International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo. June 28-30, 2006

The Polygonal Distribution

Dimitris Karlis
Department of Statistics,
Athens University of Economics and Business.
Evdokia Xekalaki*
Department of Statistics,
Athens University of Economics and Business
and
Department of Statistics,
University of California, Berkeley CA, USA.

Abstract

The triangular distribution, although simpler than the beta distribution both for mathematical treatment and for natural interpretation, has not been widely used in the literature as a modelling tool. Applications of this distribution as an alternative to the beta distribution appear to be limited in financial contexts and specifically in the assessment of risk and uncertainty and in modelling prices associated with trading single securities. One of the basic reasons is that it can have only a few shapes. In this paper, a new class of distributions stemming from finite mixtures of the triangular distribution is introduced. Their polygonal shape makes them appealing for modelling purposes since they can be used as simple approximations to several distribution functions. Properties of these distributions are studied and parameter estimation is discussed. Further, the distributions arising when using the triangular distribution as an approximation to the beta distribution, as the mixing distribution in the case of two well-known beta mixtures, the beta-binomial and the beta-negative binomial distribution are examined.

Key Words: Triangular Distribution; Binomial mixtures; Negative Binomial Mixtures; Triangular-Binomial Distribution

1 Introduction

The probability density function (pdf) of the triangular distribution is given by

$$f(x \mid \theta) = \begin{cases} \frac{2x}{\theta}, & 0 \le x \le \theta \\ \frac{2(1-x)}{1-\theta}, & \theta \le x \le 1 \\ 0, & \text{elsewhere} \end{cases} \tag{1.1}$$

*Correspondence to: Evdokia Xekalaki, Department of Statistics, Athens University of Economics and Business, 76 Patision St., 104 34, Athens, Greece.
Email: exek@aueb.gr

with “0/0” interpreted as 1. The above definition restricts the random variable X in the interval $[0, 1]$. One can define in a similar manner triangular distributions in a finite interval $[\alpha, \beta]$ by considering the transformation $Y = \frac{X-\alpha}{\beta-\alpha}$. From (1.1), one can see that the density is linearly increasing in the interval $[0, \theta)$ and linearly decreasing in the interval $[\theta, 1]$ (θ is the mode of the distribution) . The distribution is not symmetric except for the case $\theta = 1/2$. The parameter θ is allowed to take the values 0 and 1, using the appropriate part of the definition given in (1.1). More details about the triangular distribution can be found in van Dorp and Kotz (2004) and the references therein. Johnson (1997) and Johnson and Kotz (1999) refocused interest in the triangular distribution, which appeared to have been ignored as a modeling tool over the last decades, one of the most probable basic reasons being that it can have only a few shapes.

In this paper, a new class of distributions is introduced stemming from finite mixtures of the triangular distribution. Contrary to the triangular distribution, the members of this class have a shape flexibility that makes them appealing for modeling purposes. Because of their shape, which is polygonal, these distributions are termed in the sequel polygonal distributions.

The paper is organized as follows. In section 2, the polygonal distribution is defined as a finite mixture of triangular distributions. Properties of it and estimation are discussed. In section 3, mixture distributions arising when using the triangular as an approximation to a beta mixing distribution are examined. In particular, the cases of beta mixtures of binomial and negative binomial distributions are considered.

2 The Polygonal Distribution

Let $f_i(x \mid \theta_i),\ i = 1, 2, \ldots, k$ be the densities of k independent triangular variables on $[0, 1]$ with parameters $\theta_i,\ i = 1, \ldots, k$. The density of the polygonal distribution is defined as the mixture of these densities given by

$$f_k(x) = \sum_{i=1}^{k} p_i f_i(x \mid \theta_i) \tag{2.1}$$

with mixing proportions $p_i > 0$ satisfying $\sum_{i=1}^{k} p_i = 1$.

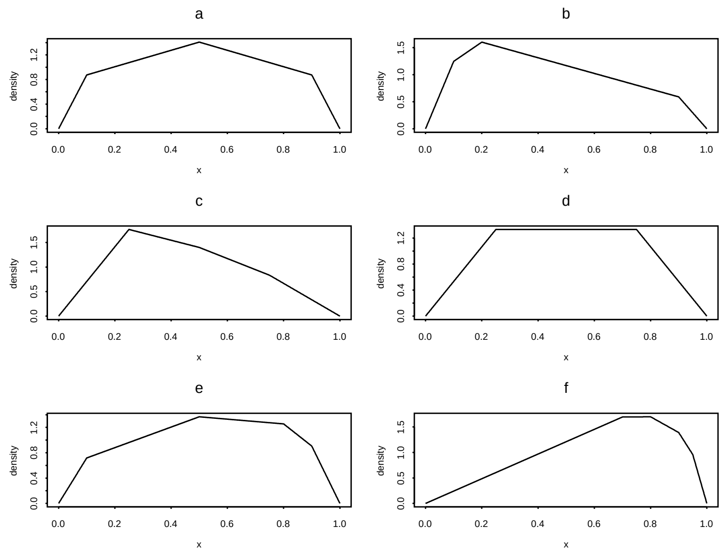

Figure 1: Examples of polygonal distributions. The depicted densities correspond to the following choices of parameter values: (a) $p_1 = p_2 = p_3 = 1/3$ *and* $\theta_1 = 0.1, \theta_2 = 0.5, \theta_3 = 0.9$*, (b)* $p_1 = 0.4, p_2 = 0.4, p_3 = 0.2$ *and* $\theta_1 = 0.1, \theta_2 = 0.2, \theta_3 = 0.9$*, (c)* $p_1 = 0.8, p_2 = 0.1, p_3 = 0.1$ *and* $\theta_1 = 0.25, \theta_2 = 0.5, \theta_3 = 0.25$*, (d)* $p_1 = p_2 = 0.5$ *and* $\theta_1 = 0.25, \theta_2 = 0.75$*, (e)* $p_1 = p_2 = p_3 = p_4 = 1/4$ *and* $\theta_1 = 0.1, \theta_2 = 0.5, p_3 = p_4 = 1/4$ *and* $\theta_1 = 0.7, \theta_2 = 0.8, \theta_3 = 0.9, \theta_4 = 0.95$.

This distribution has a polygonal form with at most k points of inflection. Some members of this family are depicted in figure 1. Note that the densities are piecewise linear. This enables one to define a broad family of distributions. When components are close together and their number becomes larger and larger, the density approaches a smooth curve. The appealing feature of this density is that it can take shapes, which are not common in other distributions. For example, a 2-polygonal distribution with $p = 0.5$ and $\theta_1 = 0.25, \theta_2 = 0.75$ is flat over the interval $(0.25, 0.75)$, having a modal interval, rather than a mode.

In the sequel, the distribution with k triangular components defined above is interchangeably referred to as the k-polygonal distribution or as the polygonal distribution. We also assume for simplicity that its components are ordered with respect to the values of their parameters θ_i.

The mean and variance of the polygonal distribution are given by

$$E(X) = \frac{1}{3} + \frac{1}{3}\sum_{j=1}^{k} p_j\theta_j \qquad \text{and}$$

$$Var(X) = \frac{1}{6} - \frac{1}{18}\sum_{j=1}^{k} p_j\theta_j + \frac{1}{6}\sum_{j=1}^{k} p_j{\theta_j}^2 - \frac{1}{9}\left(\sum_{j=1}^{k} p_j\theta_j\right)^2,$$

respectively. It can be seen that the mean lies in the interval $(1/3, 2/3)$.

Further, it can be easily verified that the polygonal distribution has always a mode. For details see Karlis and Xekalaki (2000). The position of the mode depends on the mixing proportions and it is not easy to be determined for general k. However, the mode is necessarily one of the $\theta's$ or a modal interval from θ_j to θ_{j+1}.

For the mode of the 2-polygonal distribution, in particular, the following result holds

Proposition 2.1. *For a 2-polygonal distribution we have that: If* $p_1 \leq \frac{1-\theta_1}{\theta_2-\theta_1+1}$ *the mode is at the point* θ_2*, otherwise the mode is at the point* θ_1 *. When equality holds there is a modal interval instead of a mode.*

A proof can be found in Karlis and Xekalaki (2000)

2.1 Estimation

ML estimation can be carried out using the finite mixture representation via an EM algorithm. This is comprised of the following steps.

Step 1 (E-step): Given the current values for the parameters, say ${\theta_j}^{old}$ and ${p_j}^{old}$, $j = 1, \ldots, k$, calculate

$$w_{ij} = \frac{{p_j}^{old} f(x_i \mid {\theta_j}^{old})}{f_k(x_i)},$$

where $f(x \mid \theta)$ and $f_k(x)$ are given in (1.1) and (2.1) respectively.

Step 2 (M-step): Update p_j by

$$p_j^{new} = \frac{\sum_{i=1}^{n} w_{ij}}{n}.$$

For each component j, update θ_j solving a weighted likelihood problem, namely by maximizing, the weighted likelihood

$$L_j(\theta) = \sum_{i=1}^{n} w_{ij} \log f(x_i \mid \theta)$$

for $j = 1, \ldots, k$

The maximization can be easily carried out, since the solution is one of the observations and thus evaluating L_j at all the observation points, suffices to locate the maximum. Note that the weights w_{ij} do not depend on the estimate and hence the monotonocity of the likelihood holds as in Johnson and Kotz (1999).

3 The Triangular Distribution as a Mixing Density

3.1 The Binomial - Triangular Distribution

The binomial distribution is a prominent member of the family of discrete distributions. Mixtures of the binomial distribution with respect the parameter p have been considered in the literature. Such mixtures have probability functions of the form

$$P(X = x) = \binom{n}{x} \int_0^1 p^x (1-p)^{n-x} dG(p), \qquad x = 0, 1, \ldots, n. \quad (3.1)$$

Note that $G(p)$ denotes a generic mixing distribution which can be either a finite step distribution giving positive probabilities at only a finite number of points or a continuous distribution. Some identifiability problems arise for small values of n (see, for example, Follmann and Lambert (1991)). The distribution is identifiable only up to the first n moments of the mixing distribution.

The beta-binomial is the best known member of the family of binomial mixture distributions. It arises when the parameter p follows a beta distribution (see, for example, Tripathi and Gurland (1994) and the references therein).

Only a few other binomial mixtures have been developed, mainly due to numerical difficulties (see Alanko and Duffy (1996), Horsnell (1957), Brooks et al. (1997)).

Assume that the parameter p has a triangular distribution given in (1.1). Then the resulting probability function is given by

$$P(X=x) = 2\binom{n}{x}\left(\frac{1}{\theta}\int_0^\theta p^{x+1}(1-p)^{n-x}dp + \frac{1}{1-\theta}\int_\theta^1 p^x(1-p)^{n-x+1}dp\right). \tag{3.2}$$

Both integrals are in fact incomplete beta integrals (see Abramowitz and Stegun, 1974) defined as $B_x(\alpha,\beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$. Using the represeantation

$$I_x(\alpha,\beta) = \frac{B_x(\alpha,\beta)}{B(\alpha,\beta)},$$

with $I_x(\alpha,\beta) = 1 - I_{1-x}(\beta,\alpha)$ and $I_x(\alpha,\beta) = xI_x(\alpha-1,\beta) + (1-x)I_x(\alpha,\beta-1)$, and after tedious algebraical manipulations one can write the probability function of the binomial-triangular (B-T) distribution as

$$\begin{aligned} P(X=x) &= 2\binom{n}{x}(\theta^{-1}B_\theta(x+2,n-x+1) + \\ &+ (1-\theta)^{-1}B(x+1,n-x+2) + (1-\theta)^{-1}B_\theta(x+1,n-x+2)). \end{aligned}$$

This probability function is quite awkward for calculations as it involves incomplete beta functions. One can improve by considering recurrence relationships for beta integrals and incomplete beta integrals. A simpler method can be used for calculating the probabilities, based on a finite series representation of the probability mass function.

Sivaganesan and Berger (1993) showed that for a general $G(p)$ the resulting mixed Binomial distribution can be written as

$$P(X=k) = \sum_{j=k}^{n} h(j,k)E(p^j), \qquad k = 0,1,\ldots,n, \tag{3.3}$$

where $h(j,k) = (-1)^{j-k}\frac{n!}{k!(j-k)!(n-j)!}$ for $j \geq k$ and 0 if $j < k$, $E(p^r)$ denotes the $r-th$ simple moment of the mixing distribution. For the case of the triangular distribution, we obtain

$$P(X=k) = \sum_{j=k}^{n} h(j,k)\frac{2\sum\limits_{i=0}^{j}\theta^i}{(j+1)(j+2)} \qquad , k = 0,1,\ldots,n. \tag{3.4}$$

Computationally, this form is particularly convenient,since the coefficients $h(j,k)$ can be easily computed recursively using

$$h(0,0)=1, \qquad h(j+1,j+1)=\frac{n-j}{j+1}h(j,j)$$

for $j=0,1,\ldots,n$ and

$$h(j+1,k)=-\frac{n-j}{j-k+1}h(j,k), \qquad j=k=,\ldots,n-1,$$

while the moments of the triangular distribution can be derived recursively. Evaluation of the probability function using the above form is easy and inexpensive. Even for large values of n near 200, no overflows were encountered for all the entire range of θ.

The mean and the variance of the B-T distribution are given by

$$E(X)=nE(p)=\frac{n(1+\theta)}{3} \quad \text{and} \quad Var(X)=\frac{n(n+3)}{18}-\frac{n(n-3)\theta(1-\theta)}{18},$$

respectively. Note that there is a symmetry analogous to that existing in the case of the simple binomial distribution. So if X follows a B-T distribution with parameter θ, then $Y=1-X$ follows a B-T distribution with parameter $1-\theta$.

The simple moments can be derived easily from the simple moments of the binomial distribution. It holds, in particular, that the simple moments of the B-T distribution are given by

$$\mu_r=2\sum_{j=0}^{r}\frac{S(r,j)n!}{(n-r)!}\frac{\sum\limits_{i=0}^{j}\theta^i}{(j+1)(j+2)},$$

where $S(r,j)$ denote the Stirling numbers of the second kind .

Moment estimates of the parameter θ can be obtained through equating the mean with the sample mean. This yields $\hat{\theta}=3n^{-1}\bar{x}-1$, which leads to parameter estimates whenever $\bar{x}$ is in the range $(n/3,2n/3)$. The variance of the moment estimator is given by $Var(\hat{\theta})=\frac{9}{n^2}\frac{Var(X)}{N}$, where N denotes the sample size.

3.2 The Negative Binomial - Triangular Distribution

The triangular distribution can be also used as the mixing distribution for some other discrete distributions, having a parameter defined in the interval $[0, 1]$. Such examples are the geometric and the negative binomial distributions. The negative binomial distribution has probability function given by

$$P(X = x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)x!} p^{\alpha}(1 - p)^x, \quad x = 0, 1, \ldots, n, \alpha > 0, 0 \leq p \leq 1. \tag{3.5}$$

Mixtures of the negative binomial distribution with respect the parameter p can be developed by allowing the paremeter p to vary according to some distribution $G(p)$. Such a mixture has probability function of the form

$$P(X = x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)x!} \int_0^1 p^{\alpha}(1 - p)^x dG(p), \qquad x = 0, 1, \ldots, n. \tag{3.6}$$

The literature on mixtures of the negative binomial is rather sparse. Note that one can define mixtures with respect to either of the parameters α and p. Allowing $G(p)$ to have a *beta*(α, β) form, the generalized Waring distribution arises (see for example, Xekalaki (1983)).

For a general mixture of the negative binomial distribution, one can see that

$$P(X = x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)x!} \int_0^1 p^{\alpha}(1 - p)^x dG(p).$$

Expanding $(1 - p)^x$, one obtains that

$$\int_0^1 p^{\alpha}(1 - p)^x dG(p) = \int_0^1 p^{\alpha} \sum_{k=0}^{x} \binom{x}{k} (-1)^{x-k} p^{x-k} dG(p) =$$

$$= \int_0^1 \sum_{k=0}^{x} \binom{x}{k} (-1)^{x-k} p^{\alpha+x-k} dG(p)$$

$$= \sum_{k=0}^{x} \binom{x}{k} (-1)^{x-k} \int_0^1 p^{\alpha+x-k} dG(p) = \sum_{k=0}^{x} \binom{x}{k} (-1)^{x-k} E(p^{\alpha+x-k}),$$

thus leading to

$$P(X = x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)x!} \sum_{k=0}^{x} \binom{x}{k} (-1)^{x-k} E(p^{\alpha+x-k}).$$

In other words, the probability density function can be written as a finite series of non-integral moments of the mixing distribution.

Assuming a triangular distribution as a mixing distribution, one obtains the negative binomial-triangular distribution with probability function given by

$$P(X = x) = \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)x!} \sum_{k=0}^{x} \binom{x}{k} \frac{(-1)^{x-k} 2(1 - \theta^{\alpha+x-k+1})}{(\alpha + x - k + 1)(\alpha + x - k + 2)(1 - \theta)}.$$

The above formula can be used for calculating the probability function. A similar scheme as the one proposed for the binomial-triangular distribution is applicable. However, since now the values of x are not restricted in a finite range, minor anomalies may be found at the tail. Alternatively, the probability function can be written via incomplete beta functions in a similar manner as for the simple binomial case.

Setting $\alpha = 1$, a geometric-triangular mixture is obtained. Now the moments used are of integral order and thus the recursive relationships for the moments of the triangular distribution can be used. Similar is the case when the Pascal distribution is considered.

The mean of the negative binomial-triangular distribution is

$$E(X) = \int_0^1 \frac{\alpha(1 - p)}{p} g(p)dp = \alpha E(p^{-1}) - \alpha = \alpha \left(\frac{-2\log(\theta)}{1 - \theta} - 1 \right).$$

Since $0 \leq \theta \leq 1$, it holds that $E(X) > \alpha$ for every value of θ. The variance does not exist, since it involves the second inverse moment of the triangular distribution which does not exist. The distribution exhibits a very long tail.

Note that mixtures of the negative binomial distribution with respect to the parameter p are in fact mixtures of the Poisson distribution, with mixing distribution a gamma mixture. For example, the negative binomial-triangular distribution defined above is a Poisson mixture with mixing distribution the mixture of a Gamma density with a triangular mixing density.

References

ALANKO, T. and DUFFY, J. (1996). Compound binomial distributions for modelling consumption data. *The Statistician,*, 45:269–286.

BROOKS, S. P., MORGAN, B. J. T., RIDOUT, M. S., and PACK, S. E. (1997). Finite mixture models for proportions. *Biometrics*, 53:1097–1115.

FOLLMANN, D. and LAMBERT, D. (1991). Idenitfiability of finite mixtures of logistic regression models. *Journal of Statistical Planning and Inference*, 27:375–381.

HORSNELL, G. (1957). Economic acceptance sampling schemes. *Journal of the Royal Statistical Society, Series A*, 120:148–191.

JOHNSON, D. (1997). The triangular distribution as a proxy for the beta distribution in risk analysis. *The Statistician*, 46:387–398.

JOHNSON, N. and KOTZ, S. (1999). Non-smooth sailing or triangular distributions revisited after some 50 years. *The Statistician*, 48:179–187.

KARLIS, D. and XEKALAKI, E. (2000). On some distributions stemming from the triangular distribution. *Technical Report, Department of Statistics, Athens University of Economics*, 111.

SIVAGANESAN, S. and BERGER, J. O. (1993). Robust Bayesian analysis of the binomial empirical Bayes problem. *The Canadian Journal of Statistics*, 21:107–119.

TRIPATHI, G. R., R. and GURLAND, J. (1994). Estimation of the parameters in the beta binomial model. *Annals of the Institute of Statistical Mathematics*, 46:317–331.

VAN DORP, J. and KOTZ, S. (2004). *Beyond Beta: Other continuous distributions with bounded support and applications*. World Scientific, New York.

XEKALAKI, E. (1983). The univariate generalized Waring distribution in relation to accident theory: Proneness, spells or contagion? *Biometrics*, 39:887–895.