta_1})p_1-(1-\frac{c}{\beta_2})p_2]^{-\sum_{i=1}^3 \pi_i} \end{aligned}$$

and it follows that P_1 and P_2 have the joint density function of a $BBeta^E(\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c)$

distribution, as defined in (5.1).

5.1.4 Correlation

The product moments of P_1 and P_2 are derived as

$$\begin{aligned}
& E[P_1^i P_2^j] \\
= & \int_0^1 \int_0^{1-p_2} p_1^i p_2^j f(p_1,p_2) dp_1 dp_2 \\
= & \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} \beta_1^{-\pi_1}\beta_2^{-\pi_2} c^{\pi_1+\pi_2} \\
& \times \int_0^1 \int_0^{1-p_2} p_1^{\pi_1+i-1} p_2^{\pi_2+j-1} (1-p_1-p_2)^{\pi_3-1} [1-(1-\frac{c}{\beta_1})p_1 - (1-\frac{c}{\beta_2})p_2]^{-(\pi_1+\pi_2+\pi_3)} dp_1 dp_2
\end{aligned}$$

Using equation (B.4), the integral above is proportional to a hypergeometric function of two variables. That is,

$$\begin{aligned}
& E[P_1^i P_2^j] \\
= & \frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} \beta_1^{-\pi_1}\beta_2^{-\pi_2} c^{\pi_1+\pi_2} \frac{\Gamma(\pi_1+i)\Gamma(\pi_2+j)\Gamma(\pi_3)}{\Gamma(\pi_1+\pi_2+\pi_3+i+j)} \\
& \times F_1(\pi_1+\pi_2+\pi_3, \pi_1+i, \pi_2+j, \pi_1+\pi_2+\pi_3+i+j; 1-\frac{c}{\beta_1}, 1-\frac{c}{\beta_2}) \\
= & \frac{\Gamma(\pi_1+\pi_2+\pi_3)\Gamma(\pi_1+i)\Gamma(\pi_2+j)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_1+\pi_2+\pi_3+i+j)} \beta_1^{-\pi_1}\beta_2^{-\pi_2} c^{\pi_1+\pi_2} \\
& \times F_1(\pi_1+\pi_2+\pi_3, \pi_1+i, \pi_2+j, \pi_1+\pi_2+\pi_3+i+j; 1-\frac{c}{\beta_1}, 1-\frac{c}{\beta_2})
\end{aligned} \tag{5.5}$$

where $F_1(\cdot)$ denotes the hypergeometric function of two variables.

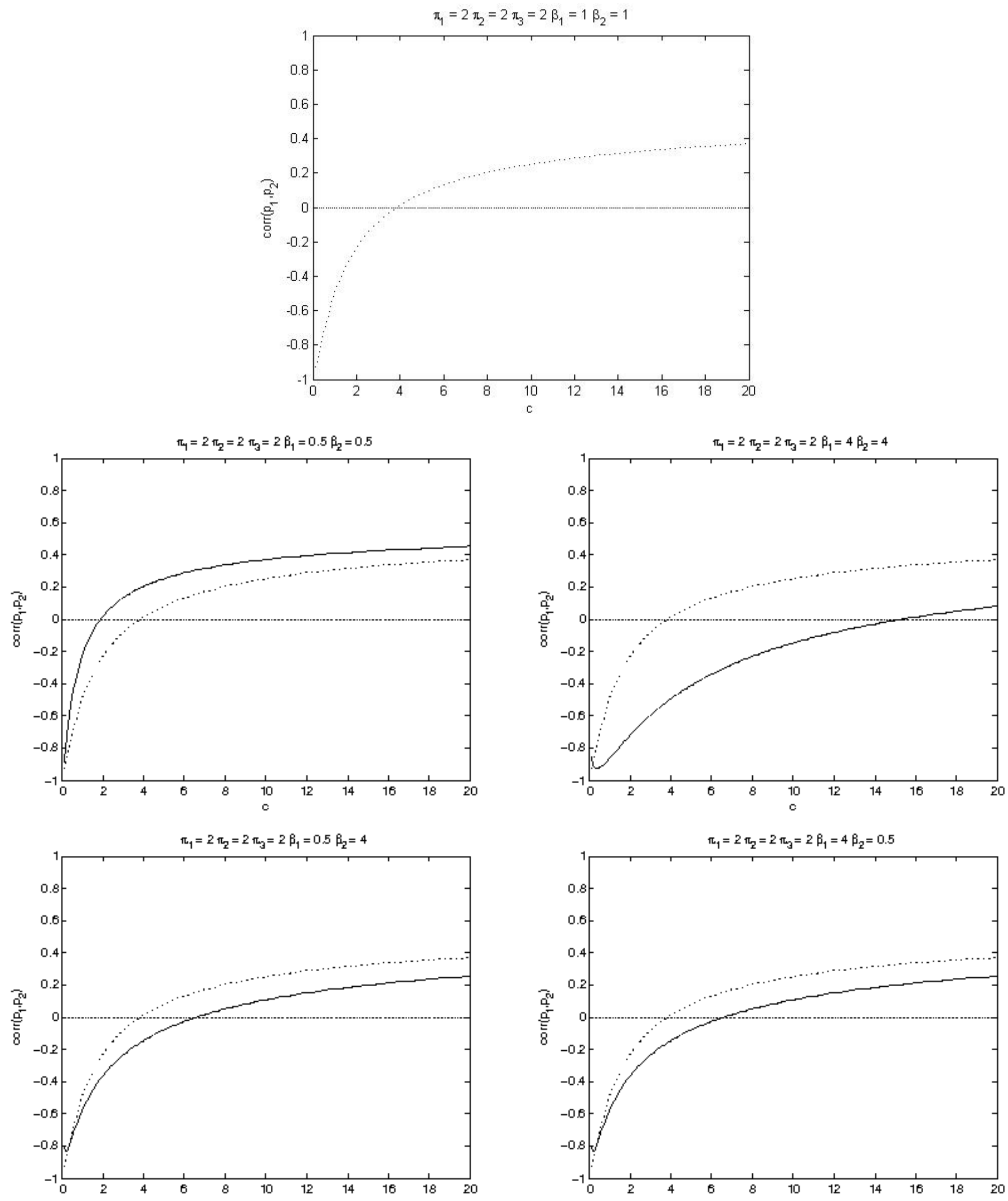
Figure 5.1 shows the correlation between P_1 and P_2 for various values of β_1, β_2 and c. Note that if $\beta_1 = \beta_2 = 1$, the extended bivariate beta distribution reduces to the bivariate beta type III distribution and the correlation, shown as the dotted curve, is therefore the same as seen in the previous chapter.

If β_1 and β_2 are both less than 1, the range of positive correlation increases. In this example, for

$\beta_1 = \beta_2 = 0.5$, positive correlation is attained from $c = 2$ onwards. If β_1 and β_2 are both large, the range of positive correlation decreases. In this example, for $\beta_1 = \beta_2 = 4$, positive correlation is attained only from $c = 15$ onwards.

The last set of graphs in Figure 5.1 shows that for $\pi_1 = \pi_2$, the correlation function is symmetric if $\pi_1 = \pi_2$. If the values of β_1 and β_2 are switched, the same function is obtained, this can also be seen in (5.5).

Figure 5.1: Extended Bivariate Beta Type I Distribution: Correlation



5.1.5 Shape Analysis

The graphs that follow show the effect of the parameters on the shape and concentration of the distribution. The reference case considered is where $\pi_1 = \pi_2 = \pi_3 = 2$ and $\beta_1 = \beta_2 = c = 1$. Using these parameters, the extended bivariate beta type I distribution in (5.1) reduces to the bivariate beta type I distribution in (2.1) with parameters $\pi_1 = \pi_2 = \pi_3 = 2$. For the impact of π_1, π_2 and π_3 on the shape and concentration of this reference case distribution, refer to Section 2.1.5.

Figure 5.2 shows the effect of β_1 on the shape and concentration of the distribution. If β_1 is decreased, the distribution shifts towards the marginal distribution of p_2 on the right axis. If β_1 is increased, the distribution shifts towards large values of P_1 and small values of P_2 along the line $p_1 + p_2 = 1$.

Figure 5.2: Extended Bivariate Beta Type I Distribution: Changing β_1

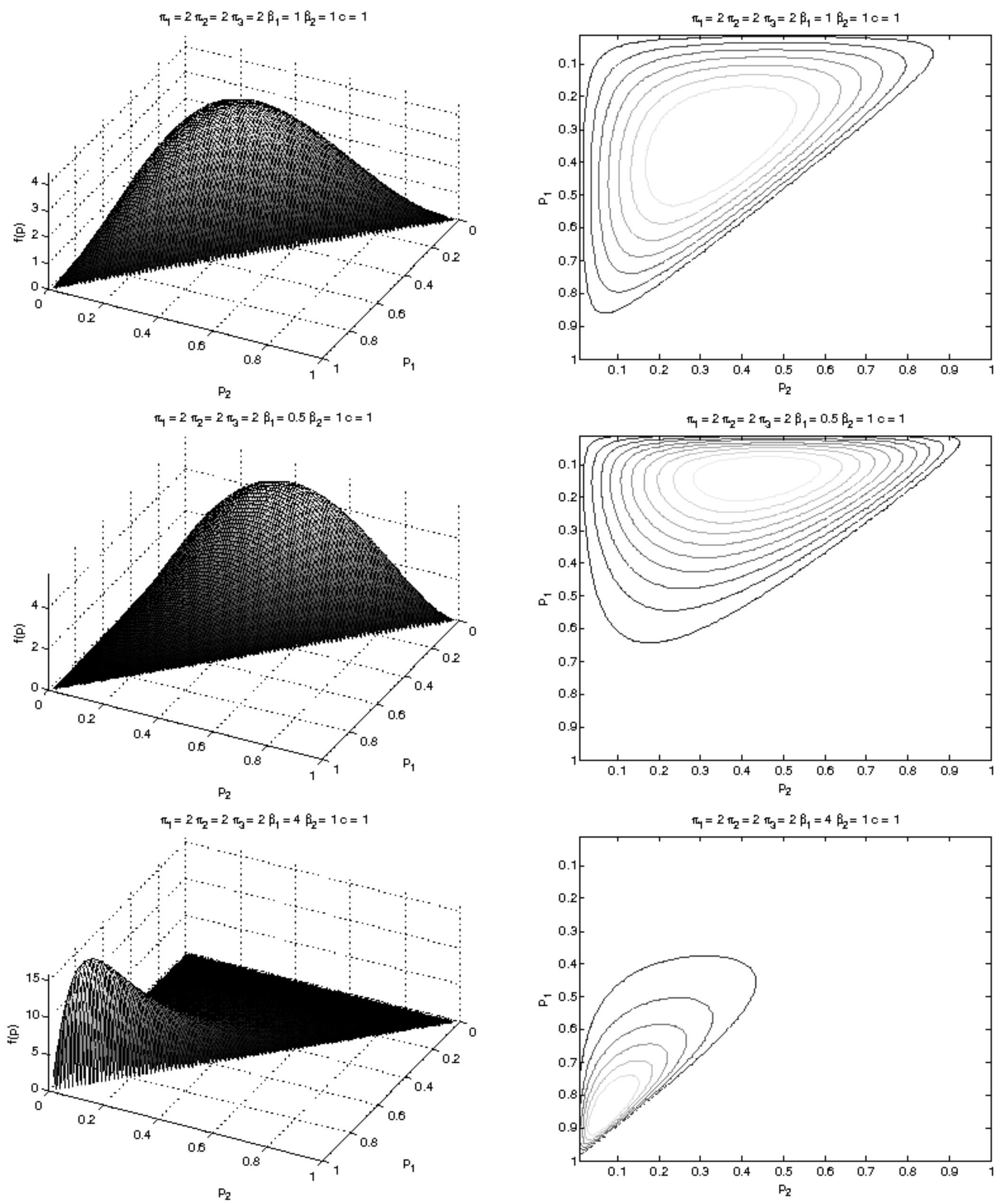


Figure 5.3 shows the effect of β_2 on the shape and concentration of the distribution. If β_2 is decreased, the distribution shifts towards the marginal distribution of P_1 on the left axis. If β_2 is increased, the distribution shifts towards large values of P_2 and small values of P_1 along the line $p_1 + p_2 = 1$.

Figure 5.3: Extended Bivariate Beta Type I Distribution: Changing β_2

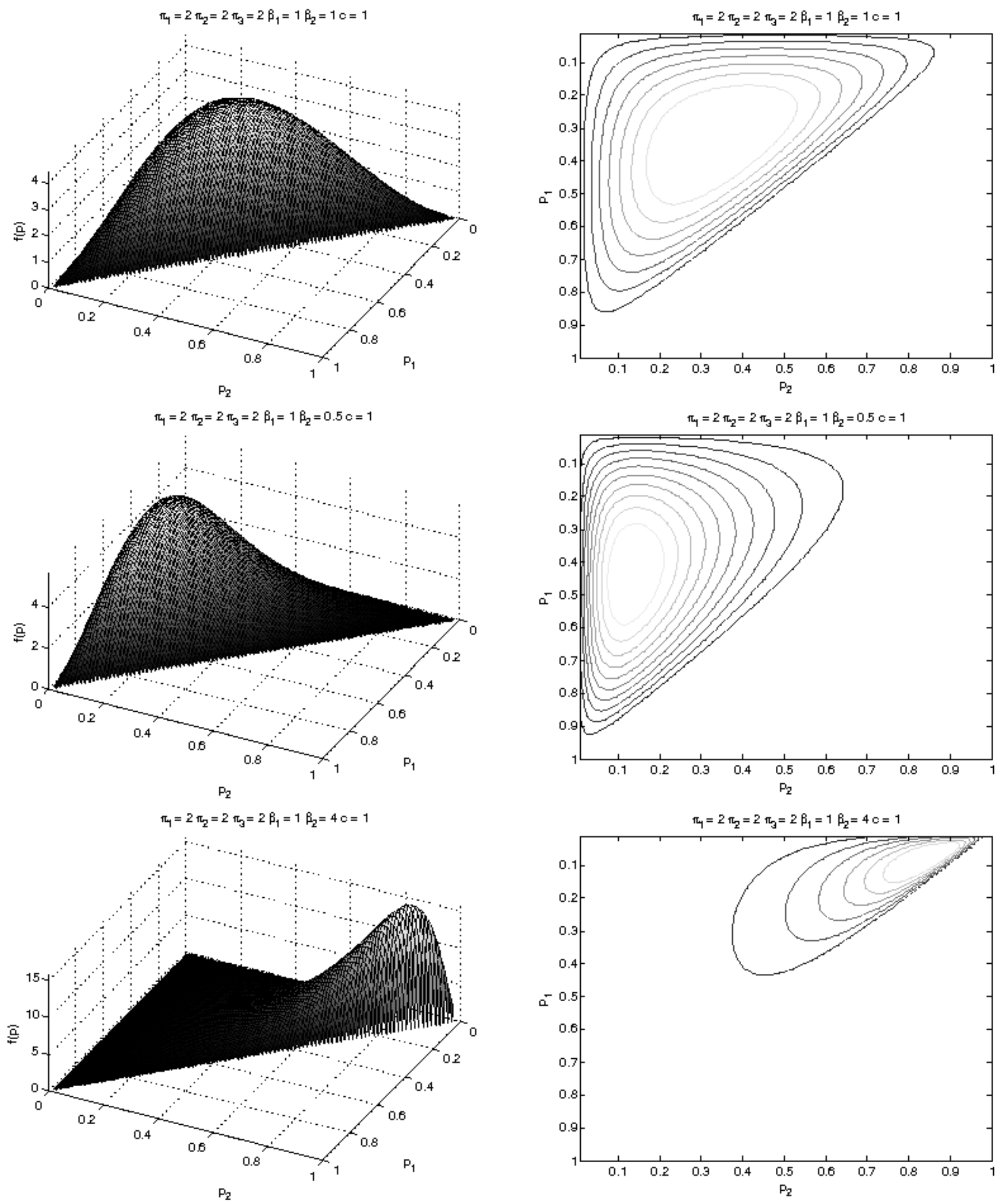


Figure 5.4 shows the effect of simultaneously changing $\beta_1$and β_2 on the shape and concentration of the distribution. If β_1 and β_2 are decreased, the distribution shifts towards small values of P_1 and P_2 in the corner. If β_1 and β_2 are increased, the distribution shifts symmetrically towards the line $p_1 + p_2 = 1$.

Figure 5.4: Extended Bivariate Beta Type I Distribution: Changing β_1 and β_2

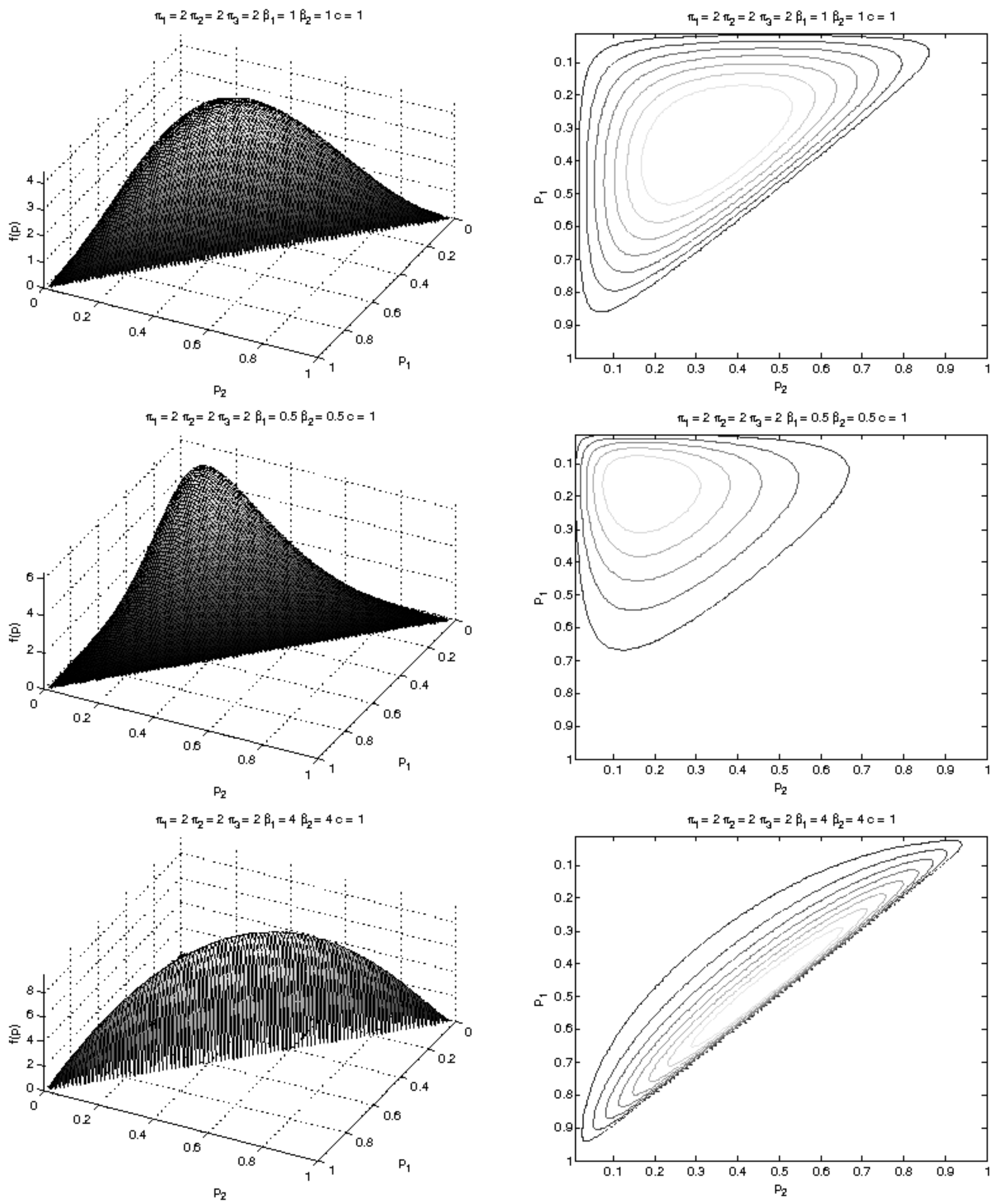
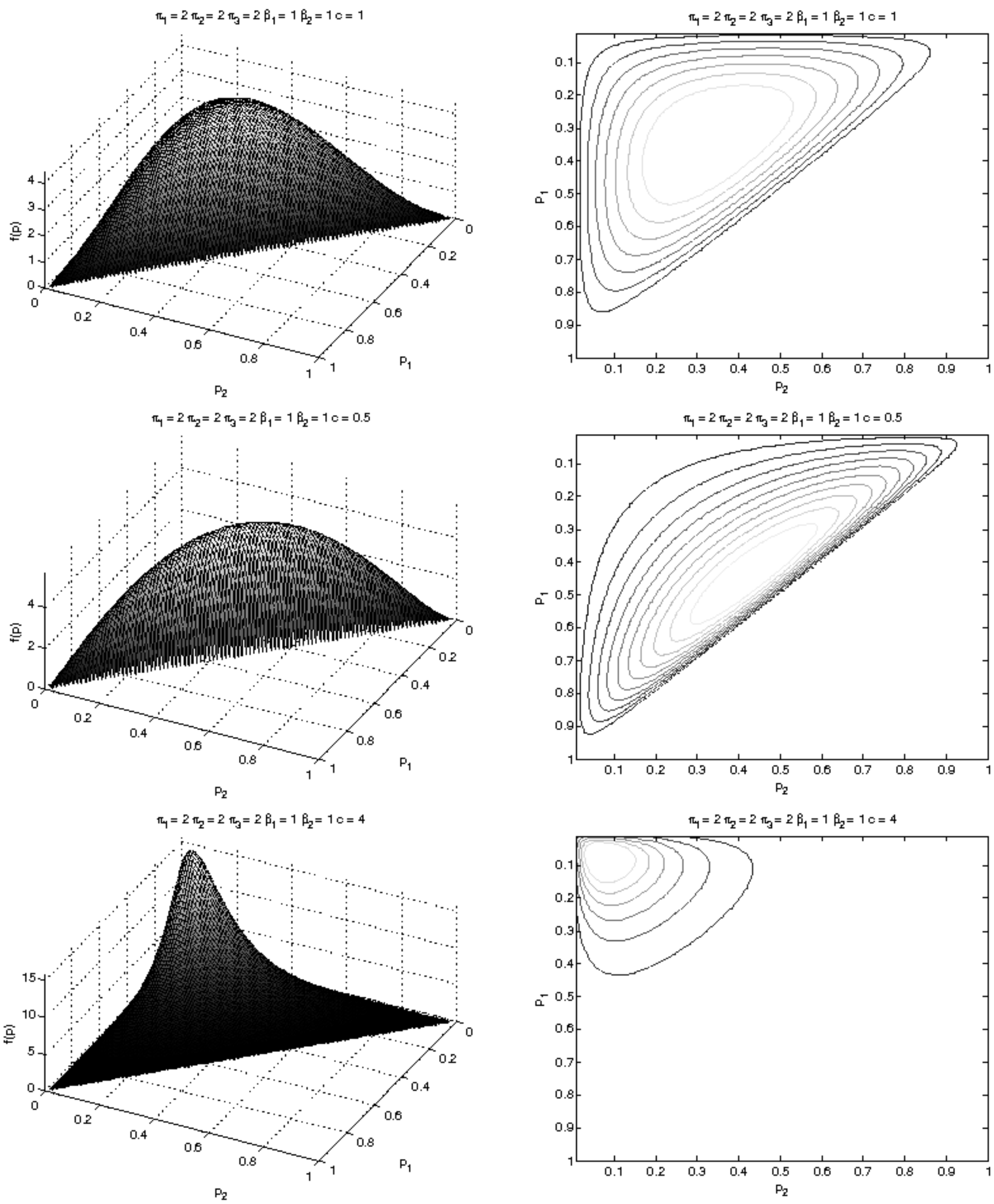


Figure 5.5 shows the effect of c on the shape and concentration of the distribution. If c is decreased, the distribution shifts towards the line $p_1 + p_2 = 1$. If c is increased, the distribution shifts towards small values of P_1 and P_2 in the corner. Note that for $\beta_1 = \beta_2 = 1$, this is the exact same behaviour as the bivariate beta type III distribution.

Figure 5.5: Extended Bivariate Beta Type I Distribution: Changing c



5.2 Bayesian Estimation of Shannon Entropy

5.2.1 Derivation

Theorem 5.1

The posterior distribution for the multinomial likelihood in (1.3) *and extended bivariate beta type I prior in* (5.1) *is given by*

$$\begin{aligned} & f(p_1, p_2|\mathbf{x}) \\ = \; & \frac{\Gamma(\pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3)}{\Gamma(\pi_1 + x_1)\Gamma(\pi_2 + x_2)\Gamma(\pi_3 + x_3)} \\ & \times [F_1(\pi_1 + \pi_2 + \pi_3, \pi_1 + x_1, \pi_2 + x_2, \pi_1 + x_1 + \pi_2 + x_2 + \pi_3 + x_3; 1 - \frac{c}{\beta_1}, 1 - \frac{c}{\beta_2})]^{-1} \\ & \times p_1^{\pi_1 + x_1 - 1} p_2^{\pi_2 + x_2 - 1} (1 - p_1 - p_2)^{\pi_3 + x_3 - 1} [1 - (1 - \frac{c}{\beta_1})p_1 - (1 - \frac{c}{\beta_2})p_2]^{-(\pi_1 + \pi_2 + \pi_3)} \end{aligned} \tag{5.6}$$

where $0 < p_i < 1$ *for* $i = 1, 2$, $0 < p_1 + p_2 < 1$, $\pi_1, \pi_2, \pi_3, \beta_1, \beta_2, c > 0$ *and* $F_1(\cdot)$ *is the hypergeometric function of two variables.*

Proof

The numerator of the posterior distribution is given by:

$$\begin{aligned} & f(p_1, p_2) f(\mathbf{x}|p_1, p_2) \\ = \; & \frac{n!}{x_1! x_2! x_3!} \frac{\Gamma(\pi_1 + \pi_2 + \pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)} \beta_1^{-\pi_1} \beta_2^{-\pi_2} c^{\pi_1 + \pi_2} \\ & \times p_1^{\pi_1 + x_1 - 1} p_2^{\pi_2 + x_2 - 1} (1 - p_1 - p_2)^{\pi_3 + x_3 - 1} [1 - (1 - \frac{c}{\beta_1}) - (1 - \frac{c}{\beta_2})]^{-(\pi_1 + \pi_2 + \pi_3)} \end{aligned} \tag{5.7}$$

The denominator of the posterior distribution is given by:

$$\begin{aligned}
& \int_0^1 \int_0^{1-p_2} f(p_1,p_2)f(\mathbf{x}|p_1,p_2)dp_1dp_2 \\
= & \frac{n!}{x_1!x_2!x_3!}\frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)}\beta_1^{-\pi_1}\beta_2^{-\pi_2}c^{\pi_1+\pi_2} \\
& \times \int_0^1 \int_0^{1-p_2} p_1^{\pi_1+x_1-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1} \\
& \times[1-(1-\frac{c}{\beta_1})-(1-\frac{c}{\beta_2})]^{-(\pi_1+\pi_2+\pi_3)}dp_1dp_2 \qquad (5.8)
\end{aligned}$$

Using equation (B.4), the integral in (5.8) simplifies to

$$\frac{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}F_1(\pi_1+\pi_2+\pi_3,\pi_1+x_1,\pi_2+x_2,\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1-\frac{c}{\beta_1},1-\frac{c}{\beta_2})$$

The denominator of the posterior distribution becomes

$$\begin{aligned}
& \int_0^1 \int_0^{1-p_2} f(p_1,p_2)f(\mathbf{x}|p_1,p_2)dp_1dp_2 \\
= & \frac{n!}{x_1!x_2!x_3!}\frac{\Gamma(\pi_1+\pi_2+\pi_3)}{\Gamma(\pi_1)\Gamma(\pi_2)\Gamma(\pi_3)}\beta_1^{-\pi_1}\beta_2^{-\pi_2}c^{\pi_1+\pi_2}\frac{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)} \\
& \times F_1(\pi_1+\pi_2+\pi_3,\pi_1+x_1,\pi_2+x_2,\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1-\frac{c}{\beta_1},1-\frac{c}{\beta_2}) \qquad (5.9)
\end{aligned}$$

Combining the numerator in (5.7) and the denominator in (5.9), the posterior distribution is given by

$$\begin{aligned}
& f(p_1,p_2|\mathbf{x}) \\
= & \frac{f(p_1,p_2)f(\mathbf{x}|p_1,p_2)}{\int\int f(p_1,p_2)f(\mathbf{x}|p_1,p_2)dp_1dp_2} \\
= & \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)} \\
& \times[F_1(\pi_1+\pi_2+\pi_3,\pi_1+x_1,\pi_2+x_2,\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1-\frac{c}{\beta_1},1-\frac{c}{\beta_2})]^{-1} \\
& \times p_1^{\pi_1+x_1-1}p_2^{\pi_2+x_2-1}(1-p_1-p_2)^{\pi_3+x_3-1}[1-(1-\frac{c}{\beta_1})p_1-(1-\frac{c}{\beta_2})p_2]^{-(\pi_1+\pi_2+\pi_3)}
\end{aligned}$$

as given in (5.6). ∎

This posterior distribution is not the same form as the prior distribution, indicating that the extended bivariate beta type I distribution is not a natural conjugate for the multinomial distribution. Using the binomial expansion, the last term in the kernel in (5.6) can be written as

$$[1-(1-\frac{c}{\beta_1})p_1-(1-\frac{c}{\beta_2})p_2]^{-(\pi_1+\pi_2+\pi_3)} = \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s}p_1^{r-s}p_2^{s}$$

which will converge if $|(\frac{c}{\beta_1}-1)p_1+(\frac{c}{\beta_2}-1)p_2|<1$, and the posterior distribution can be written as

$$\begin{aligned}
& f(p_1,p_2|\mathbf{x}) \\
= \quad & \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)} \\
& \times[F_1(\pi_1+\pi_2+\pi_3,\pi_1+x_1,\pi_2+x_2,\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1-\frac{c}{\beta_1},1-\frac{c}{\beta_2})]^{-1} \\
& \times\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
& \times p_1^{\pi_1+x_1+r-s-1}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1} \qquad (5.10)
\end{aligned}$$

Theorem 5.2

The Bayesian estimator of the Shannon entropy under squared error loss using the extended bivariate beta type I distribution as a prior for the multinomial model in (1.3) *is given by:*

$$\begin{aligned}\hat{H}_3^E &= -K\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s}\frac{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)}{\Gamma(\gamma_1+\gamma_2+\gamma_3+1)} \\ &\quad \times\sum_{i=1}^{3}\gamma_i(\psi(\gamma_i+1)-\psi(\sum_{j=1}^{3}\gamma_j+1)) \qquad (5.11)\end{aligned}$$

where

$$\begin{aligned}K &= \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)} \\ &\quad \times[F_1(\pi_1+\pi_2+\pi_3,\pi_1+x_1,\pi_2+x_2,\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1-\frac{c}{\beta_1},1-\frac{c}{\beta_2})]^{-1}\end{aligned}$$

is the normalising coefficient, and

$$\begin{aligned}\gamma_1 &= \pi_1+x_1+r-s \\ \gamma_2 &= \pi_2+x_2+s \\ \gamma_3 &= \pi_3+x_3\end{aligned}$$

denote the parameters of the posterior distribution.

Proof

Denote the Shannon entropy for the extended bivariate beta type I distribution prior as

$$H_3^E = -\sum_{i=1}^{3}p_i\ln p_i$$

The Bayesian estimator of the Shannon entropy under squared error loss is given by its expected value with respect to the posterior distribution given in (5.6). That is,

$$
\begin{aligned}
\hat{H}_3^E &= E_{f(p_1,p_2|\mathbf{x})}[H_3^E] \\
&= -\int_0^1\int_0^{1-p_2}[p_1\ln p_1+p_2\ln p_2+(1-p_1-p_2)\ln(1-p_1-p_2) \\
&\quad \times\frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)} \\
&\quad \times[F_1(\pi_1+\pi_2+\pi_3,\pi_1+x_1,\pi_2+x_2,\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1-\frac{c}{\beta_1},1-\frac{c}{\beta_2})]^{-1} \\
&\quad \times\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times p_1^{\pi_1+x_1+r-s-1}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2 \\
&= K\sum_{i=1}^{3}I_i
\end{aligned}
$$

where

$$
\begin{aligned}
K &= \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)} \\
&\quad \times[F_1(\pi_1+\pi_2+\pi_3,\pi_1+x_1,\pi_2+x_2,\pi_1+x_1+\pi_2+x_2+\pi_3+x_3;1-\frac{c}{\beta_1},1-\frac{c}{\beta_2})]^{-1}
\end{aligned}
$$

$$
\begin{aligned}
I_i &= \int_0^1\int_0^{1-p_2}p_i\ln p_i\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times p_1^{\pi_1+x_1+r-s-1}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2
\end{aligned}
$$

for $i=1,2,$ and

$$
\begin{aligned}
I_3 &= \int_0^1\int_0^{1-p_2}(1-p_1-p_2)\ln(1-p_1-p_2)\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times p_1^{\pi_1+x_1+r-s-1}p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1}dp_1dp_2
\end{aligned}
$$

The simplification of I_i will only be shown for I_1, but follows similarly for I_2 and I_3.

$$\begin{aligned}
I_1 &= \int_0^1 \int_0^{1-p_2} p_1 \ln p_1 \sum_{r=0}^{\infty}\sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times p_1^{\pi_1+x_1+r-s-1} p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \\
&= \sum_{r=0}^{\infty}\sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times \int_0^1 \int_0^{1-p_2} (\ln p_1) p_1^{\pi_1+x_1+r-s} p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \\
&= \sum_{r=0}^{\infty}\sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times \int_0^1 \int_0^{1-p_2} [\frac{\partial}{\partial \pi_1} p_1^{\pi_1+x_1+r-s}] p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2
\end{aligned}$$

since $\frac{d}{dx}a^x = a^x \ln a$. Changing the order of integration and differentiation:

$$\begin{aligned}
I_1 &= \sum_{r=0}^{\infty}\sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times \frac{\partial}{\partial \pi_1} \int_0^1 \int_0^{1-p_2} p_1^{\pi_1+x_1+r-s} p_2^{\pi_2+x_2+s-1}(1-p_1-p_2)^{\pi_3+x_3-1} dp_1 dp_2 \qquad (5.12) \\
&= \sum_{r=0}^{\infty}\sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times \frac{\partial}{\partial \pi_1} \frac{\Gamma(\pi_1+x_1+r-s+1)\Gamma(\pi_2+x_2+s)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)}
\end{aligned}$$

The integral in (5.12) is proportional to the total probability of a $BBeta^I(\pi_1+x_1+r-s+1, \pi_2+x_2+s, \pi_3+x_3)$ distribution. Using the product and chain rules for differentiation

$$\begin{aligned}
I_1 &= \sum_{r=0}^{\infty}\sum_{s=0}^{r} \binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
&\quad \times \frac{\Gamma(\pi_1+x_1+r-s+1)\Gamma(\pi_2+x_2+s)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)} \\
&\quad \times [\psi(\pi_1+x_1+r-s+1) - \psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)]
\end{aligned}$$

Similarly

$$
\begin{aligned}
I_2 = & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
& \times\frac{\Gamma(\pi_1+x_1+r-s)\Gamma(\pi_2+x_2+s+1)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)} \\
& \times[\psi(\pi_2+x_2+s+1)-\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)]
\end{aligned}
$$

and

$$
\begin{aligned}
I_3 = & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
& \times\frac{\Gamma(\pi_1+x_1+r-s)\Gamma(\pi_2+x_2+s)\Gamma(\pi_3+x_3+1)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)} \\
& \times[\psi(\pi_3+x_3+1)-\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)]
\end{aligned}
$$

Adding I_1, I_2 and I_3 together,

$$
\begin{aligned}
\sum_{i=1}^{3} I_i = & \sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s} \\
& \times\frac{\Gamma(\pi_1+x_1+r-s)\Gamma(\pi_2+x_2+s)\Gamma(\pi_3+x_3)}{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)} \\
& \times[(\pi_1+x_1+r-s)\psi(\pi_1+x_1+r-s+1)+(\pi_2+x_2+s)\psi(\pi_2+x_2+s+1) \\
& +(\pi_3+x_3)\psi(\pi_3+x_3+1) \\
& -(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r)\psi(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3+r+1)]
\end{aligned}
$$

Denote the parameters of the posterior distribution by γ_i, that is,

$$
\begin{aligned}
\gamma_1 &= \pi_1+x_1+r-s \\
\gamma_2 &= \pi_2+x_2+s \\
\gamma_3 &= \pi_3+x_3
\end{aligned}
$$

The Bayesian estimator of the Shannon entropy under squared error loss using the extended bivariate beta type I prior distribution is derived as:

$$\begin{aligned}\hat{H}_3^E &= -K\sum_{r=0}^{\infty}\sum_{s=0}^{r}\binom{-(\pi_1+\pi_2+\pi_3)}{r}\binom{r}{s}(\frac{c}{\beta_1}-1)^{r-s}(\frac{c}{\beta_2}-1)^{s}\frac{\Gamma(\gamma_1)\Gamma(\gamma_2)\Gamma(\gamma_3)}{\Gamma(\gamma_1+\gamma_2+\gamma_3+1)}\\ &\quad\times\sum_{i=1}^{3}\gamma_i(\psi(\gamma_i+1)-\psi(\sum_{j=1}^{3}\gamma_j+1))\end{aligned}$$

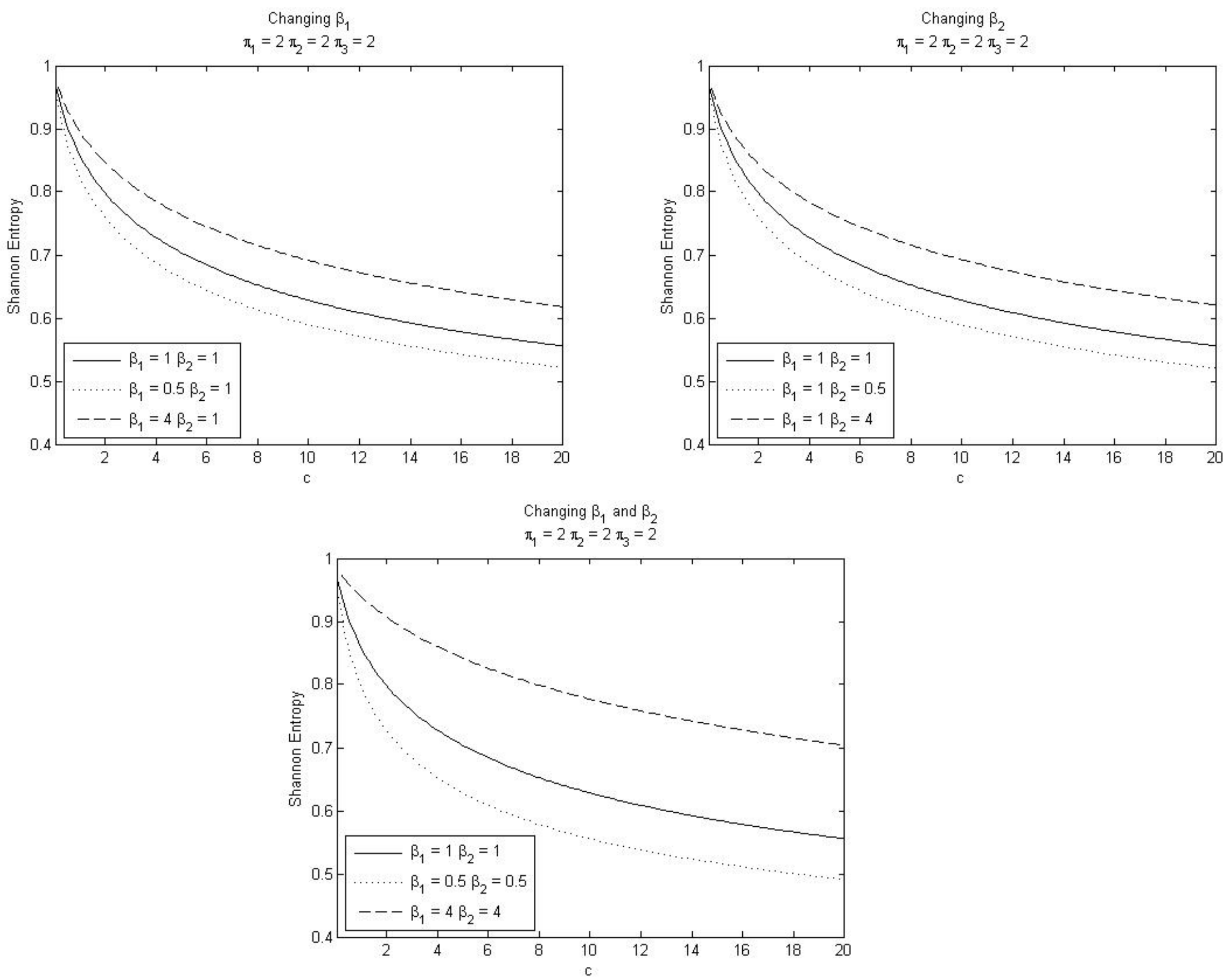
where $K = \frac{\Gamma(\pi_1+x_1+\pi_2+x_2+\pi_3+x_3)}{\Gamma(\pi_1+x_1)\Gamma(\pi_2+x_2)\Gamma(\pi_3+x_3)}[F_1(\pi_1+\pi_2+\pi_3, \pi_1+x_1, \pi_2+x_2, \pi_1+x_1+\pi_2+x_2+\pi_3+x_3; 1-\frac{c}{\beta_1}, 1-\frac{c}{\beta_2})]^{-1}$ is the normalising coefficient. ■

Note that if the prior parameters are $\beta_1 = \beta_2 = 1$ in (5.1) and using (B.5), $\hat{H}_3^E$ in (5.11) reduces to $\hat{H}_3^{III}$ in (4.13).

5.2.2 Numerical Analysis

Figure 5.6 plots the Shannon entropy values for various values of β_1, β_2 and c, with $\pi_1 = \pi_2 = \pi_3 = 2$ and multinomial frequencies $x_1 = 1$, $x_2 = 2$, $x_3 = 10$.

Figure 5.6: Bayesian Estimates of Shannon Entropy: Extended Bivariate Beta Type I Prior



As seen for the bivariate beta type III distribution, $\hat{H}_3^E$ decreases for larger values of c, indicating less uncertainty, see Section 4.2.2.

In the first two charts of Figure 5.6 it can be seen that if β_1 and β_2 are decreased respectively, the absolute level of $\hat{H}_3^E$ decreases, indicating less uncertainty. Similarly, increasing β_1 and β_2 respectively increases the absolute level of $\hat{H}_3^E$, indicating more uncertainty. The magnitude and shape of the $\hat{H}_3^E$ curves are the same for both cases, confirming the symmetry of the extended bivariate beta type I distribution.

These results are in contrast to the results from the shape analyses in Figures 5.2 and 5.3, where lower values of β_1 and β_2 are associated with less concentrated distributions, or more uncertainty; and higher values of β_1 and β_2 are associated with more concentrated distributions, or less uncertainty.

Note that in the first two charts, if $c = 1$ and $\beta_1 = \beta_2 = 1$, the extended bivariate beta type I distribution reduces to the bivariate beta type I distribution, and $\hat{H}_3^E = \hat{H}_3^I$. That is, if $\pi_1 = \pi_2 = \pi_3 = 2$, $\beta_1 = \beta_2 = 1$ (the solid curve) and $c = 1$, then, reading off the charts, $\hat{H}_3^E \approx 0.8$ and $\hat{H}_3^E \approx 0.97$. These values correspond to those provided in Table 2.1 for the bivariate beta type I distribution.

In the third chart, decreasing β_1 and β_2 simultaneously decreases the absolute level of $\hat{H}_3^E$, indicating the less uncertainty. Increasing β_1 and β_2 simultaneously increases the absolute level of $\hat{H}_3^E$, indicating more uncertainty. The directions of these shifts are the same as when changing β_1 and β_2 respectively, although the magnitude is larger, confirming the combined effect of the changes. In the shape analysis conducted in Figure 5.4, lower values of β_1 and β_2 results in a larger concentration around small values of P_1 and P_2, and higher values of β_1 and β_2 are associated with a larger concentration towards the line $p_1 + p_2 = 1$.

The difference between these intuitive and counterintuitive results observed for different combinations of β_1 and β_2 parameters may be disturbing at first, but can be explained by the location of the distribution. In the first two cases, if only β_1 or β_2 are changed, the distribution either shifts towards the marginal distribution of P_2 or P_1 respectively, or towards a specific point along the line $p_1 + p_2 = 1$, see Figures 5.2. and 5.3. Then, if β_1 and β_2 are changed simultaneously, the distribution either moves towards small values of P_1 and P_2 or towards the line $p_1 + p_2 = 1$, but is not concentrated at a specific point along this line.

This suggests that $\hat{H}_3^E$ provides information about both the concentration and location of the distribution. Lower values of $\hat{H}_3^E$ indicate a larger concentration along small values of P_1 and P_2. Larger values of $\hat{H}_3^E$ indicate lower concentration along small values of P_1 and P_2, but a larger concentration elsewhere in the feasible region, with $\hat{H}_3^E$ increasing as the concentration moves away from small value of $P_1$and P_2.

Again, it should be kept in mind that since $\hat{H}_3^E$ is the *Bayesian* estimate of the Shannon entropy, it also contains information about the likelihood function. Therefore, the multinomial frequencies will also have an effect on the behaviour of $\hat{H}_3^E$.

Chapter 6

Application to Credit Risk

6.1 Concepts

Banks are concerned with defaults occuring. A default event is defined by the Basel II captial framework (2006) as the occurence of either the bank considering the customer unlikely to repay its debt obligations in full, or the customer being past due more than 90 days on any material credit obligation.

It is thus vital to distinguish two key concepts: default rate and probability of default (PD). The default rate corresponds to the number of customers who have defaulted out of a particular population of customers, i.e. the actual observed rate at which customers default. The probability of default corresponds to the likelihood of a particular customer defaulting. In the calibration of credit risk models, the default rate is used to determine the probability of default of a particular customer.

6.2 Calibration: An Overview

Many credit analysts consider the probability of default as the most important driver for the calculation of regulatory capital. As explained in the introduction, when data is readily available it is relatively easy to estimate the probability of default. This is typically done using logistic regression.

In the environment of small low-default portfolios, it is almost impossible to construct meaningful logistic regression-type models to directly predict the likelihood of default, since adequate default information is generally not available. Instead, the probability of default of a customer is obtained through an indirect method. Customers are assigned a credit rating (irrespective of whether they defaulted or not), based on some regression model, and a probability of default is assigned to that specific rating through a model calibration process. It is intuitive that the likelihood of default of a customer is influenced, amongst others, by the state of the economic cycle, and a substantial amount of research in this field take this into account. There are various types of calibration methodologies, although most of these do not use an explicit Bayesian approach.

The simplest of these approaches is to use a moment matching approach, whereby a PD curve is fitted to the credit ratings in the calibration sample, such that the required moment of the portfolio PD is equal to the moment of the long-run observed default rate. The moment in question is typically the simple average, although it is the view of many credit risk analysts that taking the long-run average default rate without considering the economic conditions may not be an accurate representation of the risk in a portfolio, see Trück and Rachev (2005).

Sometimes expert judgement is combined with the moment matching approach, where PDs and PD bands are expertly assigned to rating classes, in particular for the very low-default investment grade rating classes. The risk here is that the PDs and the relative bands may appear to be correct, while in fact they are not, see Pluto and Tashce (2005).

Schuermann and Hanson (2004) use a cohort and duration approach to estimate PDs from transition matrices, with the focus on the last column of the transition matrix (i.e. the "default" column) . They find that the investment grade PDs do not differ much between different economic conditions, but that speculative grade PDs are much more sensitive to changes in the economic cycle.

Trück and Rachev (2005) propose the use of the entire transition matrix and a bootstrapping method to determine confidence sets for the PDs for different rating classes over different points in the economic cycle. In their bootstrap method they consider the occurrence of default within each rating class to be binomially distributed, but do not consider the distribution between rating classes.

The "most prudent estimation" methodology is contributed by Pluto and Tasche (2005), where they use upper confidence bounds to obtain PD estimates to any desired degree of conservatism, based on the assumption that PDs are monotonic between rating classes (which is generally true).

A new trend is to estimate the PD curve based on the discriminatory power of the underlying rating model, measured by the receiver operating characteristic (ROC) and cumulative accuracy profiles (CAP). This has been investigated by Van der Burgth (2008) and Tasche (2010).

The calibration methodologies discussed thus far are not explicit Bayesian calibration methodologies, which is most likely due to the fact that the majority of credit scoring and calibration methodologies used in practice follow a frequentist approach. However, the Bayesian estimation of credit risk, both the underlying credit rating model as well as the model calibration, appears to occur more and more often.

As part of the credit rating model development, Löffler et al. (2005) propose a Bayesian methodology where they use as prior information the coefficients from credit rating models from other data sets. They find that "Bayesian estimators are significantly more accurate than the straight logit estimator".

Gössl (2005) considers the development of a credit portfolio model using a Bayesian approach, and proposes the use of the joint distribution of PDs and systemic correlation between the assets in a portfolio, as opposed to the use of their point estimators.

Finally, the starting point of this analysis is obtained from Kiefer (2008). He considers the binomial distribution as an indication of the likelihood of a default or non-default event in a portfolio, and uses a univariate beta type I distribution as a prior for the binomial distribution. The parameters of his beta distribution were obtained by eliciting information from an expert, where the expert provided his/her opinion of the values to which the quantiles of the beta distribution should correspond to. Kiefer then used the method of moments to determine the beta distribution that satisfies the expert's opinion. The univariate beta type I distribution was in turn used for calibration and to obtain confidence intervals for the PDs associated with the portfolio.

As an alternative to the method of moments for estimation of the prior parameters, Kiefer considered the maximum entropy prior selection method. This approach maximises the entropy in the distribution (i.e. it maximises the uncertainty), but otherwise provides as little information as possible.

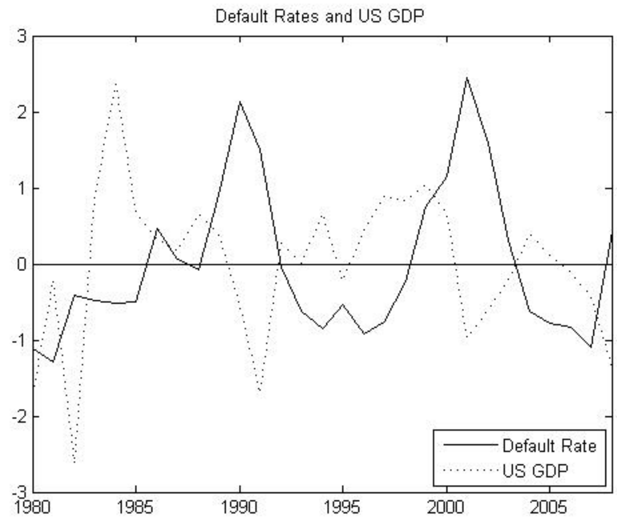
The analysis in this study differs from Kiefer's approach in two ways. Firstly, this study considers the Bayesian estimation of Shannon entropy, whereas Kiefer used the maximum entropy approach, which is a numerical optimisation of the entropy function. Secondly, the maximum entropy approach aims to *maximise* the entropy, whereas using the Bayesian estimation of the Shannon entropy allows the statistician to adjust the entropy level when selecting the prior distribution parameters. This enables the statistician to consider the Shannon entropy in conjunction with other measures.

6.3 Setting

The aim of this analysis is to illustrate how the Shannon entropy can be used as a tool to select the prior distribution as part of a Bayesian credit risk model calibration approach. Whilst Kiefer only considered the univariate situation, this analysis goes a step further and considers the bivariate case. Following suit that probabilities of default are influenced by the economic cycle, the differentiation between favourable and adverse economic conditions is considered.

The default rate distribution differs between good and bad economic conditions, which is an aspect thoroughly investigated for macroeconomic stress testing purposes. Figure 6.1 below illustrates the standardised default rates and US GDP for the last 29 years. The default rates in this graph are taken from the Moody's Annual Corporate Default Rate Study (2009), and represents the default rate out of their total rated ("All Rated") population for each year. The US GDP values are the seasonally adjusted year-on-year GDP growth rates obtained from the US Bureau of Economic Analysis (BEA). The values are standardized in order to ease graphical interpretation.

Figure 6.1: Default Rates Over Time



It is clear that in favourable economic conditions, when a high GDP (dotted line) prevails, default rates (solid line) are low. Similarly, adverse economic conditions are associated with high default rates. Throughout this analysis the GDP growth rate is assumed to represent the economic conditions. In practice other factors also indicate economic conditions, but for simplicity only one factor is used.

The default rates of the overall portfolio ("All Rated") can be divided into 21 rating classes, from AAA to C, but for the purpose of this analysis only two segments will be considered, investment grade and speculative grade. Investment grade represents all ratings between AAA and BBB-, and speculative grade represents all ratings between BB+ and C. Consider the following three events:

- Event 1: Default occurs in investment grade
- Event 2: Default occurs in speculative grade
- Event 3: Default does not occur

If a default occurs, we are concerned with its rating quality. A company that defaults in the investment grade class is more likely to have a larger exposure and therefore more significant impact on the bank's book than a company defaulting in the speculative grade class. If a default does not occur, it is not really of interest since this is what we would expect from the counterparty.

It is assumed that these three events follow the multinomial model in (1.3), and that the parameters of this model follow a bivariate beta distribution. Given the practical importance of where defaults occur, it is clear that the correct estimation of the bivariate beta distribution is very important.

6.4 Data

The default rate data used for this analysis is obtained from the Moody's "Corporate Defaults and Recovery Rates, 1920-2008" study (2009), and spans from 1930 to 2008. The first 10 years are not used, but the rest of the data is used as-is, without making any assumptions regarding the quality of the data. In order to distinguish between good and bad economic conditions, two samples are selected: a "Good" sample and a "Bad" sample. The "Good" sample consists of years where the GDP growth rate is larger than the 60th percentile of the GDP distribution spanning the same period. The "Bad" sample consists of years where the GDP growth rate is less than the 40th percentile of the GDP distribution spanning the same period. Observations with GDP growth between the 40th and 60th percentiles are not used in this analysis in order to clearly distinguish the differences between favourable and adverse economic conditions. In practice, it is recommended to use all available data. Both datasets consist of 32 observations each.

For each sample, the investment and speculative grade default rates are used. The bivariate beta distributions investigated in this study will be considered as priors to the joint distribution of the investment and speculative grade default events, as described above.

Table 6.1 lists the values of the default rates in both samples. Note that there are quite a few years in which no defaults occurred. Theoretically this violates the assumption of the bivariate beta distributions that $p_i > 0$ for $i = 1, 2$. However, it is not believed that one should calibrate to a default rate of 0. This implies another advantage of considering the bivariate beta distributions as priors, in that the distributions will be able to provide non-zero calibrated probabilities of default.

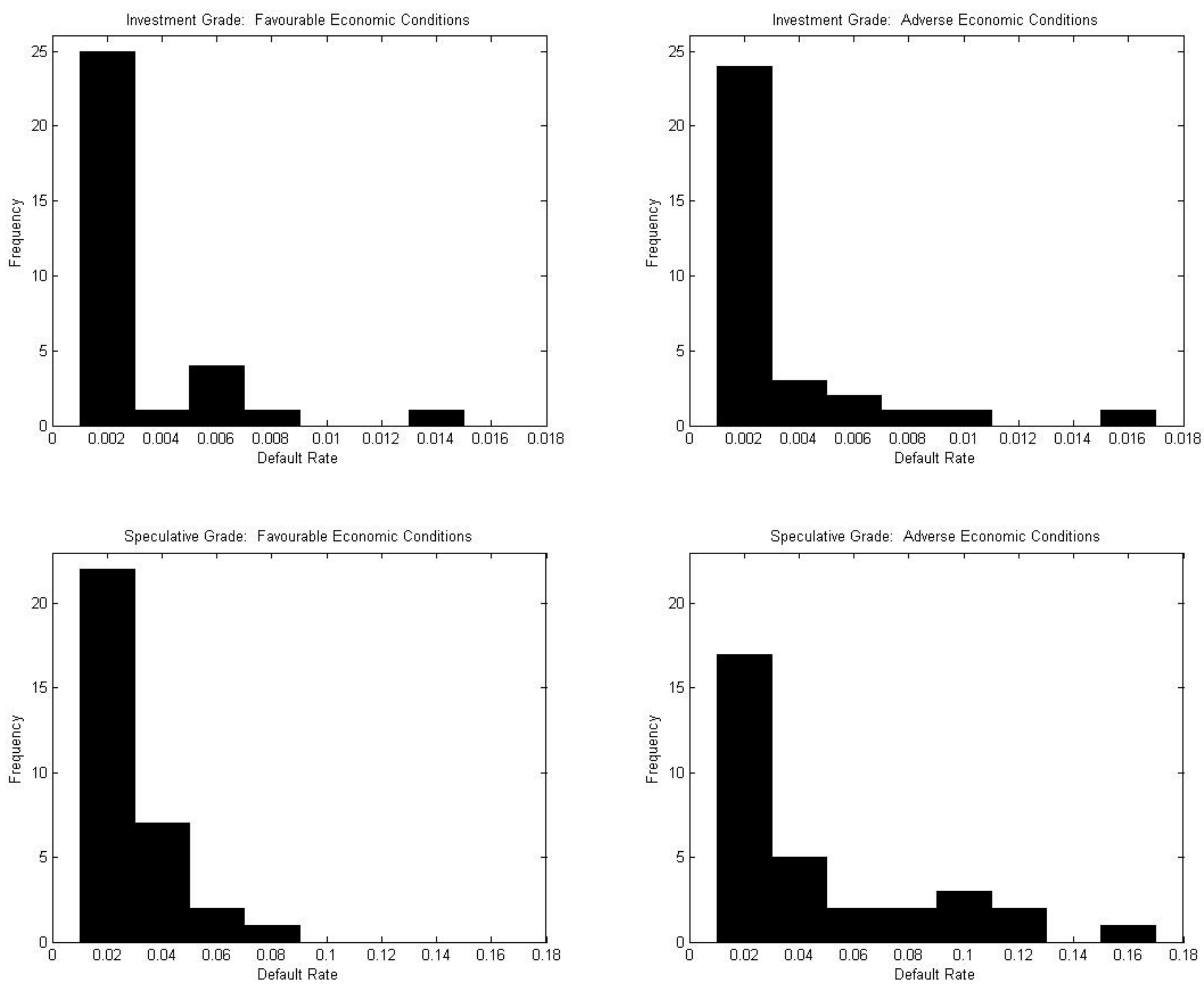
Table 6.1: Moody's Default Rates (1930-2008)

Favourable Economic Conditions				Adverse Economic Conditions		
Year	Investment Grade	Speculative Grade		Year	Investment Grade	Speculative Grade
1934	0.00578	0.05929		1930	0.00159	0.02131
1935	0.01253	0.0609		1931	0.0049	0.07845
1936	0.00465	0.02736		1932	0.0078	0.10811
1937	0.00661	0.02595		1933	0.00806	0.15391
1939	0.00402	0.01751		1938	0.01579	0.02593
1940	0.00572	0.02606		1945	0	0.00525
1941	0	0.01698		1946	0	0
1942	0	0.0075		1947	0	0.00314
1943	0	0.00615		1949	0	0.01901
1944	0	0.00679		1954	0	0.00471
1948	0	0		1956	0	0
1950	0	0		1957	0	0.00452
1951	0	0.0045		1958	0	0
1953	0	0		1960	0	0.00737
1955	0	0.00505		1961	0	0.0107
1959	0	0		1967	0	0
1962	0	0.01463		1970	0.00271	0.08772
1963	0	0.01156		1974	0	0.01330
1964	0	0		1975	0	0.01735
1965	0	0		1980	0	0.01613
1966	0	0.00415		1981	0	0.00701
1968	0	0.00387		1982	0.00212	0.03571
1972	0	0.01957		1990	0	0.09976
1973	0.00231	0.01271		1991	0.00065	0.0937
1976	0	0.00864		1993	0	0.03072
1977	0.00109	0.01339		1995	0	0.0292
1978	0	0.01798		2001	0.00132	0.10124
1983	0	0.03824		2002	0.00507	0.07921
1984	0.00096	0.03333		2003	0	0.05123
1997	0	0.02028		2006	0	0.01688
1998	0.00038	0.03152		2007	0	0.00918
1999	0.00036	0.05384		2008	0.003	0.04129

6.5 Analysis

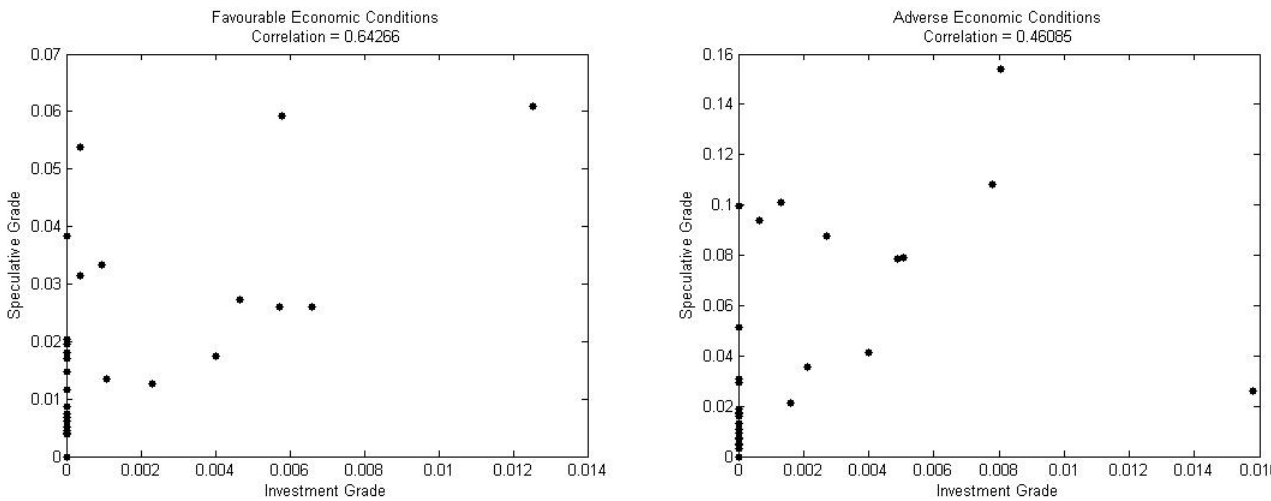
Figure 6.1 has already shown the trends between the time series for default rates and economic conditions for the overall Moody's rated population. Figure 6.2 below compares the univariate distributions of the default rates for each of the categories between favourable and adverse economic times. Adverse economic conditions indicate less concentration in the low default rate region, and the impact is particularly clear for the speculative grade default rates.

Figure 6.2: Univariate Default Rate Distributions



Having knowledge of the association between the two categories is important, and in this study the Pearson correlation coefficient is used. It is expected that the likelihood of defaults occuring in either rating class increases in bad times and decreases in good times, indicating positive correlation between the two rating categories. Figure 6.3 confirms the correlation between the default rates.

Figure 6.3: Correlation between Investment Grade and Speculative Grade

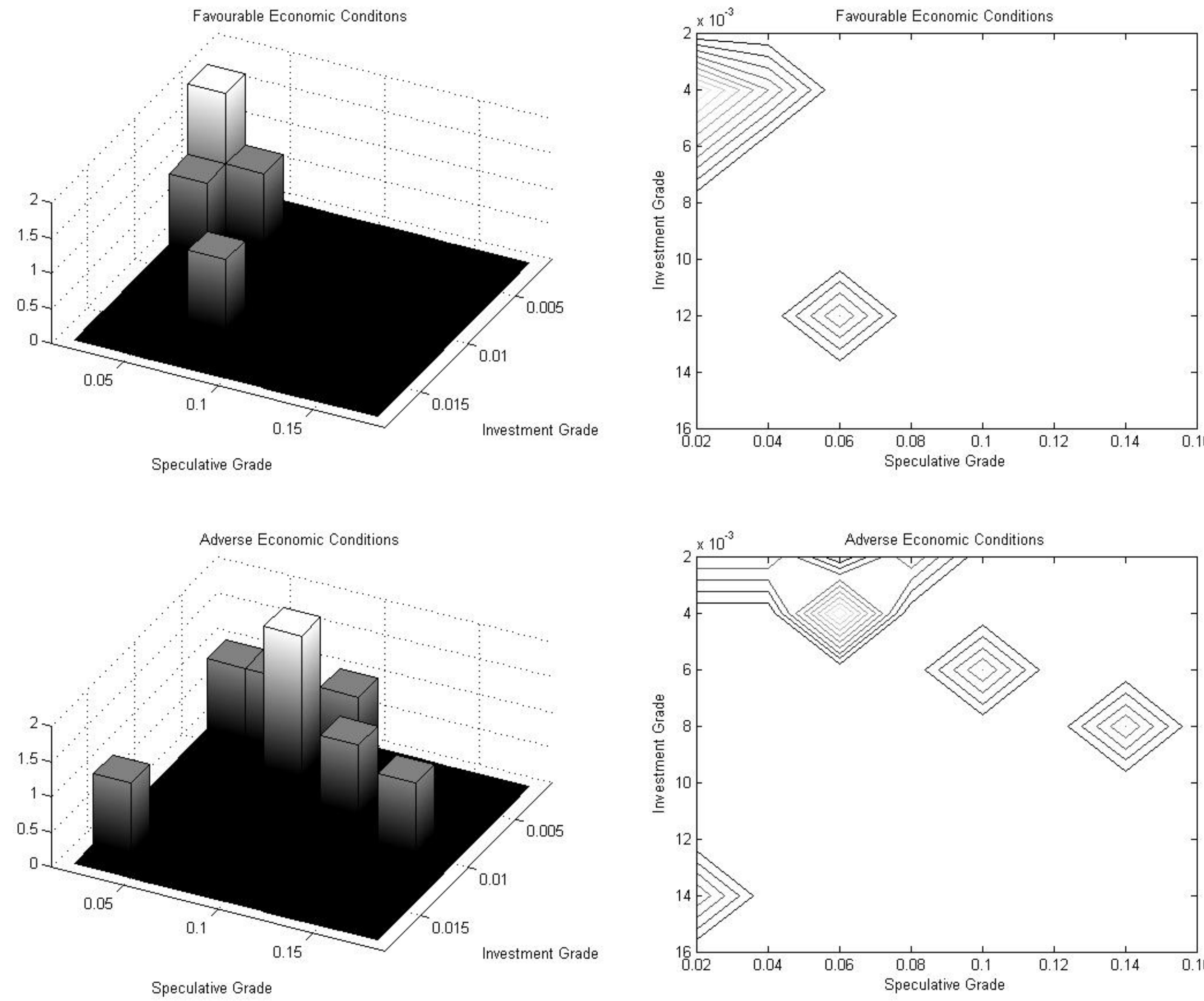


The positive correlation between investment grade and speculative grade default rates already indicates that the bivariate beta type III or extended bivariate beta type I distributions may be more appropriate as prior for this model.

Note that in 1938 there is a large investment grade default rate, which is likely to influence the correlation estimate for adverse economic conditions. However, since banks are particularly cautious about high investment grade default rates, the outlier has not been removed from the analysis.

Figure 6.4 below compares the joint distributions of default rates for the two samples. The contour plots are included for completeness.

Figure 6.4: Joint Distribution of Investment and Speculative Grade Default Rates



6.6 Prior Parameter Selection

6.6.1 Maximum Likelihood Estimation

Determining the parameters of the bivariate beta prior distributions with such little data proved to be quite challenging, as is generally the case with small samples. Using maximum likelihood estimation (MLE), the likelihood function has to be optimised numerically since explicit expressions for each of the parameters cannot be obtained. An additional problem arising with MLE is that the parameter estimates are easily influenced by the observations. For example, one very bad economic year with an investment grade default rate of 5% will have a significant impact on the default rate distribution.

Table 6.2 lists the parameter estimates obtained for the default rate distributions. Note that for the extended bivariate beta type I distributions the maximum likelihood estimates of the prior distribution parameters did not converge (indicated in the table as "DNC"). It is possible that this is due to the combination of the large number of parameters (6 parameters) to be estimated in conjunction with the small sample size (32 observations).

The parameters that could be obtained indicate that the parameter estimates obtained with MLE are consistent with the shape analysis performed in each section, although the positive correlation observed could not be captured.

Table 6.2: Parameter Selection: Maximum likelihood Estimates

	Favourable Economic Conditions				Adverse Economic Conditions			
	Type I	C and M	Type III	Ext Type I	Type I	C and M	Type III	Ext Type I
π_1	0.135	0.166	0.134	DNC	0.126	0.164	0.125	DNC
π_2	0.389	0.350	0.387	DNC	0.437	0.390	0.432	DNC
π_3	27.994	20.370	49.694	DNC	14.246	10.458	49.848	DNC
d	n/a	119.761	n/a	n/a	n/a	97.059	n/a	n/a
c	n/a	n/a	0.552	DNC	n/a	n/a	0.270	DNC
β_1	n/a	n/a	n/a	DNC	n/a	n/a	n/a	DNC
β_2	n/a	n/a	n/a	DNC	n/a	n/a	n/a	DNC
Log likelihood	376.097	379.421	376.155	n/a	341.431	346.198	341.566	n/a
Correlation	-0.008	-0.002	-0.026	n/a	-0.016	-0.003	-0.011	n/a

6.6.2 Method of Moments

An alternative to maximum likelihood estimation is to use the method of moments, where credit experts assign values to, say, the median, standard deviation, 5th percentile, 95th percentile, minimum, maximum, etc. The moments can be obtained relatively easily for the univariate case, see Kiefer (2008). Thinking in terms of "joint distributions" and "moments" is a lot more difficult for credit experts who don't have a mathematical background.

6.6.3 Bayesian Estimation of Shannon Entropy

The proposal is to use the Shannon entropy to determine the optimal values of the parameters for the various bivariate beta priors considered for the multinomial model, in conjunction with the data available and expert judgement. In this application, only the bivariate beta type III and extended bivariate beta type I will be considered as candidates due to their ability to account for positive correlation.

6.6.3.1 Bivariate beta type III distribution

The following steps are used to determine the parameters:

1. Determine the order of magnitude of the parameters using the conclusions from the shape analyses in Sections 2.1.5 and 4.1.5, where P_1 represents the investment grade default rates and P_2 represents the speculative grade default rates.

   (a) Favourable economic conditions: From the joint distributions in Figure 6.4 it is noted that for favourable economic conditions, the concentration of the distribution is towards small values of investment grade default rates (P_1) and small values of speculative grade default rates (P_2). This suggests a choice of parameters where π_1, π_2 and π_3 are less than c, see Figure 4.2.

   (b) Adverse economic conditions: From the joint distributions in Figure 6.4 it is noted that for adverse economic conditions, the concentration of the distribution is towards larger values of the speculative grade default rates (P_2), which can be obtained by choosing π_1 to be less than π_2, π_3 and c, see Figure 2.1.

2. Determine bands for the parameters using quantitative analyst expert judgement (and trial and error). For example, in the shape analysis, the combination of prior parameters $\pi_1 = \pi_2 = \pi_3 = 2$ and $c = 4$ was considered. However, comparing the contour plots between the theoretical bivariate beta type III distribution in Figure 4.2 and the observed distributions in Figure 6.4, the concentration of the distribution should be towards smaller values of P_1 and P_2, which can be obtained by:

    (a) Favourable economic conditions: Choose c to be much larger than π_1, π_2 and π_3.

    (b) Adverse economic conditions: Choose π_1 to be smaller than π_2 and π_3, and c to be larger than π_1, π_2 and π_3, but not as large as in the case of favourable economic conditions.

3. Using a grid search approach and an arbitrary step size, calculate the Shannon entropy using the Bayesian estimate in (4.13) with inputs:

    (a) Bivariate beta type III distribution parameters: The combination of parameters in the grid.

    (b) Multinomial distribution parameters: Since the focus of this analysis is on the selection of the prior distribution, $x_1 = 1$, $x_2 = 2$ and $x_3 = 10$ were used as the multinomial distribution observations. These can of course be changed as well.

4. Calculate the correlation for each combination of the parameters in the grid.

5. When selecting the parameters of the prior distributions, choose them such that:

    (a) The parameters are in a pre-specified Shannon entropy range, keeping in mind that lower Shannon entropy values are associated with less uncertainty around small values of P_1 and P_2, and therefore higher concentration in the distribution, see Figure 4.3. In this analysis, the bands can be different for the favourable and adverse economic conditions, since the concentration in the observed distributions are different. Selecting the range of Shannon entropy can be done by trial and error.

(b) The parameters are in the range of the observed correlation (0.64 for favourable economic conditions and 0.46 for adverse economic conditions), see Figure 6.3.

Table 6.3 summarises the grid search results. The first three columns provide information regarding the bounds used. Note that, in order to illustrate the results clearly, a very coarse grid has been used. In practice, it is advised to use a finer grid as this may significantly improve the accuracy of the parameter estimates. The last two columns provide the parameters chosen for the two bivariate beta type III distributions. The parameters were chosen by restricting the Shannon entropy and correlation estimates:

1. Favourable economic conditions: Choose the parameters such that the Shannon entropy is between 0.35 and 0.45, and the correlation is between 0.6 and 0.7. Deciding on the desired level of certainty (or uncertainty) in the distribution is no easy task. These bounds were chosen arbitrarily, but in practice one should investigate varying levels of Shannon entropy. The bounds for the correlation were chosen based on the observed correlation of 0.64.

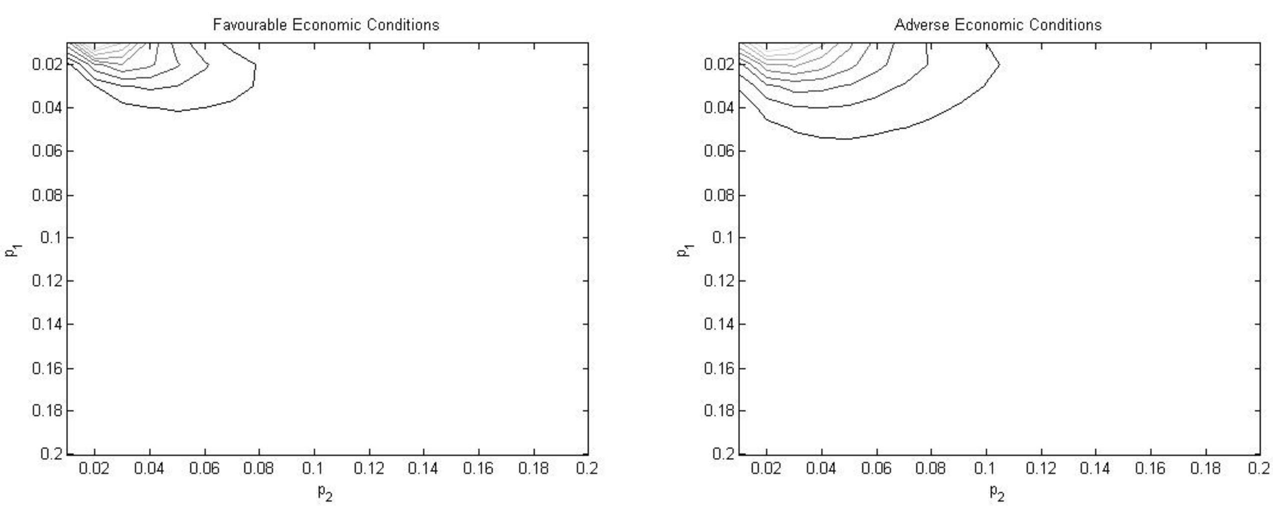
2. Adverse economic conditions: From the observed bivariate distributions in Figure 6.4 it can be seen that the concentration in the distribution reduces during adverse economic conditions, and therefore there is more uncertainty as to where the next default will occur. To account for the increased uncertainty, parameters will now only be chosen if the Shannon entropy lies between 0.45 and 0.55. In addition, the correlation has also decreased during adverse economic conditions, and the parameters now have to be such that they provide correlation between 0.4 and 0.5.

Table 6.4: Parameter Selection: Bivariate Beta Type III distribution

	Minimum	Step Size	Maximum	Favourable Economic Conditions	Adverse Economic Conditions
π_1	2	2	10	4	2
π_2	2	2	10	8	4
π_3	2	2	10	2	2
c	20	20	100	100	40
Shannon entropy				0.446	0.502
Correlation				0.698	0.491

Figure 6.5 compares the contour plots of the two fitted distributions. The parameters selected provide a bivariate beta type III distribution that successfully reflects the characteristics of the observed distributions.

Figure 6.5: Bivariate Beta Type III Parameters



6.6.3.2 Extended bivariate beta type I distribution

The steps for selecting the parameters are similar to those of the bivariate beta type III distribution, but will be presented here for completeness. They are:

1. Determine the order of magnitude of the parameters using the conclusions from the shape analyses in Sections 2.1.5 and 5.1.5, where P_1 represents the investment grade default rates and P_2 represents the speculative grade default rates.

   (a) Favourable economic conditions: From the joint distributions in Figure 6.4 it is noted that for favourable economic conditions, the concentration of the distribution is towards small values of investment grade default rates (P_1) and small values of speculative grade default rates (P_2). This suggests a possible choice of parameters where $\pi_1, \pi_2, \pi_3, \beta_1$ and β_2 are less than c, see Figure 5.5.

   (b) Adverse economic conditions: From the joint distributions in Figure 6.4 it is noted that for adverse economic conditions, the concentration of the distribution is towards larger values of the speculative grade default rates (P_2), which can be possibly be obtained by choosing β_1 to be less than $\pi_1, \pi_2, \pi_3, \beta_2$ and c, see Figure 5.2.

2. Determine bands for the parameters using quantitative analyst expert judgement (and trial and error). For example, in Figure 5.5 of the shape analysis, the combination of prior parameters $\pi_1 = \pi_2 = \pi_3 = 2$, $\beta_1 = \beta_2 = 1$, and $c = 4$ was considered. However, comparing the contour plots between the theoretical extended bivariate beta type I distributions in Section 5.1.5 and the observed distributions in Figure 6.4, the concentration of the distribution should be towards smaller values of P_1 and P_2, which can be obtained by:

   (a) Favourable economic conditions: Choose c to be much larger than $\pi_1, \pi_2, \pi_3, \beta_1$ and β_2.

   (b) Adverse economic conditions: Choose β_1 to be smaller than β_2, and c to be much larger than π_1, π_2 and π_3.

3. Using a grid search approach and an arbitrary step size, calculate the Shannon entropy using the Bayesian estimate in (5.11) with inputs:

   (a) Extended bivariate beta type I distribution parameters: The combination of parameters in the grid.

   (b) Multinomial distribution parameters: Since the focus of this analysis is on the selection of the prior distribution, $x_1 = 1$, $x_2 = 2$ and $x_3 = 10$ were used as the multinomial distribution observations. These can of course be changed as well.

4. Calculate the correlation for each combination of the parameters in the grid.

5. When selecting the parameters of the prior distributions, choose them such that:

   (a) The parameters are in a pre-specified Shannon entropy range, keeping in mind that lower Shannon entropy values are associated with less uncertainty, and therefore higher concentration, in the distribution, see Figure 5.5. In this analysis, the bands can be different for the favourable and adverse economic conditions, since the concentration in the observed distributions are different. Selecting the range of Shannon entropy can be done by trial and error.

   (b) The parameters are in the range of the observed correlation (0.64 for favourable economic conditions and 0.46 for adverse economic conditions), see Figure 6.3.

Table 6.4 summarises the grid search results. The first three columns provide information regarding the bounds used. Note that, in order to illustrate the results clearly, a very coarse grid has been used. In practice, it is advised to use a finer grid as this may significantly improve the accuracy of the parameter estimates. The last two columns provide the parameters chosen for the two extended bivariate beta type I distributions. The parameters were chosen by restricting the Shannon entropy and correlation estimates:

1. Favourable economic conditions: Choose the parameters such that the Shannon entropy is between 0.35 and 0.45, and the correlation is between 0.6 and 0.7. Deciding on the desired level of certainty (or uncertainty) in the distribution is no easy task. These bounds were chosen arbitrarily, but in practice one should investigate varying levels of Shannon entropy. The bounds for the correlation were chosen based on the observed correlation of 0.64.

2. Adverse economic conditions: From the observed bivariate distributions in Figure 6.4 it can be seen that the concentration in the distribution reduces during adverse economic conditions, and therefore there is more uncertainty as to where the next default will occur. To account for the increased uncertainty, parameters will now only be chosen if the Shannon

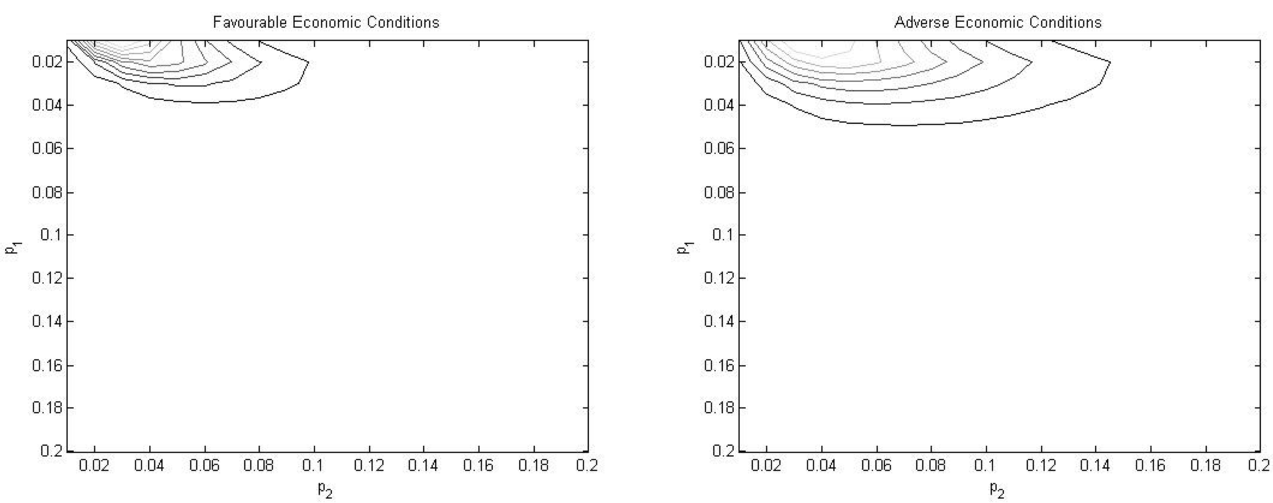
entropy lies between 0.45 and 0.55. In addition, the correlation has also decreased during adverse economic conditions, and the parameters now have to be such that they provide correlation between 0.4 and 0.5.

Table 6.5: Parameter Selection: Extended Bivariate Beta Type I Distribution

	Minimum	Step Size	Maximum	Favourable Economic Conditions	Adverse Economic Conditions
π_1	2	2	10	6	4
π_2	2	2	10	8	4
π_3	2	2	10	4	4
β_1	1	1	5	1	1
β_2	1	1	5	2	3
c	20	20	100	80	40
Shannon entropy				0.395	0.488
Correlation				0.624	0.423

Figure 6.6 compares the contour plots of the two fitted distributions. The parameters selected provide bivariate beta distributions which visually reflect the characteristics of the observed distributions.

Figure 6.6: Extended Bivariate Beta Type I Distribution



The parameters of the bivariate beta type III and extended bivariate beta type I distribution both result in distributions that compare well to the observed distributions, and also provide similar correlations. Note that the general level of the Shannon entropy for the extended bivariate beta type I distribution is lower than for the bivariate beta type III distribution. This could possibly be as a result of the additional parameters included in the bivariate beta type I distribution. Having more parameters implies that the distribution can be defined better, and therefore there is less uncertainty.

For this example, the bivariate beta type III distribution appears to be the best candidate, since it only requires one additional parameter to take into account the positive correlation between the investment grade default rates and the speculative grade default rates. This additional parameter is also flexible enough to provide different concentrations for favourable and adverse economic conditions. In practice, each situation should evaluated individually, and the appropriate distribution chosen accordingly.

6.7 Conclusion

The following important conclusions can be made:

- The bivariate beta type III and extended bivariate beta type I distributions are very flexible since they have the ability to deal with positive correlation in the underlying data.
- Using the Bayesian estimates of the Shannon entropy proved to be a useful aid in selecting the prior distribution when the sample size is small. The parameters selected for the bivariate beta prior distributions resulted not only in visually comparable distributions but also in correlation similar to what was observed.
- Now that a bivariate beta prior distribution has been selected, this can be used as part of the Bayesian calibration methodology.

Chapter 7

Conclusion

In credit risk, a statistician is often faced with the problem of having very little data. In order to determine probabilities of default for small and low-default portfolios a Bayesian calibration approach was considered, since Bayesian methods are generally useful when sample sizes are small and a lot of reliance is given on expert judgement.

This study provided a detailed analysis of the bivariate beta type I, Connor and Mosimann bivariate beta, bivariate beta type III and extended bivariate beta type I distributions. For each of the bivariate beta distributions investigated the marginal distributions, correlation, method of derivation and impact of parameters on the shape of the distribution were discussed in detail. These bivariate beta distributions were then considered as priors for the multinomial model, from which the posterior distribution and Bayesian estimators of the Shannon entropy were successfully derived.

The following concluding remarks prevail:

1. Positive correlation between P_1 and P_2 can only be attained with the bivariate beta type III and extended bivariate beta type I distributions.

   (a) Bivariate beta type I distribution: This distribution only allows for negative correlation.

(b) Connor and Mosimann bivariate beta distribution: This distribution does not allow for positive correlation, although its generalisation to more than two variables allows for positive correlation.

(c) Bivariate beta type III distribution: This distribution allows for positive and negative correlation, and the larger c is, the larger the positive value of correlation is.

(d) Extended bivariate beta type I distribution: This distribution allows for positive and negative correlation. Decreasing β_1 and β_2 simultaneously increases the range for which positive correlation can be attained, whilst increasing β_1 and β_2 simultaneously decreases the range for which positive correlation can be attained.

2. Extending the bivariate beta type I distribution by including additional parameters adds to the flexibility of the distributions.

   (a) Connor and Mosimann bivariate beta distribution: Small values of d shift the distribution towards small values of P_2 along the line $p_1 + p_2 = 1$, and large values of d shift the bivariate distribution towards the marginal distribution of P_2.

   (b) Bivariate beta type III distribution: Large values of c shift the distribution towards small values of P_1 and P_2, whereas small values of c shift the distribution towards values of P_1 and P_2 along the line $p_1 + p_2 = 1$.

   (c) Extended bivariate beta type I distribution: Small values of β_1 shift the distribution towards the marginal distribution of P_2, and large values of β_2 shift the distribution towards small values of P_2 along the line $p_1 + p_2 = 1$. Similarly, small values of β_2 shift the distribution towards the marginal distribution of P_1, and large values of β_2 shift the distribution towards large values of P_2 along the line $p_1 + p_2 = 1$. The similar results are due to the symmetry present in the extended bivariate beta type I distribution. Small values of c shift the distribution towards the line $p_1 + p_2 = 1$, and large values of c shift the distribution towards small values of P_1 and P_2.

3. The parameters of the prior distribution have a significant impact on the Bayesian estimator of the Shannon entropy (as expected).

   (a) Bivariate beta type I distribution: Decreasing π_1 or π_2 reduces $\hat{H}_3^I$, indicating less uncertainty, and increasing π_1 or π_2 increases $\hat{H}_3^I$, indicating more uncertainty. Decreasing π_3 increases $\hat{H}_3^I$, indicating more uncertainty, and increasing $\hat{H}_3^I$, indicating less uncertainty.

   (b) Connor and Mosimann bivariate beta distribution: Decreasing π_1 or π_2 reduces $\hat{H}_3^{CM}$, indicating less uncertainty in the distribution. Conversely, increasing π_1 or π_2 increases $\hat{H}_3^{CM}$, indicating more uncertainty in the distribution. Decreasing π_3 increases $\hat{H}_3^{CM}$ indicating more uncertainty, and increasing π_3 decreases $\hat{H}_3^{CM}$ indicating less uncertainty in the distribution. Larger values of d are associated with lower Shannon entropy values, indicating less uncertainty. In summary, as the concentration in the distribution remains closer to small values of P_1 and P_2, $\hat{H}_3^{CM}$ stays lower, but as soon as the concentration moves away from these small values to some point along the line $p_1 + p_2 = 1$ the uncertainty increases.

   (c) Bivariate beta type III distribution: Larger values of c are associated with lower Shannon entropy values, indicating less uncertainty.

   (d) Extended bivariate beta type I distribution: Decreasing β_1 or β_2 respectively reduces $\hat{H}_3^E$, indicating less uncertainty. Conversely, increasing β_1 or β_2 respectively increases $\hat{H}_3^E$, indicating more uncertainty in the distribution. Decreasing β_1 and β_2 simultaneously increases $\hat{H}_3^E$, indicating more uncertainty, whilst increasing β_1 and β_2 simultaneously decreases $\hat{H}_3^E$, indicating less uncertainty. In general, larger values of c are associated with lower Shannon entropy values, indicating less uncertainty. In summary, a larger concentration around small values of P_1 and P_2 is associated with lower $\hat{H}_3^E$, and as the concentration moves away $\hat{H}_3^E$ increases.

In general, the Bayesian estimator of Shannon entropy for the bivariate beta prior distributions contains information of the concentration as well as location of the distribution.

The knowledge obtained from the detailed study of the various bivariate beta distributions was used in an application to credit risk, where the selection of a bivariate beta prior as part of a Bayesian calibration methodology was considered. The Bayesian estimate of the Shannon entropy proved to be a good aid when selecting the appropriate parameters for the prior distribution.

Areas for further research include, but are not limited to:

- This study investigated the Bayesian estimation of Shannon entropy using the squared error loss function. It can be investigated using other loss functions, such as linear loss.
- The Bayesian estimation of other entropy measures, such as the Kullback-Leibler information measure.
- The posterior distribution of the Shannon entropy can be investigated in more detail.
- In credit risk, most research has primarily been done on a portfolio level due to the dimensional restrictions of the frequentist approach when little data is available. Using the Bayesian approach, the results obtained for the bivariate beta prior distributions can be generalised to Dirichlet prior distributions, which will cater for the situation of more than three categories in the multinomial distribution.
- One could use the Markov Chain Monte Carlo (MCMC) method to obtain the Bayesian estimates of the Shannon entropy for the posterior densities in Chapters 3, 4 and 5, but this was not part of this study.

In final conclusion, this study provided useful insight about various bivariate beta distributions as well as the Bayesian estimation of Shannon entropy.

Appendix A

Notation

This appendix lists the notation used in this study.

Notation	Description
$\sim$	Distributed as
X	Random variable
x	Observed value of the random variable X
$\mathbf{X} = (X_1, X_2, ..., X_k)$	Vector of random variables
$\mathbf{x} = (x_1, x_2, ..., x_k)$	Vector of observed values of the random variable $\mathbf{X}$
$f(x)$	Distribution of X
$f(x\|y)$	Distribution of X, conditional on Y
$E[X]$	Expected value of X
$E[X\|Y]$	Expected value of X, conditional on Y
$\|\mathbf{X}\|$	Determinant of a matrix $\mathbf{X}$
e^x	Exponential function
$\ln x$	Natural logarithmic function
$L(x, \hat{x})$	Loss function

$\Gamma(\alpha)$	Gamma function
${}_2F_1(\alpha, \beta; \gamma; x)$	Gauss hypergeometric function of one variable
$F_1(\alpha, \beta, \beta', \gamma; x, y)$	Hypergeometric function of two variables
H_k	Shannon entropy
$Beta(\alpha, \beta)$	Univariate beta distribution with parameters α and β
$BBeta^I(\alpha_1, \alpha_2, \alpha_3)$	Bivariate beta type I distribution with parameters α_1, α_2 and α_3
$BBeta^{CM}(\alpha_1, \alpha_2, \alpha_3, d)$	Connor and Mosimann bivariate beta distribution with parameters $\alpha_1, \alpha_2, \alpha_3$ and d
$BBeta^{III}(\alpha_1, \alpha_2, \alpha_3, c)$	Bivariate beta type III distribution with parameters $\alpha_1, \alpha_2, \alpha_3$ and c
$BBeta^E(\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, c)$	Extended bivariate beta type I distribution with parameters $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$ and c
$\chi^2(\alpha)$	Chi-squared distribution with parameter α
$Gamma(\alpha, \beta)$	Gamma distribution with parameters α and β

Appendix B

Mathematical Preliminaries

This appendix lists known definitions and relations used in this study.

Definition 1 (Gradshteyn and Ryzhik, 2007, p892)

The beta function is defined as

$$B(\alpha_1, \alpha_2, ..., \alpha_n) = \frac{\prod_{i=1}^{n} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{n} \alpha_i)}$$

Definition 2 (Gradshteyn and Ryzhik, 2007, p902)

The polygamma or psi function is defined as the derivative of the logarithmic gamma function

$$\begin{aligned} \psi(x) &= \frac{d}{dx} \ln \Gamma(x) \\ &= \frac{\Gamma'(x)}{\Gamma(x)} \end{aligned}$$

Definition 3 (Gradshteyn and Ryzhik, 2007, pXliii)

The Pochhammer coefficient is defined as

$$\begin{aligned}(\alpha)_k &= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} \\ &= \alpha(\alpha+1)(\alpha+2)...(\alpha+k-1)\end{aligned}$$

Definition 4 (Gradshteyn and Ryzhik, 2007, p1005)

The Gauss hypergeometric function is defined as

$$ {}_2F_1(\alpha,\beta;\gamma;x) = \sum_{k=0}^{\infty} \frac{(\alpha)_k(\beta)_k}{(\gamma)_k}\frac{x^k}{k!}$$

and the integral representation is

$$ {}_2F_1(\alpha,\beta;\gamma;x) = \frac{1}{B(\beta,\gamma-\beta)}\int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tx)dt$$

for $Re\,\gamma > 0$ and $Re\,\beta > 0$.

Definition 5 (Gradshteyn and Ryzhik, 2007, p1018,1021)

The hypergeometric function of two variables is defined as

$$F_1(\alpha,\beta,\beta',\gamma;x,y) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{(\gamma)_{m+n}m!n!}x^m y^n$$

for $|x| < 1$ and $|y| < 1$, and the integral representation is

$$F_1(\alpha,\beta,\beta',\gamma;x,y) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\beta')\Gamma(\gamma-\beta-\beta')} \times \iint_{\substack{u\geq 0, v\geq 0\\ u+v\leq 1}} u^{\beta-1}v^{\beta'-1}(1-u-v)^{\gamma-\beta-\beta'-1}(1-ux-vy)^{-\alpha}dudv$$

for $Re\,\beta > 0$, $Re\,\beta' > 0$, and $Re\,(\gamma-\beta-\beta') > 0$.

Relation 1 (Gradshteyn and Ryzhik, 2007, p315)

$$\int_0^u x^{\nu-1}(u-x)^{\mu-1}dx = u^{\mu+\nu-1}B(\mu,\nu) \tag{B.1}$$

for $Re\,\mu > 0$ and $Re\,\nu > 0$.

Relation 2 (Prudnikov et al., 1986, p301)

$$\int_a^b (x-a)^{\alpha-1}(b-x)^{\beta-1}(cx+d)^{\gamma-1}dx = (b-a)^{\alpha+\beta-1}(ac+d)^{\gamma}B(\alpha,\beta)_2F_1(\alpha,-\gamma;\alpha+\beta;\frac{c(a-b)}{ac+d}) \tag{B.2}$$

Relation 3 (Gradshteyn and Ryzhik, 2007, p25)

The binomial expansion of a power series is

$$(1+x)^q = \sum_{k=0}^{\infty}\binom{q}{k}x^k \tag{B.3}$$

and is a special case of the Maclaurin series (see Stewart,1999, p763). The series always converges if $|x| < 1$. Convergence at 1 and -1 depends on the value of q:

- If $q \geq 0$, the series will converge at both 1 and -1.
- If $-1 < q \leq 0$, the series will converge at 1.

If $q = n$ is a natural number, the series is reduced to the finite sum:

$$(a+x)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x^k a^{n-k}$$

Relation 4 (Prudnikov et al., 1986, p566)

$$\begin{aligned} &\int\int_{\substack{x \geq 0, y \geq 0 \\ x+y \leq 1}} x^{\beta-1} y^{\beta'-1} (1-x-y)^{\gamma-\beta-\beta'-1} (1-ux-vy)^{-\alpha} dx dy \\ = \quad & \Gamma \begin{bmatrix} \beta, \beta', \gamma-\beta-\beta' \\ \gamma \end{bmatrix} F_1(\alpha, \beta, \beta', \gamma; u, v) \end{aligned} \tag{B.4}$$

where $\Gamma \begin{bmatrix} a_1, a_2, ..., a_m \\ b_1, b_2, ... b_n \end{bmatrix} = \frac{\prod_{i=1}^{m} \Gamma(a_i)}{\prod_{j=1}^{n} \Gamma(b_j)}$,

and $F_1(\alpha, \beta, \beta', \gamma; u, v)$ is the hypergeometric function of two variables, see Definition 5.

Relation 5 (Gradshteyn and Ryzhik, 2007, p1020)

$$F_1(\alpha, \beta, \beta', \gamma; x, x) = {}_2F_1(\alpha, \beta+\beta'; \gamma; x) \tag{B.5}$$

where $F_1(\alpha, \beta, \beta', \gamma; x, x)$ is the hypergeometric function of two variables (see Definition 5) and ${}_2F_1(\alpha, \beta+\beta'; \gamma; x)$ is the Gauss hypergeometric function (see Definition 4).

Relation 6 (Gradshteyn and Ryzhik, 2007, p1008)

$$ {}_2F_1(\alpha,\beta;\gamma;z) = (1-z)^{-\beta}\,{}_2F_1(\gamma-\alpha,\beta;\gamma;\frac{z}{z-1}) \tag{B.6} $$

Relation 7 (Mathai, 1993, p3)

$$ \binom{n}{k} = \frac{(-1)^k(-n)_k}{k!} \tag{B.7} $$

Relation 8 (Gradshteyn and Ryzhik, 2007, p1008)

$$ {}_2F_1(\alpha,\beta,\gamma,1) = \frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \tag{B.8} $$

Appendix C

Computer Programs

All computations in this study were performed using Matlab.

Some computational difficulties were encountered with the gamma functions, as the argument becomes too large the results no longer converge. To resolve this problem, the continuous form representation of the Gauss hypergeometric function of one variable and the hypergeometric function of two variables were determined using rectangular numerical integration. Similarly, rectangular numerical integration was used for the valuation of the Bayesian estimates of the Shannon entropy for the Connor and Mosimann bivariate beta, bivariate beta type III and extended bivariate beta type I priors.

C.1 General Programs

Program: Num2F1.m

```
function H2F1 = Num2F1(a,b,c,z)

% Calculates the Gauss Hypergeometric function using numerical integration
if b <= 0 || c <= 0,
    disp('Error: Real(b) and Real(c) must be greater than 0');
```

```
else

    width = 0.00001;

    area = 0;

    for t = 0 :width : 1;

        if t == 1,

            length = 0;

        else

            length = (t^(b-1))*((1-t)^(c-b-1))*((1-t*z))^(-a);

        end;

        area = area + length*width;

    end;

    H2F1 = area*gamma(c)/(gamma(b)*gamma(c-b));

end;


end
```

Program: NumAppellF1.m

```
function AF1 = NumAppellF1Short(a,b1,b2,c,x,y)

% Calculates the hypergeometric function of two variables
% using numerical integration
if b <= 0 || c <= 0,
    disp('Error: Real(b) and Real(c) must be greater than 0');
else
    width = 0.0001;
    area = 0;
    for u = 0 :width : 1;
        if u == 1,
            length = 0;
        else
            length = u^(a-1)*(1-u)^(c-a-1)*(1-u*x)^-b1*(1-u*y)^-b2;
        end;
        area = area + length*width;
    end;
    AF1 = area*gamma(c)/(gamma(a)*gamma(c-a));
% end;
end
```

C.2 Bivariate Beta Type I Prior

C.2.1 Correlation

Program: Corr_BBetaI.m

```
function corr_p1p2 = Corr_BBetaI(piI)

% This function caluclates the correlation between p1 and p2
% for the bivariate beta type I distribution

pi1 = piI(1); pi2 = piI(2); pi3 = piI(3);
corr_p1p2 = -sqrt(pi1*pi2/((pi1+pi3)*(pi2+pi3)));

end
```

C.2.2 Shape Analysis

Program: ShapeAnalysis_BBetaI.m

```
close all; clear; clc;


% This program is used to perform the shape analysis for the bivariate beta

% type I distribution.


pi1 = 2; pi2 = 2; pi3 = 10;

g = gamma(pi1+pi2+pi3)/(gamma(pi1)*gamma(pi2)*gamma(pi3));


[pp1,pp2] = meshgrid(0.01:0.01:1);

f_p_all = ones(size(pp1,1),size(pp2,1));

for i = 1 : size(pp1,1);

    for j = 1 : size(pp2,1);

        if (pp1(i,j) + pp2(i,j) <=1),

            f_p = g*(pp1(i,j)^(pi1-1))*(pp2(i,j)^(pi2-1))* ...

                    ((1-pp1(i,j)-pp2(i,j))^(pi3-1));

            f_p_all(i,j) = f_p;

        else

            f_p_all(i,j) = NaN;

        end;

    end;

end;


figure;

surf(pp1,pp2,f_p_all);

title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2), ...

        ' \pi_3 = ',num2str(pi3)]);

xlabel('p_1'); ylabel('p_2'); zlabel('f(p)');
```

```
colormap(gray); axis([0 1 0 1 0 Inf]);
view([2 1 3]);

figure;
contour(pp2,pp1,f_p_all);
title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2), ...
        ' \pi_3 = ',num2str(pi3)]);
xlabel('p_2'); ylabel('p_1');
colormap(gray);
view([0 -90]);
```

C.2.3 Shannon Entropy

Program: ShEntr_BB1.m

```
function ShEntr = ShEntr_BB1(ppi,n)

% This function calculates the Shannon Entropy for the bivariate
% beta type I distribution

pi1 = ppi(1); pi2 = ppi(2); pi3 = ppi(3);
n1 = n(1); n2 = n(2); n3 = n(3);

b1 = pi1 + n1;
b2 = pi2 + n2;
b3 = pi3 + n3;
b = [b1, b2, b3];

ShEntr = 0;
for i = 1 : 3;
    sum_i = -(b(i)/(sum(b)))*(psi(b(i)+1)-psi(sum(b)+1));
    ShEntr = ShEntr + sum_i;
end;

end
```

Program: NumericalAnalysis_BBetaI.m

```
close all; clear; clc;

% This program is used to conduct the numerical analysis of the Bayesian
% estimate for the bivariate beta type I distribution.
```

```
n1 = 1; n2 = 2; n3 = 10;
n = [n1, n2, n3];

pi1 = 2; pi2 = 2; pi3 = 2;

% Changing pi1
% pi1 = 1; pi2 = 2; pi3 = 2;
% pi1 = 10; pi2 = 2; pi3 = 2;

% Changing pi2
% pi1 = 2; pi2 = 1; pi3 = 2;
% pi1 = 2; pi2 = 10; pi3 = 2;

% Changing pi3
% pi1 = 2; pi2 = 2; pi3 = 1;
% pi1 = 2; pi2 = 2; pi3 = 10;

pi = [pi1, pi2, pi3];

ShEntI = ShEntr_BB1(pi,n);
```

C.3 Connor and Mosimann Bivariate Beta Prior

C.3.1 Correlation

Program: Corr_BBetaCM.m

```
function corr_p1p2 = Corr_BBetaCM(piCM)

% This function caluclates the correlation between p1 and p2
% for the Connor & Mosimann bivariate beta distribution

pi1 = piCM(1); pi2 = piCM(2); pi3 = piCM(3); d = piCM(4);
ep1 = pi1/(pi1+d);
ep12 = (pi1+1)*pi1/((pi1+d+1)*(pi1+d));
ep2 = pi2*d/((pi2+pi3)*(pi1+d));
ep22 = (pi2+1)*pi2*(d+1)*d/((pi2+pi3+1)*(pi2+pi3)*(pi1+d+1)*(pi1+d));
ep1p2 = pi1*pi2*d/((pi2+pi3)*(pi1+d+1)*(pi1+d));

varp1 = ep12 - ep1^2;
varp2 = ep22 - ep2^2;
covp1p2 = ep1p2 - ep1*ep2;

corr_p1p2 = covp1p2/sqrt(varp1*varp2);

end
```

Program: CorrCM.m

```
close all; clear; clc;

% This program is used to determine the correlation for the Connor &
```

```
% Mosimann bivariate beta distribution.


pi1 = 2;

pi2 = 2;

pi3 = 2;


outputmat = [];

for d = 0.1 : 0.1 : 200;

    piCM = [pi1, pi2, pi3, d];

    corrp1p2 = Corr_BBetaCM(piCM);

    outputmat = [outputmat;[d,corrp1p2]];

end;


figure;

plot(outputmat(:,1),outputmat(:,2),'k');

xlim([1,200]);

title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2),' \pi_3 = ',num2str(pi3)]);

xlabel('d');

ylabel('corr(p_1,p_2)');
```

C.3.2 Shape Analysis

Program: ShapeAnalysis_BBetaCM.m

```
close all; clear; clc;


% This program is used to perform the shape analysis for the Connor and

% Mosimann bivariate beta distribution


pi1 = 2; pi2 = 2;  pi3 = 2;

d = 10;

g = gamma(pi1+d)*gamma(pi2+pi3)/(gamma(pi1)*gamma(pi2)*gamma(pi3)*gamma(d));


[pp1,pp2] = meshgrid(0.01:0.01:1);

f_p_all = ones(size(pp1,1),size(pp2,1));

for i = 1 : size(pp1,1);

    for j = 1 : size(pp2,1);

        if (pp1(i,j) + pp2(i,j) <=1),

            f_p = g*pp1(i,j)^(pi1-1)*pp2(i,j)^(pi2-1)*(1-pp1(i,j)-pp2(i,j))^(pi3-1) ...

                    *(1-pp1(i,j))^(d-pi2-pi3);

            f_p_all(i,j) = f_p;

        else

            f_p_all(i,j) = NaN;

        end;

    end;

end;


figure;

surf(pp1,pp2,f_p_all);

title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2), ' \pi_3 = ',num2str(pi3), ...

        ' d = ',num2str(d)]);
```

```
xlabel('p_1'); ylabel('p_2'); zlabel('f(p)');
colormap(gray); axis([0 1 0 1 0 Inf]);
view([2 1 3]);


figure;
contour(pp2,pp1,f_p_all);
title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2), ' \pi_3 = ',num2str(pi3), ...
        ' d = ',num2str(d)]);
xlabel('p_2'); ylabel('p_1');
colormap(gray);
view([0, -90]);
```

C.3.3 Shannon Entropy

Program: Calculate_ShEntr_BBCM.m

```
function output = Calculate_ShEntr_BBCM(n1,n2,n3,pi1,pi2,pi3,dmin,dstep,dmax)


% Numerically

output = [];


for d = dmin : dstep : dmax;


    g = gamma(pi1+d) * gamma(pi2+pi3) / (gamma(pi1) * gamma(pi2) * gamma(pi3) * gamma(d));

    n = factorial(n1+n2+n3) / (factorial(n1) * factorial(n2) * factorial(n3));


    % fix size of grid to ensure that calculations do not run too long

    [pp1,pp2] = meshgrid(0.01:0.01:1);

    f_post_all = zeros(size(pp1,1),size(pp2,1));

    f_entr_all = zeros(size(pp1,1),size(pp2,1));

    area_post = 0;

    area_entr = 0;

    for i = 1 : size(pp1,1);

        for j = 1 : size(pp2,1);

            if (pp1(i,j) + pp2(i,j) <=1),

                f_post = g * n * (pp1(i,j)^(pi1+n1-1)) * (pp2(i,j)^(pi2+n2-1)) ...

                    * ((1-pp1(i,j)-pp2(i,j))^(pi3+n3-1)) * (1-pp1(i,j))^(d-pi2-pi3);

                f_post_all(i,j) = f_post;

                area_post = area_post + f_post*0.01*0.01;


                if f_post > 0 && isnan(f_post)==0,

                    f_entr = -f_post*(pp1(i,j)*log(pp1(i,j)) + pp2(i,j)*log(pp2(i,j)) ...

                                    + (1-pp1(i,j)-pp2(i,j))*log(1-pp1(i,j)-pp2(i,j)));
```

```
                end;
                f_entr_all(i,j) = f_entr;
                area_entr = area_entr + f_entr*0.01*0.01;
            else
                f_post_all(i,j) = NaN;
            end;
        end;
    end;

    ShEntr = area_entr/area_post;

    output = [output; [d,ShEntr]];
end;

end
```

Program: ShEntr_BBCM.m

```
close all; clear; clc;

% likelihood parameters
n1 = 1; n2 = 2; n3 = 10;

% prior parameters
pi1_1 = 2; pi2_1 = 2; pi3_1 = 2;
pi1_2 = 2; pi2_2 = 2; pi3_2 = 1;
pi1_3 = 2; pi2_3 = 2; pi3_3 = 10;

dmin = 0.1;
dstep = 0.1;
```

```
dmax = 50;

ShEntr_CM = zeros((dmax-dmin)/dstep+2,6);

% Calculate
ShEntr_CM(:,1:2) = Calculate_ShEntr_BBCM(n1,n2,n3,pi1_1,pi2_1,pi3_1,dmin,dstep,dmax);
ShEntr_CM(:,3:4) = Calculate_ShEntr_BBCM(n1,n2,n3,pi1_2,pi2_2,pi3_2,dmin,dstep,dmax);
ShEntr_CM(:,5:6) = Calculate_ShEntr_BBCM(n1,n2,n3,pi1_3,pi2_3,pi3_3,dmin,dstep,dmax);

series1 = [' \pi_1 = ',num2str(pi1_1),' \pi_2 = ',num2str(pi2_1),' \pi_3 = ',num2str(pi3_1)];
series2 = [' \pi_1 = ',num2str(pi1_2),' \pi_2 = ',num2str(pi2_2),' \pi_3 = ',num2str(pi3_2)];
series3 = [' \pi_1 = ',num2str(pi1_3),' \pi_2 = ',num2str(pi2_3),' \pi_3 = ',num2str(pi3_3)];
series = {series1;series2;series3};

plot(ShEntr_CM(:,1),ShEntr_CM(:,2),'-k',ShEntr_CM(:,3),ShEntr_CM(:,4),':k',...
        ShEntr_CM(:,5),ShEntr_CM(:,6),'--k');
title(' Shannon entropy: Changing \pi_3 ');
legend(series,'Location','SW');
xlim([dmin,dmax]); xlabel('d');
ylim([0.4,1]); ylabel('Shannon Entropy');
```

C.4 Bivariate Beta Type III Prior

C.4.1 Correlation

Program: ProductMomentBetaIII.m

```
function EW1H1W2H2 = ProductMomentBetaIII(piIII,h1,h2)

pi1 = piIII(1); pi2 = piIII(2); pi3 = piIII(3); c = piIII(4);

G1 = gamma(pi1 + h1)*gamma(pi2 + h2)/(gamma(pi1)*gamma(pi2));
G2 = gamma(pi1 + pi2 + pi3)/gamma(pi1 + pi2 + pi3 + h1 + h2);
F = Num2F1(pi1+pi2+h1+h2, pi1+pi2+pi3, pi1+pi2+pi3+h1+h2, 1-c);

EW1H1W2H2 = G1*G2*F*c^(pi1+pi2);

end
```

Program: CorrBBetaIII.m

```
% This program caluclates the correlation of p1,p2 for a range of c-values

close all; clear; clc;

a1 = 2;
a2 = 2;
a3 = 2;

corr_p1p2 = [];
cmin = 0.1;
cstep = 0.1;
```

```
cmax = 20;
for c = cmin : cstep : cmax;
    a = [a1,a2,a3,c];
    E_p1_1 = ProductMomentBetaIII(a,1,0);
    E_p1_2 = ProductMomentBetaIII(a,2,0);
    var_p1 = E_p1_2 - (E_p1_1)^2;

    E_p2_1 = ProductMomentBetaIII(a,0,1);
    E_p2_2 = ProductMomentBetaIII(a,0,2);
    var_p2 = E_p2_2 - (E_p2_1)^2;

    E_p1p2 = ProductMomentBetaIII(a,1,1);
    cov_p1p2 = E_p1p2 - E_p1_1*E_p2_1;

    corr_p1p2_c = cov_p1p2/sqrt(var_p1*var_p2);
    corr_p1p2 = [corr_p1p2;corr_p1p2_c];
end;

figure;
plot(cmin:cstep:cmax,corr_p1p2,'k',cmin:cstep:cmax,0,'k');
ylim([-1,1]);
title([' \pi_1 = ',num2str(a1),' \pi_2 = ',num2str(a2),' \pi_3 = ',num2str(a3)]);
xlabel('c');
ylabel('corr(p_1,p_2)');
```

C.4.2 Shape Analysis

Program: ShapeAnalysis_BBetaIII.m

```
close all; clear; clc;

% This program is used to perform the shape analysis for the bivariate beta
% type III distribution.

pi1 = 2; pi2 = 4; pi3 = 2;
c = 40;
g = gamma(pi1+pi2+pi3)/(gamma(pi1)*gamma(pi2)*gamma(pi3))*c^(pi1+pi2);

% [pp1,pp2] = meshgrid(0.01:0.01:1);

% for application use a smaller grid, will make analysis more meaningful
[pp1,pp2] = meshgrid(0.01:0.01:0.2);

f_p_all = ones(size(pp1,1),size(pp2,1));
for i = 1 : size(pp1,1);
    for j = 1 : size(pp2,1);
        if (pp1(i,j) + pp2(i,j) <=1),
            f_p = g*(pp1(i,j)^(pi1-1))*(pp2(i,j)^(pi2-1))*((1-pp1(i,j)-pp2(i,j))^(pi3-1)) ...
                    *(1+(c-1)*pp1(i,j)+(c-1)*pp2(i,j))^(-(pi1+pi2+pi3));
            f_p_all(i,j) = real(f_p);
        else
            f_p_all(i,j) = NaN;
        end;
    end;
end;
```

```
figure;
surf(pp1,pp2,f_p_all);
% title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2), ' \pi_3 = ',num2str(pi3), ...
%          ' c = ',num2str(c)]);
title('Adverse Economic Conditions');
xlabel('p_1'); ylabel('p_2'); zlabel('f(p)');
colormap(gray); view([2 1 3]);
% axis([0 1 0 1 0 Inf]); % for shape analysis
axis([0 0.2 0 0.2 0 Inf]); % for application

figure;
contour(pp2,pp1,f_p_all);
% title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2), ' \pi_3 = ',num2str(pi3),...
%          ' c = ',num2str(c)]);
title('Adverse Economic Conditions');
xlabel('p_2'); ylabel('p_1');
colormap(gray); view([0 -90]);
```

C.4.3 Shannon Entropy

Program: ShEntr_BBetaIII.m

```
function ShEntr = ShEntr_BBetaIII(piIII,n)


% This function is used to calculate the Shannon Entropy for the bivariate

% beta type III distribution.


% Note: Numerical integration is used due to technical difficulties

% encountered when using the Bayesian Estimates.


pi1 = piIII(1); pi2 = piIII(3); pi3 = piIII(3);

c = piIII(4);

n1 = n(1); n2 = n(2); n3 = n(3);


g = gamma(pi1+pi2+pi3) / (gamma(pi1) * gamma(pi2) * gamma(pi3));

kc = c^(pi1+pi2);

n = factorial(n1+n2+n3) / (factorial(n1) * factorial(n2) * factorial(n3));


[pp1,pp2] = meshgrid(0.01:0.01:1);

f_post_all = zeros(size(pp1,1),size(pp2,1));

f_entr_all = zeros(size(pp1,1),size(pp2,1));

area_post = 0;

area_entr = 0;

for i = 1 : size(pp1,1);

    for j = 1 : size(pp2,1);

        if (pp1(i,j) + pp2(i,j) <=1),

            f_post = g * n * kc * (pp1(i,j)^(pi1+n1-1)) * (pp2(i,j)^(pi2+n2-1)) ...

                    * ((1-pp1(i,j)-pp2(i,j))^(pi3+n3-1)) ...

                    * (1+(c-1)*pp1(i,j)+(c-1)*pp2(i,j))^(-(pi1+pi2+pi3));
```

```
            f_post_all(i,j) = f_post;
            area_post = area_post + f_post*0.01*0.01;

            if f_post > 0 && isnan(f_post)==0,
                f_entr = -f_post*(pp1(i,j)*log(pp1(i,j)) + pp2(i,j)*log(pp2(i,j)) ...
                                + (1-pp1(i,j)-pp2(i,j))*log(1-pp1(i,j)-pp2(i,j)));
            end;
            f_entr_all(i,j) = f_entr;
            area_entr = area_entr + f_entr*0.01*0.01;
        else
            f_post_all(i,j) = NaN;
        end;
    end;
end;

ShEntr = area_entr/area_post;
end
```

C.5 Extended Bivariate Beta Type I Prior

C.5.1 Correlation

Program: ProductMomentBetaExt.m

```
function EW1H1W2H2 = ProductMomentBetaExt(ppi,h1,h2)

pi1 = ppi(1); pi2 = ppi(2); pi3 = ppi(3);
b1 = ppi(4); b2 = ppi(5);
c = ppi(6);

GNum = gamma(pi1+pi2+pi3)*gamma(pi1+h1)*gamma(pi2+h2);
GDenom = gamma(pi1+pi2+pi3+h1+h2)*gamma(pi1)*gamma(pi2);
Const = (b1^(-pi1))*(b2^(-pi2))*(c^(pi1+pi2));
F1 = NumAppellF1Short(pi1+pi2+pi3,pi1+h1,pi2+h2,pi1+pi2+pi3+h1+h2,1-c/b1,1-c/b2);

EW1H1W2H2 = (GNum/GDenom)*Const*F1;

end
```

Program: Calculate_CorrBBetaExt.m

```
% This program caluclates the correlation of p1,p2 for a range of c-values

function corr_p1p2 = Calculate_CorrBBetaExt(a_in,cmin,cstep,cmax)

corr_p1p2 = [];

for c = cmin : cstep : cmax;
    a = [a_in,c];
```

```
    E_p1_1 = ProductMomentBetaExt(a,1,0);
    E_p1_2 = ProductMomentBetaExt(a,2,0);
    var_p1 = E_p1_2 - (E_p1_1)^2;

    E_p2_1 = ProductMomentBetaExt(a,0,1);
    E_p2_2 = ProductMomentBetaExt(a,0,2);
    var_p2 = E_p2_2 - (E_p2_1)^2;

    E_p1p2 = ProductMomentBetaExt(a,1,1);
    cov_p1p2 = E_p1p2 - E_p1_1*E_p2_1;

    corr_p1p2_c = cov_p1p2/sqrt(var_p1*var_p2);
    corr_p1p2 = [corr_p1p2;corr_p1p2_c];

end;

end
```

C.5.2 Shape Analysis

Program: ShapeAnalysis_BBetaExt.m

```
close all; clear; clc;


% This program is used to perform the shape analysis for the extended
% bivariate beta type I distribution


pi1 = 2; pi2 = 2;  pi3 =2;
b1 = 4; b2 = 4;
c = 1;
g = gamma(pi1+pi2+pi3)/(gamma(pi1)*gamma(pi2)*gamma(pi3))*b1^(-pi1)*b2^(-pi2)*c^(pi1+pi2);



[pp1,pp2] = meshgrid(0.01:0.01:1);

% for application use a smaller grid, will make analysis more meaningful
% [pp1,pp2] = meshgrid(0.01:0.01:0.2);

f_p_all = ones(size(pp1,1),size(pp2,1));
for i = 1 : size(pp1,1);
    for j = 1 : size(pp2,1);
        if (pp1(i,j) + pp2(i,j) <=1),
            f_p = g*(pp1(i,j)^(pi1-1))*(pp2(i,j)^(pi2-1))*((1-pp1(i,j)-pp2(i,j))^(pi3-1)) ...
                        *(1-(1-c/b1)*pp1(i,j)-(1-c/b2)*pp2(i,j))^-(pi1+pi2+pi3);
            f_p_all(i,j) = f_p;
        else
            f_p_all(i,j) = NaN;
        end;
    end;
```

```
end;


figure;
surf(pp1,pp2,f_p_all);
title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2), ' \pi_3 = ',num2str(pi3), ...
        ' \beta_1 = ',num2str(b1), ' \beta_2 = ',num2str(b2),' c = ',num2str(c)]);
% title('Favourable Economic Conditions');
xlabel('p_1'); ylabel('p_2'); zlabel('f(p)');
colormap(gray); view([2 1 3]);
axis([0 1 0 1 0 Inf]); % for shape analysis
% axis([0 0.2 0 0.2 0 Inf]); % for application


figure;
contour(pp2,pp1,f_p_all);
title([' \pi_1 = ',num2str(pi1),' \pi_2 = ',num2str(pi2), ' \pi_3 = ',num2str(pi3), ...
        ' \beta_1 = ',num2str(b1), ' \beta_2 = ',num2str(b2),' c = ',num2str(c)]);
% title('Favourable Economic Conditions');
xlabel('p_2'); ylabel('p_1');
colormap(gray);
view([0 -90]);
```

C.5.3 Shannon Entropy

Program: ShEntr_BBetaExt.m

```
function ShEntr = ShEntr_BBetaExt(ppi,n)


pi1 = ppi(1); pi2 = ppi(2); pi3 = ppi(3);
b1 = ppi(4); b2 = ppi(5);
c = ppi(6);


n1 = n(1); n2 = n(2); n3 = n(3);


% Numerically
g = gamma(pi1+pi2+pi3) / (gamma(pi1) * gamma(pi2) * gamma(pi3));
kc = c^(pi1+pi2) * b1^(-pi1) * b2^(-pi2);
n = factorial(n1+n2+n3) / (factorial(n1) * factorial(n2) * factorial(n3));


% fix size of grid to ensure that calculations do not run too long
[pp1,pp2] = meshgrid(0.01:0.01:1);
f_post_all = zeros(size(pp1,1),size(pp2,1));
f_entr_all = zeros(size(pp1,1),size(pp2,1));
area_post = 0;
area_entr = 0;
for i = 1 : size(pp1,1);
    for j = 1 : size(pp2,1);
        if (pp1(i,j) + pp2(i,j) <=1),
            f_post = g * n * kc * (pp1(i,j)^(pi1+n1-1)) * (pp2(i,j)^(pi2+n2-1)) ...
                    * ((1-pp1(i,j)-pp2(i,j))^(pi3+n3-1)) * ...
                    (1-(1-c/b1)*pp1(i,j)-(1-c/b2)*pp2(i,j))^(-(pi1+pi2+pi3));
            f_post_all(i,j) = f_post;
            area_post = area_post + f_post*0.01*0.01;
```

```
            if f_post > 0 && isnan(f_post)==0,
                f_entr = -f_post*(pp1(i,j)*log(pp1(i,j)) + pp2(i,j)*log(pp2(i,j)) ...
                         + (1-pp1(i,j)-pp2(i,j))*log(1-pp1(i,j)-pp2(i,j)));
            end;
            f_entr_all(i,j) = f_entr;
            area_entr = area_entr + f_entr*0.01*0.01;
        else
            f_post_all(i,j) = NaN;
        end;
    end;
end;

ShEntr = area_entr/area_post;
end
```

C.6 Application to Credit Risk

Program: Application_1.m

```
close all; clear; clc;

% The purpose of this program is to graphically compare the observed
% default rates to GDP

load MoodysDefaultRates.txt;
x = MoodysDefaultRates;

% This data contains the Moody's default rates from 1930 to 2008 as well as
% the USA seasonally adjusted GDP change

% Standardise values in order to make them comparable
n1 = size(x);
lastn = 29;
z = zscore(x(n1-lastn+1:n1,:));
n = size(z,1);

figure;
plot(1980:2008,z(:,10),'-k',1980:2008,z(:,11),':k',1980:0.05:2008,0,'-k');
xlim([1980,2008]);
title('Default Rates and US GDP');
legend('Default Rate','US GDP','Location','SE');
```

Program: Application_2.m

```
close all; clear; clc;

load MoodysDefaultRates.txt;

InvGr = MoodysDefaultRates(:,8);
SpecGr = MoodysDefaultRates(:,9);
GDP = MoodysDefaultRates(:,11);

GoodYears = find(GDP>=prctile(GDP,60));
BadYears = find(GDP<=prctile(GDP,40));

% Divide by 100 since default rates are given as % in report
DRGood = [InvGr(GoodYears),SpecGr(GoodYears)]/100;
DRBad = [InvGr(BadYears),SpecGr(BadYears)]/100;

% Compare univariate distributions - inv grade
figure;
bar([0.002:0.002:0.016],BinInvGrade(DRGood(:,1)),1);
title('Investment Grade:  Favourable Economic Conditions');
set(findobj(gca,'Type','patch'),'FaceColor','k');
xlim([0 0.018]); ylim([0 26]);
xlabel('Default Rate'); ylabel('Frequency');

figure;
bar([0.002:0.002:0.016],BinInvGrade(DRBad(:,1)),1);
title('Investment Grade:  Adverse Economic Conditions');
set(findobj(gca,'Type','patch'),'FaceColor','k');
xlim([0 0.018]); ylim([0 26]);
```

```
xlabel('Default Rate'); ylabel('Frequency');

% Compare univariate distributions - Spec grade
figure;
bar([0.02:0.02:0.16],BinSpecGrade(DRGood(:,2)),1);
title('Speculative Grade:  Favourable Economic Conditions');
set(findobj(gca,'Type','patch'),'FaceColor','k');
xlim([0 0.18]); ylim([0 23]);
xlabel('Default Rate'); ylabel('Frequency');

figure;
bar([0.02:0.02:0.16],BinSpecGrade(DRBad(:,2)),1);
title('Speculative Grade:  Adverse Economic Conditions');
set(findobj(gca,'Type','patch'),'FaceColor','k');
xlim([0 0.18]); ylim([0 23]);
xlabel('Default Rate'); ylabel('Frequency');

% Correlation between good and bad times
CorrGood = corrcoef(DRGood);
figure;
plot(DRGood(:,1),DRGood(:,2),'.k','MarkerSize',14);
title({'Favourable Economic Conditions';['Correlation = ',num2str(CorrGood(1,2))]});
CorrGood = corrcoef(DRGood);
xlabel('Investment Grade'); ylabel('Speculative Grade');

CorrBad = corrcoef(DRBad);
figure;
plot(DRBad(:,1),DRBad(:,2),'.k','MarkerSize',14);
title({'Adverse Economic Conditions';['Correlation = ',num2str(CorrBad(1,2))]});
xlabel('Investment Grade'); ylabel('Speculative Grade');
```

```
% Plot observed bivariate distribution

xInv = 0.002:0.002:0.016;
ySpec = 0.02:0.02:0.16;
edg = {xInv;ySpec};

figure;
hist3(DRGood,'FaceAlpha',0.85,'Edges',edg);
title('Favourable Economic Conditons');
xlabel('Investment Grade'); ylabel('Speculative Grade');
surfHandle = get(gca, 'child');
set(surfHandle,'FaceColor','interp', 'CdataMode', 'auto');
colormap(gray);
view([2 1 3]);

figure;
hist3(DRBad,'FaceAlpha',0.85,'Edges',edg);
title('Adverse Economic Conditions');
xlabel('Investment Grade'); ylabel('Speculative Grade');
surfHandle = get(gca, 'child');
set(surfHandle,'FaceColor','interp', 'CdataMode', 'auto');
colormap(gray);
view([2 1 3]);

% Contourplots

hGood = hist3(DRGood,'FaceAlpha',0.85,'Edges',edg);
figure;
contour(ySpec,xInv,hGood); colormap(gray);
```

```
title('Favourable Economic Conditions');
xlabel('Speculative Grade'); ylabel('Investment Grade');
view([0 -90]);

hBad = hist3(DRBad,'FaceAlpha',0.85,'Edges',edg);
figure;
contour(ySpec,xInv,hBad); colormap(gray);
title('Adverse Economic Conditions');
xlabel('Speculative Grade'); ylabel('Investment Grade');
view([0 -90]);

% Fit BBetaI
paramBetaI_0 = [2; 2; 2];
% A and b are the constraints of beta I distribution
A = [-1, 0, 0; 0, -1, 0; 0, 0, -1];
b = ones(3,1)*(0);
[paramBetaI_Good,LLHIval_Good] = fmincon(@(paramBetaI_0)LogLikelihoodBetaI ...
                (paramBetaI_0,DRGood(:,1),DRGood(:,2)),paramBetaI_0,A,b);
[paramBetaI_Bad,LLHIval_Bad] = fmincon(@(paramBetaI_0)LogLikelihoodBetaI ...
                (paramBetaI_0,DRBad(:,1),DRBad(:,2)),paramBetaI_0,A,b);

corr_BetaI_Good = Corr_BBetaI(paramBetaI_Good);
corr_BetaI_Bad = Corr_BBetaI(paramBetaI_Bad);

% Fit CMBBeta
paramBetaCM_0 = [2; 2; 2; 2];
% A and b are the constraints of beta I distribution
A = [-1, 0, 0, 0; 0, -1, 0, 0; 0, 0, -1, 0];
b = ones(3,1)*(0);
[paramBetaCM_Good,LLHCMval_Good] = fmincon(@(paramBetaCM_0)LogLikelihoodBetaCM ...
```

```
                   (paramBetaCM_0,DRGood(:,1),DRGood(:,2)),paramBetaCM_0,A,b);
[paramBetaCM_Bad,LLHCMval_Bad] = fmincon(@(paramBetaCM_0)LogLikelihoodBetaCM ...
                   (paramBetaCM_0,DRBad(:,1),DRBad(:,2)),paramBetaCM_0,A,b);

corr_BBetaCM_Good = Corr_BBetaCM(paramBetaCM_Good);
corr_BBetaCM_Bad = Corr_BBetaCM(paramBetaCM_Bad);

% Fit BBetaIII
paramBetaIII_0 = [0.5; 0.5; 50; 0.5];
% A and b are the constraints that the estimated parameters must be
% positive
A = [-1 0 0 0;
    0 -1 0 0;
    0 0 -1 0;
    0 0 0 -1];
b = ones(4,1)*(0);
[paramBetaIII_Good,LLHIIIval_Good] = fmincon(@(paramBetaIII_0)LogLikelihoodBetaIII ...
                    (paramBetaIII_0,DRGood(:,1),DRGood(:,2)),paramBetaIII_0,A,b);
[paramBetaIII_Bad,LLHIIIval_Bad] = fmincon(@(paramBetaIII_0)LogLikelihoodBetaIII ...
                    (paramBetaIII_0,DRBad(:,1),DRBad(:,2)),paramBetaIII_0,A,b);

corr_BBetaIII_Good = Corr_BBetaIII(paramBetaIII_Good);
corr_BBetaIII_Bad = Corr_BBetaIII(paramBetaIII_Bad);

% Fit BBetaExt
paramBetaExt_0 = [0.9; 0.9; 20; 2; 2; 0.5];
% A and b are the constraints of beta I distribution
A = [-1, 0, 0, 0, 0, 0;
    0, -1, 0, 0, 0, 0;
    0, 0, -1, 0, 0, 0;
```

```
    0, 0, 0, -1, 0, 0;
    0, 0, 0, 0, -1, 0;
    0, 0, 0, 0, 0, -1];
b = ones(6,1)*(0);
[paramBetaExt_Good,LLHExtval_Good] = fmincon(@(paramBetaExt_0)LogLikelihoodBetaExt ...
                    (paramBetaExt_0,DRGood(:,1),DRGood(:,2)),paramBetaExt_0,A,b);
[paramBetaExt_Bad,LLHExtval_Bad] = fmincon(@(paramBetaExt_0)LogLikelihoodBetaExt ...
                    (paramBetaExt_0,DRBad(:,1),DRBad(:,2)),paramBetaExt_0,A,b);

% Save the files
dlmwrite('DRGood.txt',DRGood);
dlmwrite('DRBad.txt',DRBad);
```

Program: LogLikelihoodBetaI.m

```
function LLHBetaI=LogLikelihoodBetaI(param,p1,p2)

pi1 = param(1);
pi2 = param(2);
pi3 = param(3);

n = size(p1,1);

% add small number to factors to avoid problems with zero default rates
s1 = sum(log(p1 + 0.000001));
s2 = sum(log(p2 + 0.000001));
s3 = sum(log(1 - p1 - p2 + 0.000001));

g = gamma(pi1 + pi2 + pi3)/(gamma(pi1)*gamma(pi2)*gamma(pi3));

LLHTmp = n*log(g) + (pi1-1)*s1 + (pi2-1)*s2 + (pi3-1)*s3;

% Matlab optimisation constraint: can only do minimisation
% For MLE: minimise negative LLH
LLHBetaI = -LLHTmp;
end
```

Program: LogLikelihoodBetaCM.m

```
function LLHBetaCM = LogLikelihoodBetaCM(param,p1,p2)

pi1 = param(1);
pi2 = param(2);
pi3 = param(3);
```

```
d = param(4);

n = size(p1,1);

% add small number to factors to avoid problems with zero default rates
s1 = sum(log(p1 + 0.000001));
s2 = sum(log(p2 + 0.000001));
s3 = sum(log(1 - p1 - p2 + 0.000001));
s4 = sum(log(1 - p1 + 0.000001));

g = (gamma(pi1 + d)*gamma(pi2 + pi3))/(gamma(pi1)*gamma(pi2)*gamma(pi3)*gamma(d));

LLHTmp = n*log(g) + (pi1-1)*s1 + (pi2-1)*s2 + (pi3-1)*s3 + (d-pi2-pi3)*s4;
LLHBetaCM = -LLHTmp;
end
```

Program: LogLikelihoodBetaIII.m

```
function LLHBetaIII = LogLikelihoodBetaIII(param,p1,p2)

pi1 = param(1);
pi2 = param(2);
pi3 = param(3);
c = param(4);

n = size(p1,1);

% add small number to factors to avoid problems with zero default rates
s1 = sum(log(p1 + 0.000001));
s2 = sum(log(p2 + 0.000001));
```

```
s3 = sum(log(1 - p1 - p2 + 0.000001));
s4 = sum(log(1 + (c-1)*p1 + (c-1)*p2 + 0.000001));

g = gamma(pi1 + pi2 + pi3)/(gamma(pi1)*gamma(pi2)*gamma(pi3));

LLHTmp = n*log(g) + n*(pi1+pi2)*log(c) + (pi1-1)*s1 + (pi2-1)*s2 + (pi3-1)*s3 ...
    - (pi1+pi2+pi3)*s4;
LLHBetaIII = -LLHTmp;
end
```

Program: LogLikelihoodBetaExt.m

```
function LLHBetaIII = LogLikelihoodBetaExt(param,p1,p2)

pi1 = param(1);
pi2 = param(2);
pi3 = param(3);
c = param(4);
b1 = param(5);
b2 = param(6);

n = size(p1,1);

% add small number to factors to avoid problems with zero default rates
s1 = sum(log(p1 + 0.000001));
s2 = sum(log(p2 + 0.000001));
s3 = sum(log(1 - p1 - p2 + 0.000001));
s4 = sum(log(1 - ((1-c)/b1)*p1 + ((1-c)/b2)*p2 + 0.000001));

g = gamma(pi1 + pi2 + pi3)/(gamma(pi1)*gamma(pi2)*gamma(pi3));
```

```
const = -pi1*log(b1) - pi2*log(b2) + (pi1+pi2)*log(c);


LLHTmp = n*log(g) + n*const + (pi1-1)*s1 + (pi2-1)*s2 + (pi3-1)*s3 - (pi1+pi2+pi3)*s4;

LLHBetaIII = -LLHTmp;

end
```

Program: Application_3_III.m

```
close all; clear; clc;

tic;
% Assume multinomial parameters are constant (for this example)
n1 = 1;
n2 = 2;
n3 = 10;
n = [n1, n2, n3];

% Create parameter matrix
pi1min = 2; pi1step = 2; pi1max = 10;
pi2min = 2; pi2step = 2; pi2max = 10;
pi3min = 2; pi3step = 2; pi3max = 10;
cmin = 20; cstep = 20; cmax = 100;

xpi1 = pi1min:pi1step:pi1max;
xpi2 = pi2min:pi2step:pi2max;
xpi3 = pi3min:pi3step:pi3max;
c = cmin : cstep : cmax;

npi1 = (pi1max - pi1min)/pi1step + 1;
npi2 = (pi2max - pi2min)/pi2step + 1;
npi3 = (pi3max - pi3min)/pi3step + 1;
nc = (cmax - cmin)/cstep + 1;

xc = repmat(c',npi3,1);
x3 = sortrows(repmat(xpi3',nc,1));
x3c = repmat([x3,xc],npi2,1);
```

```
x2 = sortrows(repmat(xpi2',npi3*nc,1));
x23c = repmat([x2,x3c],npi1,1);


x1 = sortrows(repmat(xpi1',npi2*npi3*nc,1));
x123c = [x1,x23c,zeros(npi1*npi2*npi3*nc,2)];


% calculate Shannon entropy and correlation
for i = 1 : npi1*npi2*npi3*nc;
    ShEntr_III_i = ShEntr_BBetaIII(x123c(i,1:4),n);
    corr_i = Corr_BBetaIII(x123c(i,1:4));
    x123c(i,5:6) = [ShEntr_III_i, corr_i];
end;
dlmwrite('BBetaIII_Comb5.txt',x123c);
toc;
```

Program: Application_3_Ext.m

```
close all; clear; clc;

tic;
% Assume multinomial parameters are constant (for this example)
n1 = 1;
n2 = 2;
n3 = 10;
n = [n1, n2, n3];

% Create parameter matrix
pi1min = 2; pi1step = 2; pi1max = 10;
pi2min = 2; pi2step = 2; pi2max = 10;
pi3min = 2; pi3step = 2; pi3max = 10;
cmin = 20; cstep = 20; cmax = 100;
b1min = 1; b1step = 1; b1max = 5;
b2min = 1; b2step = 1; b2max = 5;

xpi1 = pi1min:pi1step:pi1max;
xpi2 = pi2min:pi2step:pi2max;
xpi3 = pi3min:pi3step:pi3max;
b1 = b1min : b1step : b1max;
b2 = b2min : b2step : b2max;
c = cmin : cstep : cmax;

npi1 = (pi1max - pi1min)/pi1step + 1;
npi2 = (pi2max - pi2min)/pi2step + 1;
npi3 = (pi3max - pi3min)/pi3step + 1;
nb1 = (b1max - b1min)/b1step + 1;
```

```
nb2 = (b2max - b2min)/b2step + 1;
nc = (cmax - cmin)/cstep + 1;

xc = repmat(c',nb2,1);
xb2 = sortrows(repmat(b2',nc,1));
xb2c = repmat([xb2,xc],nb1,1);

xb1 = sortrows(repmat(b1',nb2*nc,1));
xb1b2c = repmat([xb1,xb2c],npi3,1);

x3 = sortrows(repmat(xpi3',nb1*nb2*nc,1));
x3b1b2c = repmat([x3,xb1b2c],nb2,1);

x2 = sortrows(repmat(xpi2',npi3*nb1*nb2*nc,1));
x23b1b2c = repmat([x2,x3b1b2c],nb1,1);

xb1 = sortrows(repmat(xpi1',npi2*npi3*nb1*nb2*nc,1));
x123cb1b2 = [xb1,x23b1b2c ,zeros(npi1*npi2*npi3*nb1*nb2*nc,2)];

% calculate Shannon entropy and correlation
for i = 1 : npi1*npi2*npi3*nb1*nb2*nc;
    ShEntr_Ext_i = ShEntr_BBetaExt(x123cb1b2(i,1:6),n);
    corr_i = Corr_BBetaExt(x123cb1b2(i,1:6));
    x123cb1b2(i,7:8) = [ShEntr_Ext_i, corr_i];
end;
dlmwrite('BBetaExt_Comb2.txt',x123cb1b2);
toc;
```

References

Balakrishnan, N. and Lai, C. D., 2009, *Continuous bivariate distributions*, 2^{nd} ed., New York: Springer.

Bank for International Settlements, 2006, *Basel II: Revised international capital framework*, viewed 31 March 2010, `http://www.bis.org/publ/bcbs128.htm`.

Berger, J.O., 1980, *Statistical decision theory: foundations, concepts and methods*, New York: Springer.

Berger, A.L., Della Pietra, S. A., Della Pietra, V.J., 1996, 'A maximum entropy approach to natural language processing', Computational Linguistics, vol. 22, no. 1., pp. 1-36.

Berger, Z.D., 2003, 'Bayesian and frequentist models: legitimate choices for different purposes', *Annals of epidemiology*, vol. 13, no. 8, p. 583.

Bernardo, J.M. and Smith, A.F.M., 2000, *Bayesian theory*, Chichester, England: Wiley.

Bishop, C.M., 2007, *Pattern recognition and machine learning*, New York: Springer.

Box, G.E.P. and Tiao, G.C., 1992, *Bayesian inference in statistical analysis*, New York: Wiley.

Cardeño, L., Nagar, D.K., Sànchez, L.E., 2005, 'Beta type 3 distribution and its multivariate generalization', *Tamsui Oxford Journal of Mathematical Sciences*, vol. 21, no. 2, pp. 225-241.

Connor R.J. and Mosimann, J.E., 1969, 'Concepts of independence for proportions with a generalization of the Dirichlet distribution', *Journal of the American Statistical Association*, vol. 64, no. 325, pp. 194-206.

Ebrahimi, N., Habibullah, M. and Soofi, E.S., 1992, 'Testing exponentiality based on Kullback-Leibler information', *Journal of the Royal Statistical Society, Series B*, vol. 54, no. 3, pp. 739-748.

Ehlers, R. and Bekker, A., 2010, 'Extended bimatrix beta type I distribution', *In preparation*.

Ehlers, R. Bekker, A. and Roux, J.J.J., 2009, 'The central and non-central matrixvariate Dirichlet type III distribution', *South African Statistical Journal*, vol. 43, no. 2, pp. 97-116.

Encyclopedia of Statistical Sciences, 2006, 2^{nd} ed., Volume 3, Hoboken, N.J.: Wiley-Interscience.

Giudici, P., 2003, *Applied data mining: statistical methods for business and industry*, England: Wiley.

Gössl, C., 2005, 'Predictions based on certain uncertainties - a Bayesian credit portfolio approach', viewed 31 March 2010, `http://www.defaultrisk.com/pp_model119.htm`.

Kiefer, N., 2008, 'Default estimation for low-default portfolios', viewed 31 March 2010, `http://www.arts.cornell.edu/econ/kiefer/SmallPFinal3_08.pdf`.

Kotz, S., Balakrishnan, N. and Johnson, N.L., 2000, *Continuous multivariate distributions, volume 1: models and applications*, 2^{nd} ed., New York: Wiley.

Gradshteyn, I.S. and Ryzhik, I.M., 2007, *Tables of integrals, series and products*, 7^{th} ed., London: Academic Press.

Lee, P.M., 2004, *Bayesian statistics: an introduction*, 3^{rd} ed., London: Wiley.

Lochner, R.H., 1975, 'A generalized Dirichlet distribution in Bayesian life testing', *Journal of the Royal Statistical Society*, Series B, vol. 37, no. 1, pp. 103-113.

Löffler, G., Posch, P.N., and Schöne, C., 2005, 'Bayesian methods for improving credit scoring models', viewed 31 March 2010, `www.defaultrisk.com/pp_score_46.htm`.

Mathai, A.M., 1993, *A handbook of generalized special functions for statistical and physical sciences*, Oxford: Claredon Press.

Martz, H.F. and Waller, R.A., 1982, *Bayesian reliability analysis*, New York: Wiley.

Moody's, 2009, 'Corporate Defaults and Recovery Rates, 1920-2008', viewed 31 March 2010, `http://www.moodys.com/cust/content/content.ashx?source=StaticContent/Free%20Pages/Credit%20Policy%20Research/documents/current/2007400000578875.pdf`.

Nadarajah, S. and Kotz, S., 2005, 'Some bivariate beta distributions', *Journal of Statistics*, vol. 39, no. 5, pp. 457-466.

O' Hagan, A. and Forster, J., 2004, *Kendall's advanced theory of statistics, Volume 2B: Bayesian Inference*, London: Arnold.

Ong, M.K., 2002, *Credit Ratings: Methodologies, Rationale and Default Risk*. London: Risk Books.

Pielou, E.C., 1967, 'The use of information theory in the study of the diversity of biological populations', in *Proceedings of the fifth Berkely symposium on mathematical statistics and probability, Vol. 4: Biology and problems of health*, Berkely, Calif.: University of California Press, pp. 163-177.

Pluto, K. and Tasche, D., 2005, 'Thinking positively', *Risk*, vol. 18, no. 8, pp. 72-78.

Press, J.S., 1989, *Bayesian statistics: principles, models and applications*, New Jersey: Wiley.

Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I., 1986, *Integrals and Series, volume I: Elementary functions*, New York: Gordon and Breach.

Raiffa, H. and Schlaifer, R., 1961, *Applied Statisical Decision Theory*, Boston: Harvard University Press.

Schuermann, T. and Hanson, S., 2004, 'Estimating probabilities of default', viewed 31 March 2010, `http://www.newyorkfed.org/research/staff_reports/sr190.html`.

Shannon, C.E., 1948, 'A mathematical theory of communication', *Bell System technical journal*. vol. 27, July and Oct., pp. 379-423, and pp. 623-656.

Simion, E., 1999, 'Remarks on the Bayesian estimation of Shannon entropy using prior truncated Dirichlet distributions', *Mathematics report,* vol. 1, no. 2, pp. 227-288.

Simion, E., 2000, 'Asymptotic behaviour of some statistical estimators used in cryptograhpy', *Analele Universitii din Bucaresti*, vol. 49, no. 1, pp. 85-98.

Stephanides, G., 2005, 'Bayesian estimates in cryptography', *WSEAS transactions on communications*, vol. 4, no. 7, pp. 449-454.

Stewart, J., 1999, *Calculus: early transcendentals*, 4th ed., Pacific Grove, Calif.: Brooks/Cole.

Tasche, D., 2010, 'Estimating discriminatory power and PD curves when the number of defaults is small', viewed 31 March 2010, `http://www.defaultrisk.com/pp_test_45.htm`.

Trück, S. and Rachev, S., 2005, 'Credit portfolio risk and PD confidence sets through the business cycle', *Journal of Credit Risk*, vol. 1, no. 4, pp. 61-88.

Vallverdu, J., 2008, 'The false dilemma: Bayesian vs. Frequentist', *Electronic Journal for Philosophy*, viewed 31 March 2010, `http://nb.vse.cz/kfil/elogos/science/vallverdu08.pdf`.

Van der Burght, M.J., 2008, 'Calibrating low-default portfolios, using the cumulative accuracy profile', *Journal of risk model validation*, vol. 1, no. 4, pp. 1-17.

Vasicek, O., 1976, 'A test for normality based on sample entropy', *Journal of the Royal Statistical Society, Series B*, vol. 38, no. 1, pp. 54-59.

Wong, T.-T., 1998, 'Generalized Dirichlet distribution in Bayesian analysis', *Journal of applied mathematics and computation*, vol. 97, no. 2-3, pp. 165-181.

Zellner, A., 1996, 'Models, prior information, and Bayesian analysis', *Journal of econometrics*, vol 75, no 1., pp. 51-68.