Two Moments of the Logitnormal Distribution

Patrizio Frederic and Frank Lad

University of Modena, Italy, and University of Canterbury, New Zealand

Abstract

We display the first two moment functions of the Logitnormal(μ, σ^2) family of distributions, conveniently described in terms of the Normal mean, μ, and the Normal signal-to-noise ratio, μ/σ, parameters that generate the family. Long neglected on account of the numerical integrations required to compute them, awareness of these moment functions should aid the sensible interpretation of logistic regression statistics and the specification of "diffuse" prior distributions in hierarchical models, which can be deceiving. We also use numerical integration to compare the correlation between bivariate Logitnormal variables with the correlation between the bivariate Normal variables from which they are transformed.

AMS 2000 Subject Classifications: 60E99, 60-04, 62E15.

Key Words: Numerical integration, S_B distributions, adherent masses, extendible exchangeability, logistic regression, hierarchical models.

Corresponding author: Frank Lad, Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, NEW ZEALAND, 8020.

Email: F.Lad@math.canterbury.ac.nz

Short title for running head: Logitnormal moments

1 Introduction and background

An analytic characterisation of the Logitnormal family of distributions first appeared in the pathbreaking paper of Johnson (1949a). The features of this important family of distributions are unknown to many practicing statisticians because they are classified under the minimally informative name of S_B distributions. Following a suggestion of Edgeworth, Johnson characterised these distributions in terms of a "system of curves" constructed from transformations of variables that are distributed Normal. This development occurred in the tradition of the Pearson system and the Charlier system of curves, which were already current at the time. The appellation S_B family is due to the fact that these distributions have a "bounded domain", which distinguishes them from other families. In the case of the Logitnormal distributions, the domain is [0,1], while a four-parameter specification of the family allows the endpoints to vary. While distinguishing this family from others characterised over unbounded or partially bounded domains, the nomenclature disguises the relevance of the details of this family to many important areas of applied statistics.

While the Lognormal Distribution is accorded a Chapter in the classic compilation of Johnson, Kotz and Balakrishnan (1994), the Logitnormal is found within a sub-subsection of (General) results in Chapter 12. The Logitnormal does not appear in any form in the popular Wikipedia, nor are its important properties mentioned in the widely-used text of Agresti (1996) in the relevant Chapter on Logistic Regression.

In the Appendix to his original article, Johnson (1949a, p. 174) remarked that while the moments of this family of distributions are complicated in form, their numerical computation is "straightforward though tedious." A few computations were displayed in a Table.

Our present note merely augments Johnson's extensive accomplishments of research on this family of distributions that have been known for more than half a century. The minimal theoretical detail we review serves only for background. We provide a graphical display of the first two moment functions over the parameter space, conveniently specified in terms of the signal-to-noise ratio, μ/σ, and the standard deviation, σ, of the Normal family from which the Logitnormal distributions are derived. The systematic numerical integrations that provide these Figures and their constant contours are readily programmed using R or MATLAB or any number of software systems.

2 The Logitnormal density and its moments

A variable X is distributed *Logitnormal* (μ, σ^2) over the interval $[0, 1]$ if its logit transformation, $log[X/(1-X)]$, is distributed Normal (μ, σ^2). The density function for X can be derived from the Normal density for $log[X/(1-X)]$ using standard transformation methods. If $X \sim Logitnormal(\mu, \sigma^2)$ then its density function is

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

This density function for X is not symmetric unless $\mu = 0$, in which case $E(X) = 1/2$. Generally, the expectation and variance for X are determined by functions of the Normal parameters, $M(\mu, \sigma^2)$ and $V(\mu, \sigma^2)$, that do not have algebraic solutions in closed form. However, we can easily compute values for these functions using numerical integration.

The left side of Figure 1 displays the mean of the Logitnormal(μ, σ^2) densities as a function of the signal to noise parameter, μ/σ, and σ; the right side of the Figure displays the variances of Logitnormal densities using the same arguments. The left side of Figure 2 displays contours of $(\mu/\sigma, \sigma)$ pairs that support constant values of the mean at $E(X) = 0.6, 0.75$, and 0.9, while the right side shows contours supporting values of the variance at $V(X) = 0.01, 0.04$, and 0.10. This display is limited to non-negative values of μ/σ due to the obvious result that if $X \sim Logitnormal(\mu, \sigma^2)$, then $(1-X) \sim Logitnormal(-\mu, \sigma^2)$. The functions $E(X) = M(\mu, \sigma^2)$ and $V(X) = V(\mu, \sigma^2)$ are invertible, so that a specification of $E(X)$ would determine a contour of supporting $(\mu/\sigma, \sigma)$ pairs as shown in the left side of Figure 2, and a specification of $V(X)$ would then identify the appropriate single $(\mu/\sigma, \sigma)$ pair, selected from this contour. In applications involving the specification of a prior distribution, once $E(X)$ is specified, it is usually easier to specify $V(X)$ by examining an array of densities that support this mean value, as we shall now see.

Figure 3 displays graphs of three Logitnormal densities that support $E(X) = .60$, to indicate what members of this family of densities can look like. (The three densities in Figure 3 correspond to the three parameter pairs identified by small circles along the $E(X) =$ .6 contour on the left side of Figure 2.) It is interesting to see how the Logitnormal densities 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 $x \to 0^+$ and as $x \to 1^-$. 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 unbounded asymptotes at 0^+ and at 1^-. The Logitnormal family contains no members that resemble the Beta(α, β) densities when $\alpha \leq 1$ and $\beta \geq 1$, or vice versa. These characteristics feature in Appendix B of the wider-ranging book of Kotz and van Dorp (2004). The authors allude to the difficulty of evaluating the moment integrals. However, their discussion focuses largely on moment estimation procedures which are similarly cumbersome.

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 X with a large variance; small values of σ^2 specify unimodal distributions with a small variance. Specifically, if $X \sim Logitnormal(\mu, \sigma^2)$ then the derivative of the density equals 0 at solutions to the transcendental equation

$$log\,(\tfrac{x}{1-x}) \;=\; \sigma^2\;(2x-1)\;+\;\mu\;\;.$$

There are either one or three solutions for x depending on the relative sizes of μ and σ^2. When there are three, the first and third support modal points while the second supports a relative minimum of the density. This characterisation of bimodality is explicit in Johnson (1949a, pp. 158-159).

Two of the three example densities with $E(x) = 0.6$ displayed in Figure 3 are unimodal with small values of σ. Logitnormal densities become bimodal when σ is large enough. The scales of the variances for these three densities (noted next to their graphs) are recognisable from the variance contours and the position of the same three density identifying points which are circled on the right side of Figure 2.

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. The Bernoulli arises as the limit of the Logitnormal family as $\sigma^2 \to \infty$. Interestingly, in this form the limiting density displays adherent masses (agglutinated masses) at 0 and 1. This analytic concept was developed in de Finetti (1955).

3 Bivariate Logitnormal Correlations

The bivariate S_{BB} distribution was also introduced in Johnson (1949b), defining a bivariate Logitnormal distribution. A vector of variables $(X_1, X_2, ...X_N)$ is distributed *Logitnormal* $(\mu, \mathbf{\Sigma})$ if the vector of their logit transformations is distributed Multivariate Normal $(\mu, \mathbf{\Sigma})$. Our contribution here is merely to compare the numerical relation between the correlation of two Logitnormal variables with the correlation between their bivariate Normal transformations. Our specific computations pertain to exchangeable bivariate distributions, having two equal means and variances. This is the context in which we have recently applied the distribution, and this serves to exemplify a feature of extendible exchangeable distributions. We again used numerical integration to identify the correlation between the two variables X_1 and X_2 when their joint distribution is specified as bivariate Logitnormal$(\mu_1 = \mu_2 \equiv \mu, \sigma_1^2 = \sigma_2^2 \equiv \sigma^2, \sigma_{12} = \rho\sigma^2)$.

The exchangeable bivariate density function for X_1 andX_2 was printed in Quintana and Newton (1998). It has the form

$$g(x_1, x_2) \;=\; \frac{\exp\{-\;Q\,/\,2\;\}}{2\pi\,\sigma^2(1-\rho^2)^{1/2}\,x_1\,x_2\,(1-x_1)\,(1-x_2)} \qquad \text{for} \;\; (x_1, x_2) \;\in\; (\;0,\;1\;)^2 \qquad (2)$$

where

$$Q \;=\; (\sigma^2\,(1-\rho^2))^{-1}\;\;[(logit(x_1)-\mu)^2 + (logit(x_2)-\mu)^2 - 2\,\rho\,(logit(x_1)-\mu)(logit(x_2)-\mu)]\;.$$

This result 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 extend*ible* to dimension K, then the bivariate correlation is still constrained by $\rho \geq -1/(K-1)$. Infinitely extendible exchangeable distributions must have non-negative correlations.

To study the correlations between the two **variables** themselves, we have computed the value of $Cor(X_1, X_2)$ for an array of $(\mu/\sigma, \sigma)$ parameters, while specifying various values of ρ, the correlation between the logits of X_1 and X_2. Numerical results for three (μ, σ) configurations supporting $EX = 0.6$ appear in Table 1. For each configuration of (μ, σ) pairs, the values of ρ entertained were -0.3 through +0.8 in gradations of 0.1. We choose this range because in the application that motivated this work, the bivariate distribution was exchangeably extendible to dimension 3.

Notice that when the correlation between the logits of X_1 and X_2 equals 0, the corre-

Table 1: Correlations between X_1 and X_2 computed by numerical integrations when the logits of X_1 and X_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 column 1 of each row.

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

lation between X_1 and X_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. Although the transformations of the variables X_i to their logits is not linear, the correlations between X_1 and X_2 and between their logit transforms are quite close over a broad range of μ, σ and ρ parameters. The fact that the correlations between the Logitnormal variables never exceeds the correlation between the Normal variables from which they are transformed exemplifies a theorem of Gebelein, as reported by Koyak (1987, pp. 1215-1216). **That is, the maximal correlation between all possible transformations of two Normal random variables is the simple correlation between the variables themselves. This is not true of all bivariate random variables. General multivariate results of this type have been developed in the theory of maximal canonical correlation coefficients.**

4 Remarks

The Logitnormal distribution has been used in hierarchical Bayesian modelling as early as Leonard (1972) and in many applications such as that of Kass and Steffey (1989). In these cases it was 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. The direct analysis of the moments using numerical integration yields a clear understanding of the possible choice of a family member as a mixing distribution in applications, avoiding misrepresentations. For example, it is common to select a large value of σ^2 for a Logitnormal specification, supposedly to represent diffuse prior information. To the contrary, in many cases such a choice would represent rather strong prior information that the relevant proportion is very likely to be close to 0 or to 1, as has been exhibited here. Caution and understanding are important in this regard.

Another useful aspect of the results presented here pertains to general applications of logistic regression. The use of statistical results for subsequent forecasting would benefit from a graphical display of the conditional forecasting distributions that are appropriate to estimated regression coefficients and the variance.

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).

Acknowledgments

This research was conducted while Frank Lad was a visiting professor of statistics at the Institute for Advanced Study, University of Bologna, Department of Statistics. Thanks especially to an unknown referee and to Wes Johnson and Val Johnson for helpful comments.

REFERENCES

Agresti, A. (1996). *An Introduction to Categorical Data Analysis.* New York: Wiley.

DiBacco, M., P. Frederic and F. Lad (2003). Learning from the probability assertions of experts, University of Canterbury Department of Mathematics and Statistics Research Report, www.math.canterbury.ac.nz

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

Johnson, N.L. (1949a). Systems of frequency curves generated by methods of translation. Biometrika 36:149-176.

Johnson, N.L. (1949b). Bivariate distributions based on simple translation systems. Biometrika 36:297-304.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions.* New York: 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 84:717-726.

Kotz, S. and van Dorp, J.R. (2004). *Beyond Beta: other continuous families of distributions with bounded support and applications.* New Jersey: World Scientific.

Koyak, R.A. (1987), On measuring internal dependence in a set of random variables. Annals of Statistics 15:1215-1228.

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

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

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

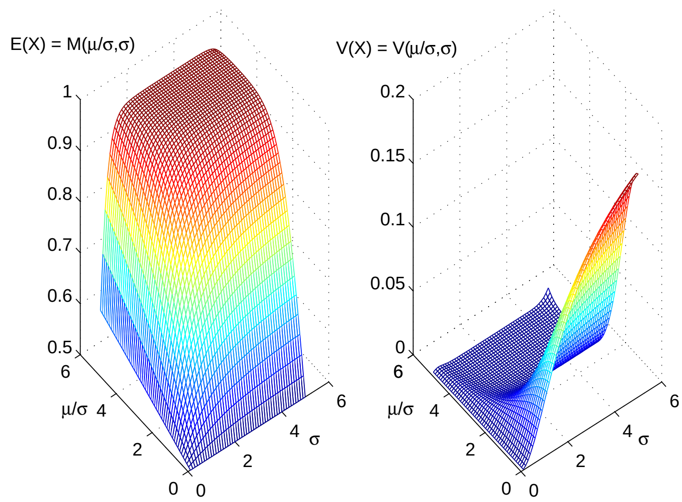

Figure 1: The graph at left depicts values of $E(X)$ as a function of the Logitnormal μ/σ and σ parameters. At right appears the Variance of X as a function of these same parameters.

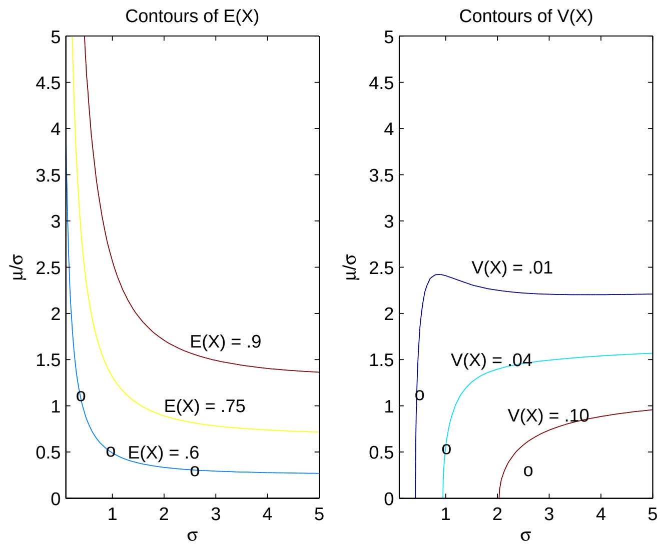

Figure 2: The graph at left shows contours of $(\mu/\sigma, \sigma)$ combinations that support constant values of $E(X)$ at $0.6, 0.75$, and 0.9. The mean $E(X) = .5$ whenever $\mu/\sigma = 0$. At right appear contours of constant values of the variance, $V(X)$, at $0.01, 0.04$ and 0.10.

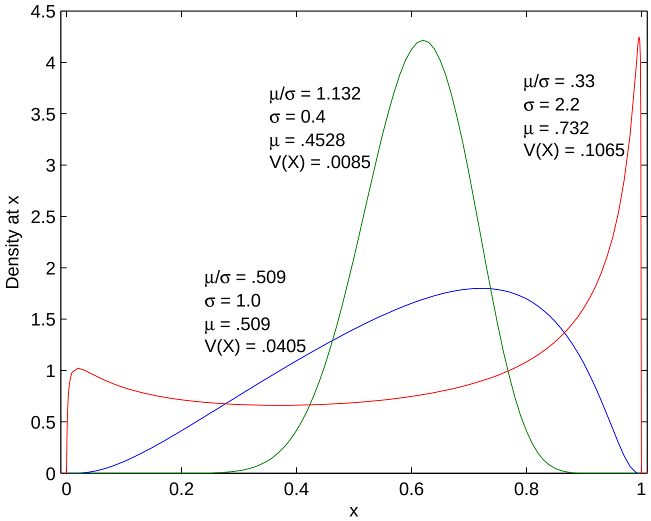

Figure 3: Logitnormal (μ, σ^2) densities with expectation $E(X) = 0.60$ are specified by various (μ, σ) pairs. The largest values of σ specify a bimodal Logitnormal density.