A Technical Note on the Logitnormal Distribution

Patrizio Frederic* and **Frank Lad†**

Abstract

We present the technical details of the logitnormal distribution of a variable supported on the open unit-interval. Although this distribution has been used widely in a variety of statistical applications under a transformation, its direct analysis has not been exposited. The density function can be expressed algebraically, and the moments can be computed easily by numerical integration as functions of the parameters of the transformed normal distribution. The family of densities has characteristics different from the Beta family especially near 0 and 1, although some members of each family resemble each other. Limiting members of the family provide examples of distributions with adherent masses at 0 and 1. We also examine the bivariate exchangeable logitnormal distribution.

Key Words: Numerical integration, logitnormal moments, adherent masses.

1 The Logitnormal density and its moments

While the logit transformation of a variable on the unit-interval is well-known and has been studied in myriad applications, the Logitnormal distribution has escaped direct exposition. This is somewhat surprising given the widespread use of logistic regression analysis. The distribution has also been used in hierarchical Bayesian modelling as early as Leonard (1972) and in applications such as that of Kass and Steffey (1989). In these cases it is applied in the form of the normal distribution for the logit transformation of a variable in the unit-interval, rather than directly in the form of a distribution over that interval. In this note we examine the density directly in the univariate and exchangeable bivariate form, exhibiting its structure and its properties. The direct analysis of the density via numerical integration yields much clearer understanding of the possible choice of a family member as a mixing distribution in applications, and it also can aid the interpretation of logistic regression results.

Definition: A variable Θ is distributed *Logitnormal* (μ, σ^2) if its logit transformation, $log(\Theta/(1-\Theta))$, is distributed Normal (μ, σ^2).

The density function for Θ can be derived from the normal density for $log(\Theta/(1-\Theta))$ using standard transformation methods.

Theorem: If $\Theta \sim Logitnormal(\mu, \sigma^2)$ then its density function is

$$g(\theta) \;=\; \frac{1}{\sqrt{2\pi}\,\sigma\,\theta\,(1-\theta)}\;\exp\{-\frac{1}{2}\,[\,\frac{log\,(\frac{\theta}{1-\theta})-\mu}{\sigma}\,]^2\} \qquad \text{for}\ \ \theta\ \in\ (\ 0,\ 1\ )\ . \qquad (1)$$

This density function for θ is not symmetric unless $\mu = 0$, in which case $E(\Theta) = 1/2$. Generally, the expectation and variance for Θ are determined by functions $M(\mu, \sigma^2)$ and $V(\mu, \sigma^2)$ that do not have algebraic solutions in closed form. However, we can easily compute values for these mean and variance functions using numerical integration. Figure 1 displays the mean and variance functions using the signal to noise parameter, μ/σ, and σ as arguments. Figure 2 displays contours of $(\mu/\sigma, \sigma)$ pairs that support constant values of the

*Researcher, University of Bologna
†Research Associate, University of Canterbury

mean at $E(\Theta) = .6, .75$, and $.9$, and values of the variance at $V(\Theta) = .01, .04$, and $.10$. We have limited this display to non-negative values of μ/σ due to the obvious result that if $\Theta \sim Logitnormal(\mu, \sigma^2)$, then $(1 - \Theta) \sim Logitnormal(-\mu, \sigma^2)$.

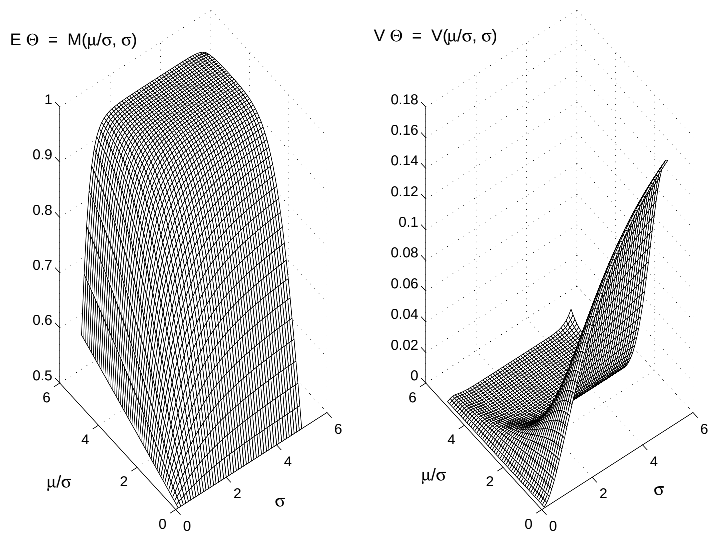

Figure 1: The graph at left depicts values of $E(\Theta)$ as a function of the Logitnormal μ/σ and σ parameters. At right appears the Variance of Θ as a function of these same parameters.

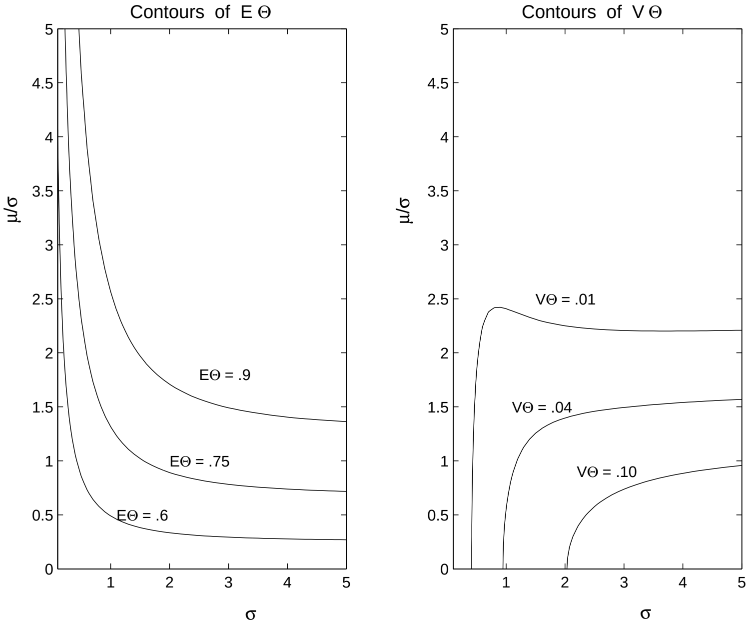

Figure 2: The graph at left shows contours of $(\mu/\sigma, \sigma)$ combinations that support constant values of $E(\Theta)$ at $.6, .75$, and $.9$. The mean $E(\Theta) = .5$ whenever $\mu/\sigma = 0$. At right appear contours of constant values of the variance, $V(\Theta)$, at $.01, .04$ and $.10$.

The displayed functions $E(\Theta) = M(\mu, \sigma^2)$ and $V(\Theta) = V(\mu, \sigma^2)$ are invertible, of course, so that a specification of $E(\Theta)$ would determine a contour of supporting (μ, σ^2) values from Figure 2, and a specification of $V(\Theta)$ would then identify the appropriate single (μ, σ^2) pair, selected from this contour. Once $E(\Theta)$ is specified, it is probably easier to specify $V(\Theta)$ by examining an array of densities that support this mean value, as we shall now see.

Figures 3 and 4 display graphs of some logitnormal densities that support $E(\Theta) = .5$ and $E(\Theta) = .60$, respectively, to indicate what members of this family of densities can look like. It is interesting to see how they are different from the Beta family of densities. When

σ^2 is small, the densities are unimodal and look similar to the Beta(α, β) densities when α and β both exceed 1. However the logitnormal densities always converge to 0 as $\theta \to 0^+$ and as $\theta \to 1^-$, and when σ^2 is large enough the densities are bimodal. In this way they differ markedly from the family of Beta(α, β) densities when α and β are both less than 1. These Beta densities have asymptotes at 0^+ and at 1^-. The logitnormal family contains no members that resemble the Beta densities when $\alpha \leq 1$ and $\beta \geq 1$, or vice versa.

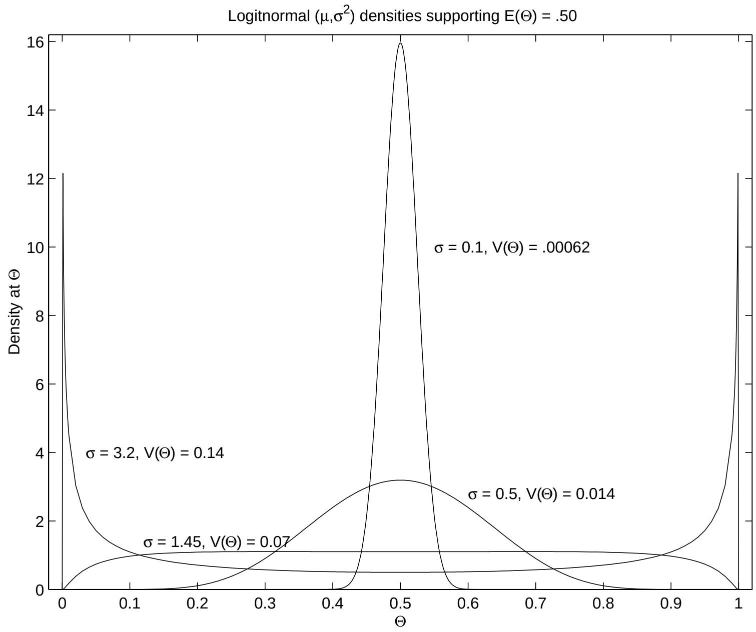

Figure 3: Logitnormal (μ, σ^2) densities with expec tation $E(\Theta) = .5$ are specified by $\mu = 0$ and the values of σ indicated.

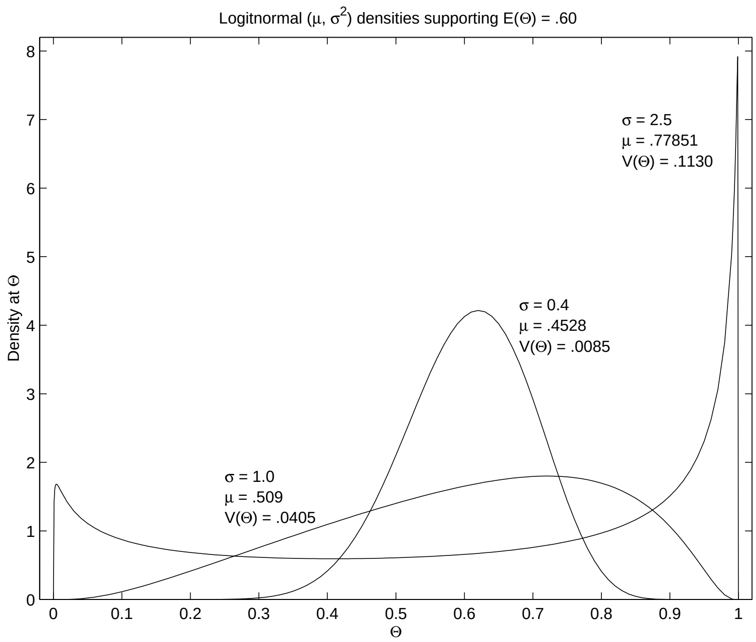

Figure 4: Logitnormal (μ, σ^2) densities with expectation $E(\Theta) = .60$ are specified by various (μ, σ) pairs. Again, higher values of σ specify a bimodal logitnormal density.

Depending on the size of σ^2, a logitnormal density may be unimodal or bimodal. Large values of σ^2 in the Logitnormal (μ, σ^2) specification characterise bimodal distributions for Θ with a large variance; small values of σ^2 specify unimodal distributions with a small variance. For example, in the graph of the densities with mean $E(\theta) = .5$ (meaning the logitnormal parameter $\mu = 0$), the density for θ bunches near 0 and near 1 when $\sigma = 3.2$, while it is concentrated near .5 when $\sigma = 0.1$. Figure 4 shows a similar feature of the family of densities with expectation .60. The extreme values for the variance of a logitnormal density with expectation M are 0 and $M(1-M)$. The upper bound variance agrees with the variance of a Bernoulli distribution, though the limiting member of the Logitnormal family is slightly different, as we shall now see.

2 Limiting family members are delta functions or deploy adherent masses

The limiting members of the Logitnormal (μ, σ^2) family of distributions are unusual. Suppose μ and σ are constrainted to lie on a contour supporting some $E(\Theta) = M(\mu, \sigma^2)$. Then as $\sigma^2 \to 0$, the density approaches a delta function, infinite at $\theta = E(\Theta)$ and 0 elsewhere. At the same time, the distribution function for Θ approaches the step function that indicates $(\theta \geq E(\Theta))$. Here and following we use parenthetic notation to denote an indicator function. Parentheses around any mathematical expression that may be true and may be false denote the function that equals 1 if the interior expression is true, and 0 if it is false.

Under the same constraint that μ and σ lie on a contour supporting a specific value of $E(\Theta)$, as $\sigma^2 \to \infty$ the density for θ approaches a function that equals 0 everywhere on the interval $(0, 1)$, but has an *adherent mass* of $E(\Theta)$ at 1, and $1 - E(\Theta)$ at 0. An alternative terminology is an agglutinated mass. An adherent mass is a feature of finitely additive distributions that was developed by Bruno de Finetti (1955). In the context of a distribution over (0,1) with adherent masses at 0 and 1, it is defined as follows:

Definition: A probability distribution for Θ over (0,1) is said to have adherent masses of $E(\Theta)$ and $1 - E(\Theta)$ at 1 and 0, respectively, if $P(\Theta = 0) = P(\Theta = 1) = 0$, yet for any numbers a and b for which $0 < a < b < 1$, $P[\Theta \in \ (b, 1)] = E(\Theta)$ and $P[\Theta \ \in \ (0, a)] = 1 - E(\Theta)$.

This property of distributions with adherent masses appears unusual, because such distributions are only finitely additive, not countably additive. The image of adherence is that the total probability of 1 does not attach itself to any open intervals that are separated from 0 and 1, because the probability that Θ lies in any open interval inside of $(0, 1)$ is zero. Yet the entire probability of 1 adheres to the end-points of the unit-interval without amassing at these points themselves.

The reason that these masses merely adhere to 0 and 1 in the limiting logitnormal case derives from the fact that the regular family density values always converge to 0 as $\theta \to 0^+$ and as $\theta \to 1^-$. Moreover for any choice of $E(\Theta)$, as $\sigma^2 \to \infty$ the densities converge to 0 at all interior points of the interval as well. Thus, for any finite σ^2 and $\epsilon > 0$, there always exist ordered positive numbers $a_L < a_U < b_L < b_U$ and open intervals $(0, a_L), (0, a_U), (b_L, 1)$, and $(b_U, 1)$ with the properties that $P[\Theta \in (0, a_L) \ or \ \Theta \in (b_U, 1)] < \epsilon$ and $P[\Theta \in (0, a_U) \ or \ \Theta \in (b_L, 1)] > 1 - \epsilon$. As σ^2 increases, the total measure of 1 becomes concentrated in smaller and smaller intervals that are getting moved closer and closer to 0 and 1, but for any choice of σ^2 they are buffered from the endpoints by intervals with negligible probabilities. In the limit, all probability gets pushed off every open interval inside of $(0, 1)$, but it never gets to attach itself to the endpoints 0 and 1.

3 The Multivariate Logitnormal Distribution

Definition: The vector of variables $(\Theta_1, \Theta_2, ...\Theta_N)$ is distributed *Logitnormal* $(\mu, \mathbf{\Sigma})$ if the vector of their logit transformations is distributed Multivariate Normal $(\mu, \mathbf{\Sigma})$.

We limit our exposition here to the bivariate exchangeable distribution, having equal means and variances for the two components. This is the context in which we have recently applied the distribution, and this serves to exemplify a feature of extendible exchangeable distributions. What we need to do is to identify the correlation between the two variables Θ_1 and Θ_2 when the distribution has been specified in terms of the mean, variance, and the correlation *between the logits of these two variables.* Determining this function $Cor(\Theta_1, \Theta_2) = C(\mu, \sigma^2, \rho)$ is a little more intricate than the analysis of the univariate mean and variance.

The correlation between two variables jointly distributed Logitnormal in this way must also be determined by numerical integration. The exchangeable bivariate density function for Θ_1 and Θ_2 is printed in Quintana and Newton (1998). It has the form

$$g(\theta_1, \theta_2) \;=\; \frac{\exp\{-\; Q\,/\,2\;\}}{2\pi\,\sigma^2(1-\rho^2)^{1/2}\,\theta_1\,\theta_2\,(1-\theta_1)\,(1-\theta_2)} \qquad \text{for}\;\; (\theta_1,\theta_2)\;\in\;(\;0,\;1\;)^2 \qquad (2)$$

where

$$Q \;=\; (\sigma^2\,(1-\rho^2))^{-1}\;\;[(logit(\theta_1)-\mu)^2\;+(logit(\theta_2)-\mu)^2\;-\;2\,\rho\,(logit(\theta_1)-\mu)(logit(\theta_2)-\mu)]$$

which is easily derived from the quadratic form of the bivariate normal density of the joint logitnormal, along with the Jacobian of the transformation.

There are specific constraints on the correlation between two variables with an exchangeable bivariate normal distribution that need to be considered. When the covariance matrix of a K-variate normal distribution is expressed in the form $\mathbf{\Sigma} = a\,\mathbf{I} + b\,\mathbf{1}$ where $\mathbf{I}$ and $\mathbf{1}$ are the identity matrix and a matrix of 1's, respectively, then it is required that $b \geq -\,a/K$ in addition to the positivity of a and $a+b$. (See Lad, 1996, p. 387). This implies that the correlation between any two components of the K-variate normal vector must exceed $-1/(K-1)$. In the case of a bivariate distribution that is not exchangeably extendible, the correlation coefficient is not constrained further than the usual constraint that $\mid \rho \mid \leq 1$. However, if it were specified further that the bivariate distribution be exchangeably extendible to dimension K, then the bivariate constraint on any two components would be that $\rho \;\geq -1/(K-1)$. Infinitely extendible exchangeable distributions must have non-negative correlations.

To study the correlations between the two thetas themselves, we have computed the value of $Cor(\Theta_1, \Theta_2)$ for an array of $(\mu/\sigma, \sigma)$ parameters, while specifying various values of ρ, the correlation between the logits of Θ_1 and Θ_2. Numerical results for three (μ, σ) configurations supporting $E\Theta = .6$ appear in Table 1. For each configuration of (μ, σ) pairs, the values of ρ entertained were -.3 through +.8 in gradations of .1. Notice that when the correlation between the logits of Θ_1 and Θ_2 equals 0, the correlation between Θ_1 and Θ_2 is also 0. This is evident from equation 2 when $\rho = 0$, noticing that this function is then a product of two functions in the form of equation 1.

It is also of interest to view some of the exchangeable bivariate densities specified by different values of ρ. I am still not sure which figures we should print, nor in what form. For now I shall just print three figures of symmetric densities specified by $(\mu = .6, \sigma = 2)$ each with a different value of $\rho = -.25, 0$ or $.25$.

4 Remarks

We have had reason to study the Logitnormal distribution on account of its relevance to the problem of combining information elicited from experts in the form of probabilities. This research is reported in DiBacco, Frederic, and Lad (2003). We were surprised to find that the distribution has not been studied directly in the form we have reported here. We surmise that its neglect has been due to the extensive effort that the numerical integrations would have required prior to the contemporary ease of numerical computing.

Acknowledgements

This research was begun while Frank Lad was a visiting professor of statistics at the Institute for Advanced Study, University of Bologna, during May and June, 2002.

Table 1: Correlations between Θ_1 and Θ_2 computed by numerical integrations when the logits of Θ_1 and Θ_2 are distributed exchangeably bivariate normal with the specified values of μ and σ displayed in the headings of columns 2-4, and the correlation value ρ specified in the row beginning with column 1.

ρ	$\mu = .77851, \sigma = 2.5$	$\mu = .509, \sigma = 1$	$\mu = .4528, \sigma = .4$
-.3	-0.266	-0.292	-0.299
-.2	-0.177	-0.195	-0.199
-.1	-0.089	-0.097	-0.100
0	0	0	0
.1	0.089	0.098	0.100
.2	0.179	0.196	0.199
.3	0.270	0.294	0.299
.4	0.363	0.393	0.399
.5	0.458	0.492	0.499
.6	0.556	0.592	0.599
.7	0.657	0.692	0.699
.8	0.764	0.794	0.799

Acknowledgements

This research was begun while Frank Lad was a visiting professor of statistics at the Institute for Advanced Study, University of Bologna, during May and June, 2002.

REFERENCES

DiBacco, M., P. Frederic and F. Lad (2003) Learning from the probability assertions of experts, University of Bologna Research Report, Dipartimento di Statistiche, Bologna.

de Finetti, B. (1955) The structure of distributions on abstract spaces, reprinted in English translation in *Probability, Induction and Statistics*, 1972, New York: John Wiley.

Kass, R. and D. Steffey (1989) Approximate Bayesian inference in conditionally independent hierarchical models (Parametric empirical Bayes models), *Journal of the American Statistical Association*, **84**, pp. 717-726.

Lad, F. (1996) *Operational Subjective Statistical Methods: a mathematical, philosophical, and historical introduction*, New York: John Wiley.

Leonard, T. (1972) Bayesian methods for binomial data, *Biometrika*, **59**, pp. 581-589.

Quintana, F.A. and M. Newton (1998) Assessing the order dependence for partially exchangeable binary sequences, *Journal of the American Statistical Association*, **93**, pp. 194-202.

Figures of some exchangeable bivariate densities

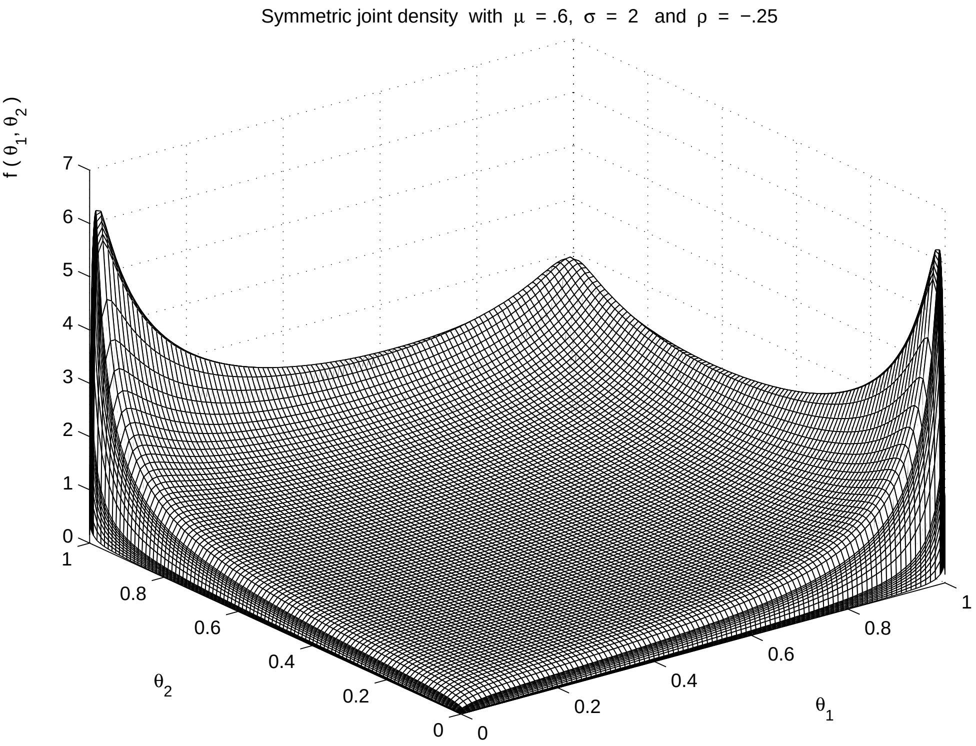

Figure 5: $\rho = -.25$

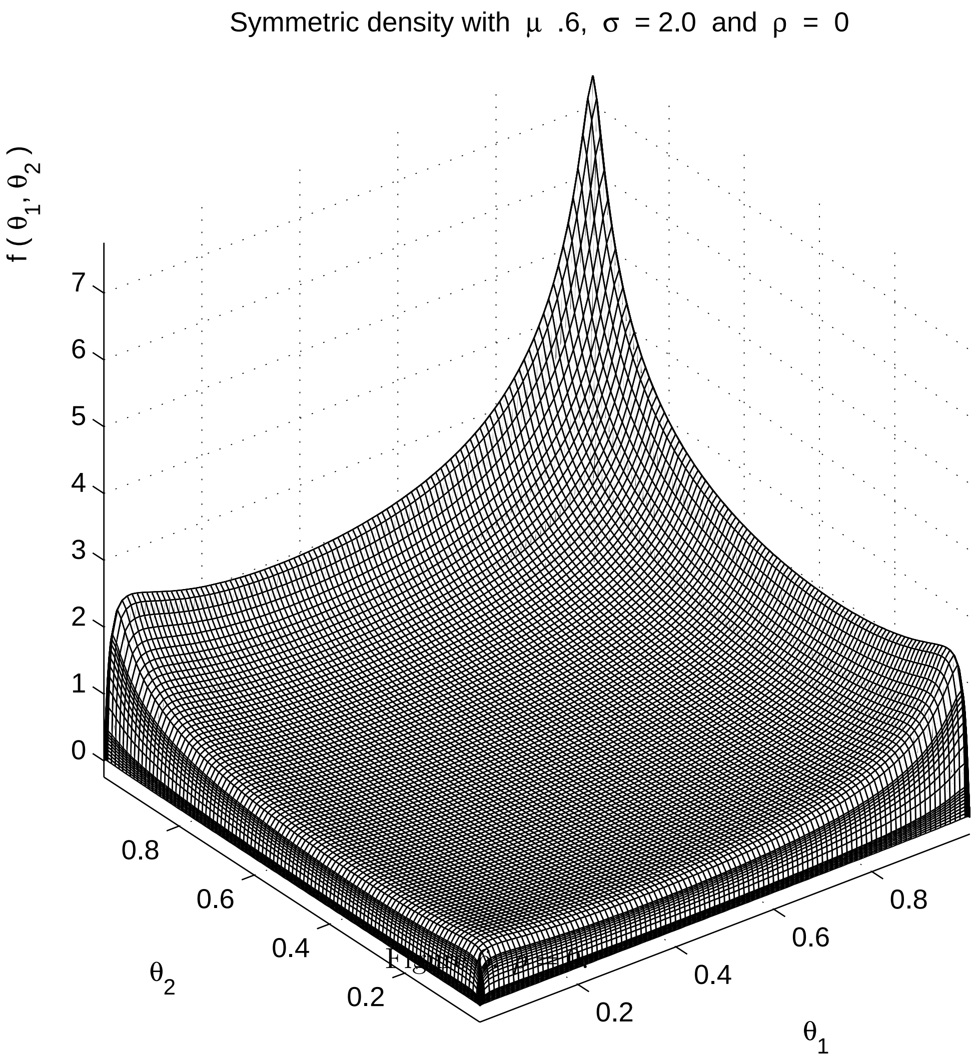

Figure 6: $\rho = 0$.

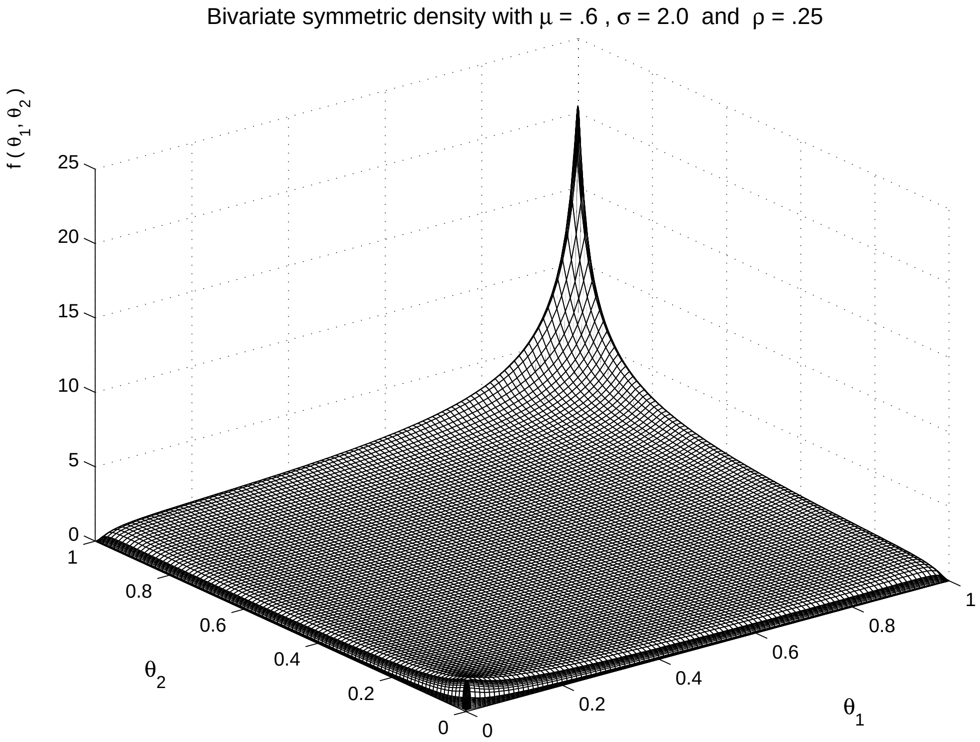

Figure 7: $\rho = .25$.