Quaderni di Statistica
Vol. 5, 2003

On the Moments of a Mixture of Uniform and Shifted Binomial random variables

Domenico Piccolo
Dipartimento di Scienze Statistiche, Università di Napoli Federico II
E-mail: domenico.piccolo@unina.it

Summary: In this note, we study the first four moments of the MUB random variable, that is a mixture of two discrete random variables, recently introduced for the fitting of ranks data models. After a brief review of the location and variability indexes, the paper derives and discusses the asymmetry and the kurtosis measures, investigating the shape properties of this distribution on the admissible parametric space. Finally, the usefulness of the parameters moment estimators is shown in order to get starting values for the maximum likelihood estimation procedure.

Keywords: Moments, MUB random variable, Asymmetry, Kurtosis.

1. Introduction

In a series of recent works, it has been established the usefulness of a mixture distribution for the modelling of ranks. This approach helps in the analysis and the interpretation of statistical data arising from preferences or evaluation contexts.

In this area, a stochastic model can be generated by the convex combination of the probability distributions of a discrete Uniform and a shifted Binomial random variables, both defined on the support $\{1, 2, \dots\ m\}$. This model has been called MUB distribution. It allows many distributional shapes depending on the parameter values; thus, it seems relevant to study their interpretative content via the first four moments evaluations.

From inferential and computational point of views, maximum likelihood (ML) parameters estimation is obtained in an efficient way via the E-M algorithm, as proposed by D'Elia (2003) and Piccolo (2003). However, the convergence of the algorithms involved in the E-M estimation towards the final estimates may be significantly accelerated starting from some preliminary consistent values[1]. Thus, the introduction of the moments estimates is a significant issue for this aim.

All these considerations support the opportunity to deepen the interpretation of the parameters, their relationships with the moments and their role in the specification of the structure and the shape of the MUB distribution.

The paper is organized as follows. In the next section, we introduce the main probabilistic structure for the MUB random variable and discuss some fundamental issues about the meaning of the parameters, while section 3 is devoted to notations and some algebraic results. Then, in section 4 we discuss the expectation and the variance of such model, and in section 5 we derive an asymmetry measure; some relationships about these moments are also presented. In section 6, a formula for the kurtosis is obtained and the results are commented with regard to the shape of the distribution. Then, in section 7 consistent parameters estimators are derived from the first two sample moments; their usefulness as initial values is discussed with the empirical support of several models estimated from a real data set. Some concluding remarks end the paper.

2. The MUB random variable

We suppose that a set of m objects (or a set of m ordered evaluation degrees) has been defined, and r is the rank assigned by a single rater to a given item. Thus, our analysis concerns just the preference towards a single object (or the evaluation of a single item). Also, throughout the

[1] The E-M algorithm has the drawback of a lengthy convergence to the ML maximum. Thus, many solutions have been proposed in the literature for accelerating this convergence: McLachlan and Krishnan (1997), pp.141-151; McLachlan and Peel (2000), pp.70-73.

paper, we assume that $r = 1$ means "most preferred" while $r = m$ means "least preferred". Since the number of objects (or the grades in the evaluation process) to be ranked is generally known in advance, m is a prefixed number.

Then, D'Elia and Piccolo (2003) defined a MUB random variable and denoted it by $R \sim MUB\,(m, \pi, \xi)$ if:

$$Pr\,(R = r) = \pi \binom{m-1}{r-1} (1-\xi)^{r-1}\, \xi^{m-r} + (1-\pi)\,\frac{1}{m},\; r = 1, 2, ..., m.$$

Thus, the generating process is determined by the triple (m, π, ξ), where $m \geq 3$ is a positive integer[2] and the parametric space of (π, ξ) is $[0, 1] \times [0, 1]$.

Many interpretations can be derived from the elicitation mechanism that drives the raters' choice; for instance, it is immediate to convince oneself that π is a measure that is inversely related to the uncertainty of the choice, while ξ is positively related to the degree of liking expressed by the raters towards the prefixed object. We note that $\pi \to 0$ supports a shifted Binomial distribution while $\pi \to 1$ supports an *indifference preference* choice since any rank value included in $\{1, 2, \ldots\ m\}$ may be chosen with the same probability.

In some circumstances, it may be convenient to express the previous formula of the probability distribution in the following manner[3]:

$$Pr\,(R = r) = (unc) + (imp) \binom{m-1}{r-1} (dis)^r\,,\; r = 1, \ldots, m;$$

where:

$$unc = \frac{1-\pi}{m};\;\; imp = \frac{\pi}{1-\xi}\,\xi^m;\;\; dis = \frac{1-\xi}{\xi}.$$

In fact, the parameter unc is a measure of the *uncertainty share* that drives the choice, the parameter imp is an *impact coefficient* that raises

[2]The requirement $m \geq 3$ is a formal one; indeed, we need 2 degrees of freedom for the parameters and one more for satisfying the unit sum of the probabilities.

[3]Although this parametrization has been found useful in some comparative studies, it is not a one-to-one transformation, since the new quantities are obviously related each other.

the values of the probability distribution, and the parameter dis is a sort of *disliking odd measure* since it is defined as the ratio of a non-preference to a preference quantity.

In this paper, we discuss the expectation $\mu(\pi, \xi)$ and the variance $\sigma^2(\pi, \xi)$ of the MUB random variable; then, we will derive the formulae for an asymmetry measure $Asym(\pi, \xi)$ and for the kurtosis $Kurt(\pi, \xi)$ of $R \sim MUB(m, \pi, \xi)$. Generally, for obtaining a scale-invariant measure, the asymmetry coefficient is defined as the ratio of $Asym(\pi, \xi)$ to $[\sigma(\pi, \xi)]^3$; however, since we are just interested to the sign of the asymmetry measures we prefer to consider -throughout this note- only the numerator of the coefficient of asymmetry, that is the third central moment of R.

Formally, we let:

$$\mu(\pi, \xi) = \mathbb{E}(R); \quad \sigma^2(\pi, \xi) = \mathbb{E}(R - \mu(\pi, \xi))^2;$$

$$Asym(\pi, \xi) = \mathbb{E}(R - \mu(\pi, \xi))^3; \quad Kurt(\pi, \xi) = \frac{\mathbb{E}(R - \mu(\pi, \xi))^4}{[\sigma^2(\pi, \xi)]^2}.$$

To get some idea about this aspect, we let $m = 12$ and in Table 1 we show the main characteristics of some MUB random variables for some admissible parameters values.

Table 1. Some MUB distributions and their main characteristics.

	$(\pi, \quad \xi)$	unc	imp	dis	μ	σ^2	$Asym$	$Kurt$
A	0.75, 0.90	0.021	2.11822	0.111	3.200	7.352	35.623	5.440
B	0.80, 0.50	0.017	0.00039	1.000	6.500	4.583	0.000	3.223
C	0.05, 0.50	0.079	0.00002	1.000	6.500	11.458	0.000	1.840
D	0.50, 0.25	0.042	0.00000	3.000	7.875	8.880	-20.840	2.700

The corresponding plots of the probability distributions are presented in Figure 1.

Figure 1 confirms that the MUB random variable is able to take into accounts distributions with several different shapes including symmetric, positively and negatively skewed (in a moderate or strong degree), peaked, bell-shaped and platykurtic shapes. Moreover, the amount of uncertainty is inversely related to π as can be seen by comparing models B and

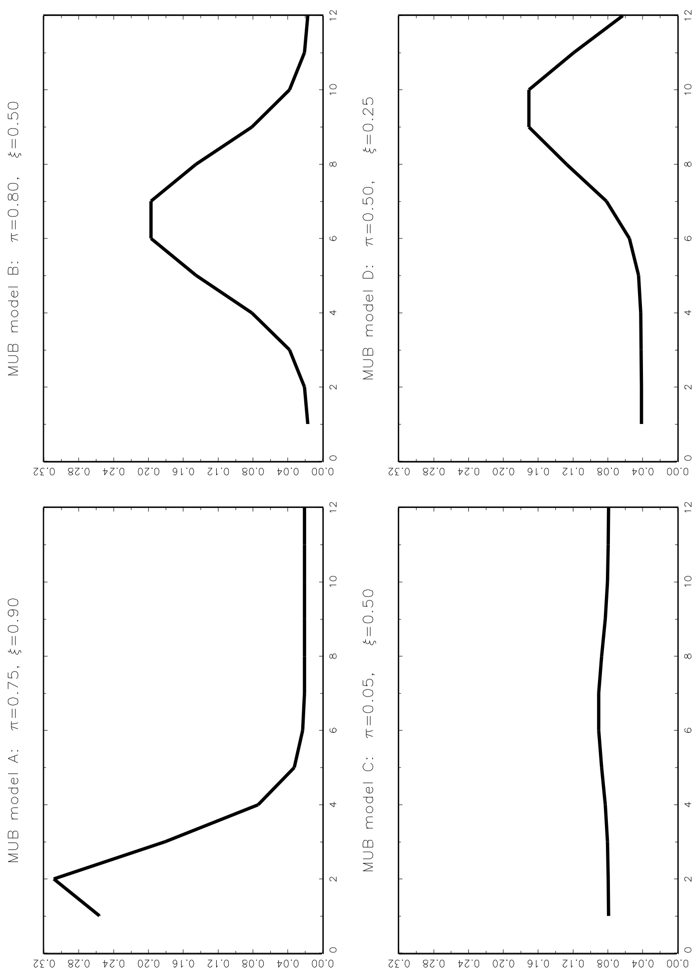

Figure 1. Probability distributions for the MUB models of Table 1

C, for instance. Of course, these characteristics are all related to the first moments of the MUB random variable.

3. Notation and preliminary results

In this note, we study the first four moments of the MUB random variable and discuss the effect of the parameters π and ξ on the location, the variability and the shape of the mixture distribution. Then, in this section we establish some results about the moments of a mixture distributions and the main algebraic result about the moments of a shifted Binomial distribution.

Let $B \sim h(r;\theta_1)$ and $U \sim g\,(r;\theta_2)$ be two discrete random variables, both defined on the support $\{1,2,\ldots,m\}$, and characterized by some parameters vectors θ_1, θ_2, respectively. Then, a mixture random variable R is defined by the probability mass function:

$$f_R(r) = \pi\, h(r;\theta_1) + (1-\pi)\; g\,(r;\theta_2)\,,\; r = 1,\ldots,m,$$

and the parameter $\pi \in [0,1]$.

The generating functions $G(t),\; G_B(t),\; G_U(t)$ of the random variables R, B, U are related, respectively, by:

$$G(t) = \pi\, G_B(t) + (1-\pi)\; G_U(t), \qquad \forall t \in \mathbb{R}.$$

Then, the k-th moments from the origin of R are:

$$\mathbb{E}\left(R^k\right) = \pi\, \mathbb{E}_B\left(X^k\right) + (1-\pi)\; \mathbb{E}_U\left(U^k\right),\; k = 0,1,\ldots,$$

where the expectation operators $\mathbb{E},\, \mathbb{E}_B,\, \mathbb{E}_U$ are to be applied with respect to the distributions of the random variables R, B, U, respectively. From these results, central and standardized moments can be obtained by standard formulae.

We denote by $Bin\,(n,\psi)$ a classical Binomial distribution characterized by the parameters n, ψ, where n is the number of independent experiments and ψ is the constant probability of the event of interest in each experiment. Then, in our case, B is a shifted Binomial random variable related to $X \sim Bin\,(m-1, 1-\psi)$ by: $B = X + 1$. Instead, $U \sim Ud\,(m)$

is a Uniform (rectangular) random variable defined over $\{1, 2, \ldots, m\}$. Their moments are well known in the literature (see: Johnson and Kotz, 1969, for instance).

Using a convenient matrix-oriented notation, the basic result for the following pages is the formula:

$$\sum_{r=1}^{m} \begin{bmatrix} r \\ r^2 \\ r^3 \\ r^4 \end{bmatrix} \binom{m-1}{r-1} (1-\xi)^{r-1} \xi^{m-r} = (m-1) \; \boldsymbol{B}_{(m)} \begin{bmatrix} \xi \\ \xi^2 \\ \xi^3 \\ \xi^4 \end{bmatrix} + \begin{bmatrix} m \\ m^2 \\ m^3 \\ m^4 \end{bmatrix},$$

where the non-zero elements b_{ij} of the $\boldsymbol{B}_{(m)}$ matrix depend only upon m and are defined by:

$$\begin{aligned}
b_{11} &= -1; \\
b_{21} &= -(2m-1); \\
b_{31} &= -\left(3m^2 - 3m + 1\right); \\
b_{41} &= -(2m-1)\left(2m^2 - 2m + 1\right); \\
b_{22} &= (m-2); \\
b_{32} &= 3(m-1)(m-2); \\
b_{42} &= (m-2)\left(6m^2 - 12m + 7\right); \\
b_{33} &= -(m-2)(m-3); \\
b_{43} &= -2(m-2)(m-3)(2m-3); \\
b_{44} &= (m-2)(m-3)(m-4).
\end{aligned}$$

The previous formula has been obtained, after some algebraic manipulations, from the moments of the Binomial random variable via the moment or the factorial generating functions[4].

4. Expectation and variance of the MUB random variable

Following the previous approach, it is immediate to derive the expectation of the MUB random variable:

$$\mu(\pi,\xi) = \pi\ (m-1)\left(\frac{1}{2}-\xi\right)+\frac{(m+1)}{2}.$$

Starting from the mid-range of the distribution, $(m+1)/2$, the mean value of R increases (decreases) as long as ξ is less than (more than) $\frac{1}{2}$.

Figure 2 shows the behavior of the expectation of R over the parametric space. It confirms that when both π and ξ increase towards 1 the mean value converges to 1, and thus the MUB model implies a greater preference for the given object.

In a similar way, we obtain the variance of the MUB distribution:

$$\sigma^2(\pi,\xi) = (m-1)\left\{\pi\,\xi\,(1-\xi)+(1-\pi)\left[\frac{m+1}{12}+\pi\,(m-1)\left(\frac{1}{2}-\xi\right)^2\right]\right\}.$$

The variance reaches its minimum when both π and ξ increase to 1; for a given π, the variance has a maximum when $\xi=\frac{1}{2}$.

Re-arranging the formula of the variance, we can write:

$$\frac{\sigma^2(\pi,\xi)}{(m-1)} = \pi\xi\,(1-\xi)\,[2-m+\pi\,(m-1)]+(1-\pi)\,\frac{[3\pi\,(m-1)+(m+1)]}{12},$$

and from this alternative formulation it is immediate to realize that:

$$\sigma^2(\pi,\xi)=\sigma^2(\pi,1-\xi)\,,$$

confirming that the variability of the R distribution is symmetric around $\xi=1/2$ (Figure 3).

[4] All the results of this paper have been checked from a symbolic point of view by the Maple© language and from a computational point of view by the Gauss© programming language.

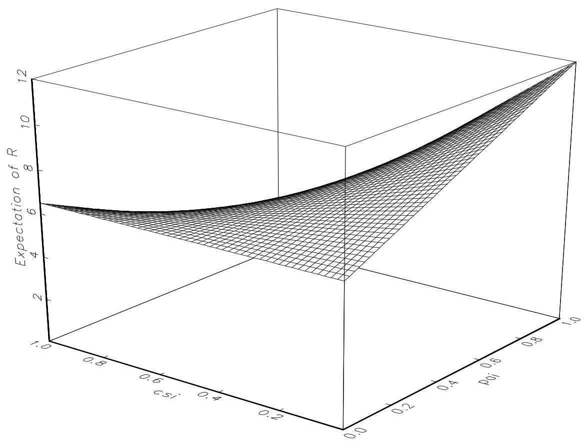

Figure 2. Expectation of the MUB random variable

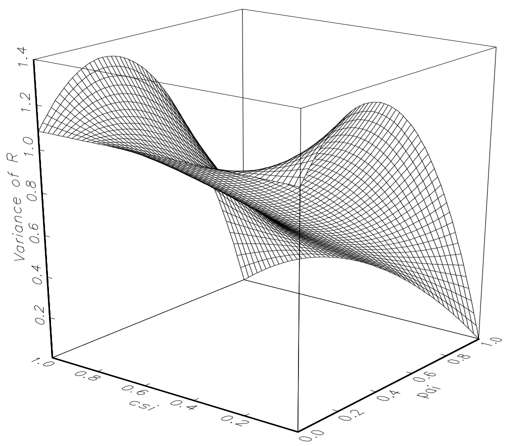

Figure 3. Variance of the MUB random variable

5. Asymmetry of the MUB random variable

In this section, we derive and discuss an asymmetry measure for the MUB random variable. In fact, as it concerns the sign of the asymmetry coefficient, it is sufficient to study only the third central moment as defined in the section 2.

After a lengthy algebra, we found that:

$$Asym\,(\pi, \xi) = \pi\ (m-1)\,(2\xi - 1)\,\frac{[A\,(\pi)\ \xi\,(1-\xi) + B\,(\pi)]}{4}\,,$$

where:

$$\begin{aligned} A\,(\pi) &= 4\,(m-1)^2\,\pi^2 - 6\,(m-1)\,(m-2)\,\pi + 2\,(m-2)\,(m-3)\,; \\ B\,(\pi) &= (m-1)\,[(1-\pi)\,(1+\pi\,(m-1))]\,; \end{aligned}$$

are functions only of the parameter π.

From this formula, we can derive some general results:

i) $Asym\left(\pi, \frac{1}{2}\right) = 0,\ \forall\pi \in [0,1]\,;$

ii) $Asym\,(\pi, \xi) = -Asym\,(\pi, 1-\xi)\,,\ \forall\pi \in [0,1]\,.$

Figure 4 shows the behavior of $Asym\,(\pi, \xi)$ over the parametric space.

There is an immediate relationship between the asymmetry and the expectation of the MUB random variable, as the mean value of this distribution increases (decreases) as long as the asymmetry is markedly negative (positive). This fact has an interpretative consequence since it shows that the mean preference of the raters towards a fixed object increases (decreases) with respect to the mid-range together with the negative (positive) value of the asymmetry measure. Briefly, a positive (negative) asymmetry is accompanied by a preference (adversity) toward the object.

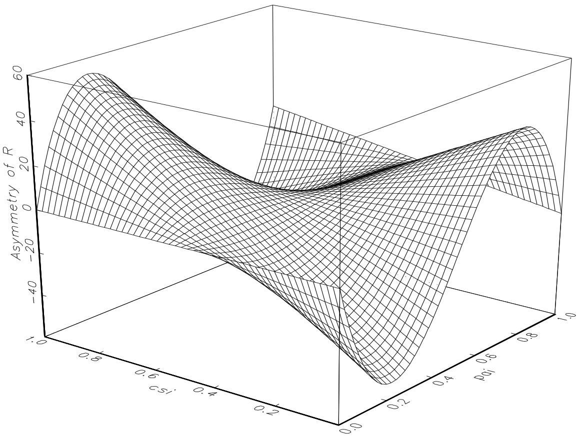

Figure 4. Asymmetry measure of the MUB random variable

6. Kurtosis of the MUB random variable

In this section, we derive and discuss the kurtosis of the MUB random variable defined as the fourth standardized moment; for this measure it is more convenient to take into account also the effect of the denominator. Indeed, we need a scale invariant index for referring our measure to the kurtosis of the standardized Gaussian random variable.

The relevance of this aspect stems from three facts:

i) peaked ranks distributions are those with a neat preference (adversity) depending upon the positive (negative) sign of $(\xi - 1/2)$;

ii) in the family of MUB random variables, very peaked distributions are allowed only when there is a mode at $R = 1$ or at $R = m$;

iii) a large number of items increases the performance of the Central Limit theorem on the elicitation mechanism since, in this case, the uncertainty factor is greater.

Proceeding as in the previous sections, we get the following for the numerator of the kurtosis[5]:

$$Kurt(m, \pi, \xi) = \frac{3\, N\,(\pi, \xi)}{5\ (m-1)\ D\,(\pi, \xi)},$$

where

$$\begin{aligned} N\,(\pi, \xi) &= A_0 + A_1(\xi)\pi + A_2(\xi)\pi^2 + A_3(\xi)\pi^3 + A_4(\xi)\pi^4; \\ D\,(\pi, \xi) &= \left(B_0 + B_1(\xi)\pi + B_2(\xi)\pi^2\right)^2; \end{aligned}$$

and the coefficients are defined by:

$$\begin{aligned} A_0 &= (m+1)\left(3m^2-7\right); \\ A_1(\xi) &= 4\,(m-2)\left[c_0 + c_1\xi + c_2\xi^2 + c_3\xi^3\,(2-\xi)\right]; \\ A_2(\xi) &= 30\,(m-1)\,(m-2)\,(2\xi-1)^2\,[8\xi\,(1-\xi)\,(m-2)-(m-1)]; \\ A_3(\xi) &= 60\,(m-2)\,(m-1)^2\,(2\xi-1)^2\left(6\xi^2-6\xi+1\right); \\ A_4(\xi) &= -45\,(m-1)^3\,(2\xi-1)^4; \\ B_0 &= m+1; \\ B_1(\xi) &= 2\,(m-2)\left(6\xi^2-6\xi+1\right); \\ B_2(\xi) &= -3\,(m-1)\,(2\xi-1)^2. \end{aligned}$$

For simplifying the expression of $A_1(\xi)$, we let:

$$c_0 = \left(3m^2-6m+1\right); \qquad c_1 = -30\left(m^2-4m+5\right);$$

$$c_2 = 30\left(3m^2-18m+29\right); \; c_3 = -60\,(m-3)\,(m-4)\,.$$

Then, we examine the plot of $Kurt\,(\pi, \xi)$, for varying π and ξ over their parametric space (Figure 5). The surface enhances the symmetry of the kurtosis and the quick increase of this measure towards the borders of the parametric space; indeed, the kurtosis coefficient can not be computed on the border since, for $\xi = 0$ and $\xi = 1$, we obtain a zero variance.

[5]Although the denominator of this formula is related to the squared variance of the MUB distribution, we prefer a more compact expression for computational purposes.

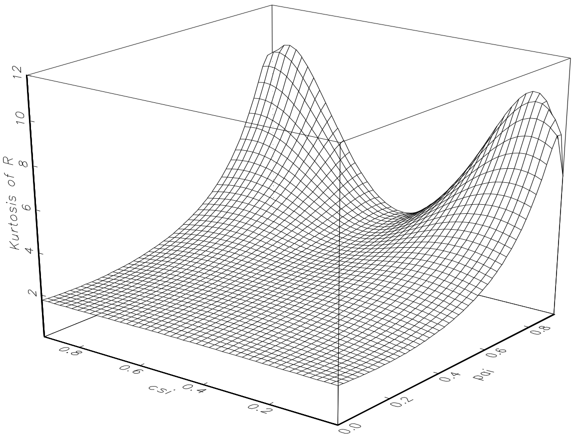

Figure 5. Kurtosis of the MUB random variable

From the kurtosis formula, we found that:

$$Kurt\,(\pi, \xi) = Kurt\,(\pi, 1 - \xi)\,,\, \forall m \geq 3.$$

Thus, it turns out that it is useful to study kurtosis only for $0 \leq \xi \leq 1/2$.

A direct interpretation of the kurtosis formula is extremely cumbersome; thus, we limit ourselves to quote some useful properties obtained in several specific cases.

When $\xi = 1/2$, the distribution is symmetric and the kurtosis coefficient can be greatly simplified:

$$Kurt\left(\pi, \frac{1}{2}\right) = \frac{3\left[(m+1)\left(3m^2 - 7\right) - (m-2)\left(3m^2 + 9m - 34\right)\pi\right]}{5\ (m-1)\left[(m-2)\,\pi - (m+1)\right]^2}.$$

When $\xi = 1/2$ and the kurtosis approaches the value of 3 we expect a bell-shaped distribution. Indeed, the equation: $Kurt\left(\pi, \frac{1}{2}\right) = 3$, has two real and admissible solutions only for $m \geq 10$:

$$\pi_i = \frac{7m^2 - 9m + 24 \pm \sqrt{9m^4 - 126m^3 + 417m^2 - 432m + 616}}{10\,(m-1)\,(m-2)}\,,\ i = 1, 2.$$

Note that:

$$\lim_{m\to\infty} \pi_{_1} = 1; \ \lim_{m\to\infty} \pi_{_2} = 2/5 = 0.4.$$

Table 2 gives the solutions of $Kurt\left(\pi_i, \frac{1}{2}\right) = 3$ for some common values of m. The values of π reported in the table are those implied by a MUB distribution whose shape is mostly symmetric and similar to the Gaussian random variable.

Table 2. Solutions of $Kurt\left(\pi, \frac{1}{2}\right) = 3$ *for some values of* m*.*

$m = 10$	$\pi_1 = 0.94261$	$\pi_2 = 0.81850$
$m = 11$	$\pi_1 = 0.97100$	$\pi_2 = 0.74455$
$m = 12$	$\pi_1 = 0.98193$	$\pi_2 = 0.69807$
$m = 13$	$\pi_1 = 0.98770$	$\pi_2 = 0.66381$
$m = 14$	$\pi_1 = 0.99117$	$\pi_2 = 0.63703$
$\ldots\ldots$	$\ldots\ldots\ldots$	$\ldots\ldots\ldots$
$m \to \infty$	$\pi_1 \to 1$	$\pi_2 \to 2/5 = 0.4$

The fact that a kurtosis of 3 is admissible only when $m \geq 10$ is a confirmation of the empirical founding that in the preference/evaluation studies the distributions are symmetric, unimodal and bell-shaped only when the number of items is moderately large.

Moreover,

$$Kurt\,(0, \xi) = \frac{3\,(3m^2 - 7)}{5\,(m^2 - 1)};$$

this result is consistent with the fact that $R \sim MUB\,(m, 0, \xi)$ is indeed the Uniform discrete random variable[6] over $\{1, 2, \ldots, m\}$. We note also that if $m \to \infty$, then $Kurt\,(0, \xi) \to 1.8$, that is the kurtosis of the continuous Uniform random variable over any finite interval.

[6]Here, we write $R \sim MUB\,(m, 0, \xi)$ for consistency with the previous notation. However, this distribution is function only of m, since the effect of ξ is cancelled out when $\pi = 0$.

Finally, we obtain:

$$Kurt\,(1,\xi) = \frac{1 + 3\ (m-3)\ \xi\ (1-\xi)}{(m-1)\,\xi\ (1-\xi)};$$

and $Kurt\,(1,\xi) \rightarrow 3$ if $m \rightarrow \infty$. Indeed, when $\pi = 1$ the MUB distribution is a Shifted Binomial random variable, that is a converging distribution to the Gaussian one for $m \rightarrow \infty$.

As far as it concerns the kurtosis and asymmetry relationships, the previous discussion confirms that (given a relatively high value of m) the MUB random variable converges to the Gaussian distribution if it is symmetric and if π is not too small; in fact, small values of π induce a platykurtic shape in the distribution (like model C in Figure 1).

7. Moment estimators for the MUB parameters

In this section we deal with the derivation of the moment estimators of π and ξ from the first two sample moments. These relationships produce consistent estimators of the parameters and thus they can be quite effective for getting starting values for an efficient maximum likelihood estimation procedure[7].

If we compute the sample moments m_1, m_2 from the observed ranks data, then the solutions of the two non-linear equations:

$$\begin{aligned} \mathbb{E}\,(R) &= \mu_1\,(\pi,\xi) = m_1; \\ \mathbb{E}\left(R^2\right) &= \mu_2\,(\pi,\xi) = m_2; \end{aligned}$$

are expressed by[8]

$$\begin{cases} \hat{\xi} = \frac{2\left(1+3m_1 m - m^2\right) - 3(m_1+m_2) \pm \sqrt{\Delta}}{3(m-2)(2m_1-m-1)}, & \text{if } m_1 \neq \frac{m+1}{2}; \\ \hat{\pi} = \frac{2m_1-(m+1)}{(m-1)\left(1-2\hat{\xi}\right)}, & \text{if } \hat{\xi} \neq \frac{1}{2}; \end{cases}$$

[7]The effects of good starting points are more relevant in the planning of simulation experiments.

[8]The solutions are a couple of two quadratic expressions for each parameter. If these solutions are different, only one of them produces admissible solutions for the parameters.

where

$$
\begin{aligned}
\Delta &= d_0 + d_1 m_1 + d_2 m_1^2; \\
d_0 &= m^4 + (3m_2+1)\,m^2 + 3\,(3m_2+2)\,m + \left(4 + 6m_2 + 9m_2^2\right); \\
d_1 &= -3\,(m+1)\left[m\,(2m-1) + 6\,(m_2+1)\right]; \\
d_2 &= 3\left(4m^2 + 2m + 7\right).
\end{aligned}
$$

We observe that if the empirical distribution of the ranks is perfectly symmetric, the first moment is the midrange: $m_1 = (m+1)/2$, and this implies $\hat{\xi} = \frac{1}{2}$. Thus, neither of the previous solutions are defined.

In this case, we let $\hat{\xi} = \frac{1}{2}$ and we derive the estimates for π from the second moment equation (in fact, in this case, the first equation is not informative for π). Specifically, for an empirical symmetrical distribution of the ranks the solutions are:

$$
\begin{cases}
\hat{\xi} = \frac{1}{2}, & \text{if } m_1 = \frac{m+1}{2}; \\
\hat{\pi} = \frac{2\left(2m^2+3m+1-6m_2\right)}{(m-1)(m-2)}, & \text{if } \hat{\xi} = \frac{1}{2}.
\end{cases}
$$

If we let $\hat{\theta}_{mom} = (\hat{\pi}_{mom}, \hat{\xi}_{mom})'$ and $\hat{\theta}_{ML} = (\hat{\pi}_{ML}, \hat{\xi}_{ML})'$ for the moment and maximum likelihood estimates, respectively, then the quantity:

$$
dist = \sqrt{\left(\hat{\theta}_{mom} - \hat{\theta}_{ML}\right)' \left(\hat{\theta}_{mom} - \hat{\theta}_{ML}\right)}
$$

is a measure of the Euclidean distance between the two different estimates. Of course, its range is $\left[0, \sqrt{2}\right]$.

Then, to show the effectiveness of using the moment estimates, we examine both the statistical efficiencies of the moment estimates and the computational advantages in using these values as a starting points for the E-M algorithm. We choose the 'Cities preferences' expressed by $n = 183$ young people, a data set fully discussed by D'Elia and Piccolo (2003).

From a statistical point of view, we compare the moment estimates and the maximum likelihood estimates for 12 estimated distributions. These models exhibit quite different locations, variabilities and shapes; thus, this experiment includes a large variety of real ranking situations.

Table 3 presents the different estimates obtained and the Euclidean distance (*dist*) previously defined.

Table 3. A comparison between moment and ML estimates.

Cities	$\hat{\pi}_{mom}$	$\hat{\xi}_{mom}$	$\hat{\pi}_{ML}$	$\hat{\xi}_{ML}$	*dist*
Florence	0.89279	0.86971	0.83369	0.87884	0.05980
Venice	0.53843	0.56439	0.53480	0.57872	0.01478
Turin	0.69113	0.23111	0.33974	0.14190	0.36254
Bari	0.81228	0.18621	0.67249	0.16480	0.14142
Bologna	0.64931	0.82385	0.53869	0.84661	0.11294
Catania	0.79513	0.22272	0.63518	0.20490	0.16094
Genoa	0.67658	0.45479	0.64187	0.48942	0.04903
Palermo	0.76662	0.27114	0.63137	0.28688	0.13616
Milan	0.60985	0.34815	0.15460	0.30267	0.45752
Naples	0.67649	0.84161	0.56582	0.86156	0.11245
Rome	0.83037	0.87346	0.74971	0.88884	0.08211
Verona	0.51337	0.20128	0.16803	0.01364	0.39302

Although in some limited case (Turin, Milan, Verona) there is a substantial discrepancy between the two estimates (mainly for the π parameter), the results confirm that the preliminary moments estimates are generally quite good as a starting point for the final maximum likelihood estimates. With a few exceptions, the distance between initial and final estimates on the parametric space is a very small quantity.

Then, from a computational point of view, we compare the number of iterations that the same E-M algorithm required in order to reach the convergence to the final maximum likelihood estimates. Indeed, we examine the effects of the following starting points:

$$\begin{aligned}
\theta^{(0)}_{init} &= (0.1,\, 0.1)'\,; \\
\theta^{(0)}_{DP} &= \left(0.5,\, \frac{m-m_1}{m-1}\right)'\,; \\
\theta^{(0)}_{mom} &= \left(\hat{\pi}_{mom},\, \hat{\xi}_{mom}\right)'\,.
\end{aligned}$$

The rationale for these values derives from the following considerations:

- the $\theta^{(0)}_{init}$ values set the starting points at 0.1, assuming no preliminary information on the parameters estimates;

- the $\theta^{(0)}_{DP}$ values were proposed by D'Elia and Piccolo (2003). They set the π initial estimate at its midrange and the ξ at the ML estimate of the shifted Binomial component of the mixture. Thus, the ξ initial estimates uses the sample information contained in the sample rank average m_1;

- the $\theta^{(0)}_{mom}$ values are the moment estimates previously discussed.

Finally, we let $NIT\left(\theta^{(0)}_{init}\right)$, $NIT\left(\theta^{(0)}_{DP}\right)$, $NIT\left(\theta^{(0)}_{mom}\right)$ be the number of iterations required for the convergence in the three starting values approaches, respectively. In Table 4, the numbers in bold italics indicates for each model the best result in terms of minimum number of iterations required for the convergence.

Table 4. Number of iterations for the convergence of E-M algorithms.

Cities	$NIT\left(\theta^{(0)}_{init}\right)$	$NIT\left(\theta^{(0)}_{DP}\right)$	$NIT\left(\theta^{(0)}_{mom}\right)$
Florence	42	20	***17***
Venice	47	20	***13***
Turin	***74***	89	86
Bari	29	***20***	23
Bologna	54	***20***	21
Catania	33	***21***	26
Genoa	41	26	***22***
Palermo	35	25	***24***
Milan	188*****	90	***85***
Naples	57	***19***	21
Rome	48	18	***17***
Verona	***59***	81	70

(*****) *Convergence achieved at a singular point.*

The results confirm the relevance of the starting values in carrying out the E-M algorithm. As a matter of fact, Table 4 shows no definite dominance of a single approach, although the second and the third method for the starting values are the only suitable candidates, with a distinct preference for the moment estimator in the case of well behaved data. The fact that in one case (Milan data) generic starting values implied a convergence of the E-M algorithm towards a singular point is a further evidence of the usefulness of a consistent preliminary estimates in this kind of numeric procedures.

8. Concluding remarks

In this paper, we have derived the formulae for the asymmetry and the kurtosis of the MUB random variable and we have discussed their interpretations over the admissible parametric space.

Finally, we obtained the moment estimates for the MUB distribution parameters and we checked their usefulness with respect to several models estimated on a real data set.

In this regard, further developments include the possibility to set the moment solutions as automatic starting points in the E-M algorithms of the maximum likelihood procedure. The statistical and numerical efficiencies of this proposal should be evaluated by a simulation experiment to be planned by generating data from several MUB models well spread over the parametric space.

Acknowledgments: We thank the support received by the Dipartimento di Scienze Statistiche, University of Naples Federico II and the research structures offered by the Centro per la Formazione in Economia e Politica dello Sviluppo Rurale, Portici. We thank a referee for helpful comments to a preliminary version of the paper.

References

D'Elia A. (2003), A mixture model with covariates for ranks data: some inferential developments, *Quaderni di Statistica*, **5**, *this issue*.

D'Elia A., Piccolo D. (2003), A mixture model for preference data analysis, *submitted*.

Johnson N.L., Kotz S. (1969), *Discrete Distributions*, J. Wiley & Sons, New York.

McLachlan G., Krishnan G. J. (1997), *The EM Algorithm and Extensions*, J. Wiley & Sons, New York.

McLachlan G., Peel G. J. (2000), *Finite Mixture Models*, J. Wiley & Sons, New York.

Piccolo D. (2003), Computational issues in the E-M algorithm for ranks model estimation with covariates, *Quaderni di Statistica*, **5**, *this issue*.