An. Şt. Univ. Ovidius Constanţa Vol. 11(2), 2003, 189–194

ARITHMETIZING CONTINUOUS DISTRIBUTIONS: NUMERICAL ASPECTS

Raluca Vernic

Abstract

This paper deals with the arithmetization of univariate and bivariate distributions, i.e. with their transformation into discrete distributions defined on the non-negative integers. Numerical aspects are investigated.

1 Introduction

An **arithmetic distribution** is a discrete distribution defined on the non-negative integers, while an equispaced arithmetic distribution is defined only on multiples of an unit of measurement $h > 0$, called the span (see [2]). It is sometimes necessarily to transform a non-arithmetic distribution into an arithmetic one and such a transformation is called discretization or arithmetization. For example, in actuarial mathematics, the total claims distribution of a portfolio is frequently evaluated using recursive methods which need an arithmetic form for the claim severity distribution. Since this distribution is usually continuous (e.g. Exponential, Lognormal, Pareto etc.), one must first arithmetize it.

From the existing methods for the arithmetization of continuous distributions (see [1], [2], [3], [4], [5]), in this paper we will investigate some practical aspects on the method of rounding.

2 Arithmetizing the marginals of a bivariate distribution

Let us first recall the method of rounding (mass dispersal) for both univariate and bivariate case. We denote by X the random variable (r.v.) to be arithmetized, by F_X its distribution function (d.f.) and by f_j the probability placed at jh, $j = 0, 1, ...$ Then we set

$$f_0 = P\left(X < \frac{h}{2}\right) = F_X\left(\frac{h}{2} - 0\right) \tag{1}$$

and for $j \geq 1$:

$$f_j = P\left(jh - \frac{h}{2} \leq X < jh + \frac{h}{2}\right) = F_X\left(jh + \frac{h}{2} - 0\right) - F_X\left(jh - \frac{h}{2} - 0\right). \tag{2}$$

Remark. The notation $F_X(x-0)$ indicates that the discrete probability at x should not be included. For continuous distributions this will make no difference.

In the bivariate case, we denote by (X, Y) the random vector to be arithmetized, by $F_{X,Y}$ its d.f. and let f_{ij} be the probability placed at (ih, jk), $i, j = 0, 1, ...$ and $h, k > 0$. Then (see [4])

$$f_{00} = F_{X,Y}\left(\frac{h}{2} - 0, \frac{k}{2} - 0\right) \tag{3}$$

$$f_{0j} = F_{X,Y}\left(\frac{h}{2} - 0,\, jk + \frac{k}{2} - 0\right) - F_{X,Y}\left(\frac{h}{2} - 0,\, jk - \frac{k}{2} - 0\right),\ j = 1, 2, ... \tag{4}$$

$$f_{i0} = F_{X,Y}\left(ih + \frac{h}{2} - 0, \frac{k}{2} - 0\right) - F_{X,Y}\left(ih - \frac{h}{2} - 0, \frac{k}{2} - 0\right),\ i = 1, 2, ... \tag{5}$$

$$\begin{aligned} f_{ij} = {} & F_{X,Y}\left(ih + \frac{h}{2} - 0,\, jk + \frac{k}{2} - 0\right) - F_{X,Y}\left(ih - \frac{h}{2} - 0,\, jk + \frac{k}{2} - 0\right) - \\ & -F_{X,Y}\left(ih + \frac{h}{2} - 0,\, jk - \frac{k}{2} - 0\right) + F_{X,Y}\left(ih - \frac{h}{2} - 0,\, jk - \frac{k}{2} - 0\right). \end{aligned} \tag{6}$$

Proposition. *We consider a bivariate distribution with known marginal distributions. If we are interested in arithmetizing its marginal distributions, then the two following methods give the same result:*

1. *one directly arithmetizes the univariate marginal distributions;*
2. *one first arithmetizes the bivariate distribution, then evaluates the marginal distributions by summing.*

Proof. We will prove the property for the marginal distribution of X, the proof for Y being similar. Using the second method and (3), (4), we get

$$f_0^{(X\,biv)} = f_{00} + \sum_{j=1}^{\infty} f_{0j} = F_{X,Y}\left(\frac{h}{2} - 0, \frac{k}{2} - 0\right) +$$

$$+\sum_{j=1}^{\infty}\left[F_{X,Y}\left(\frac{h}{2}-0,\, jk+\frac{k}{2}-0\right)-F_{X,Y}\left(\frac{h}{2}-0,\, jk-\frac{k}{2}-0\right)\right]=$$

$$=F_{X,Y}\left(\frac{h}{2}-0,\infty\right)=F_X\left(\frac{h}{2}-0\right)=f_0^{(X\,univ)}.$$

Similarly, from (5) and (6), for $i\geq 1$,

$$f_i^{(X\,biv)}=f_{i0}+\sum_{j=1}^{\infty}f_{ij}=...=F_{X,Y}\left(ih+\frac{h}{2}-0,\infty\right)-F_{X,Y}\left(ih-\frac{h}{2}-0,\infty\right)=$$

$$=F_X\left(ih+\frac{h}{2}-0\right)-F_X\left(ih-\frac{h}{2}-0\right)\overset{(2)}{=}f_i^{(X\,univ)}.\square$$

Remark. Following the above property, one may choose the best method considering his particular framework.

3 A distance between the exact and the arithmetized distributions

3.1 The univariate case

Let X be a r.v. and $\hat{X}$ its arithmetized corresponding r.v. We denote by F and respectively $\hat{F}$ their d.f.s. In order to evaluate the error made when arithmetizing a distribution, we introduce the following distance

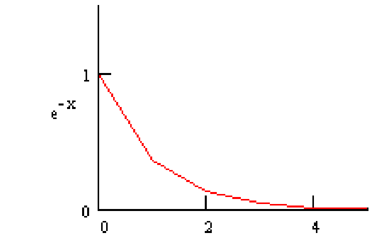
$$\begin{aligned} d(F,\hat{F}) &= \sup_{x\in\mathbb{R}}\left|F(x)-\hat{F}(x)\right| = \sup_{a\in\mathbb{N}^*}\sup_{(a-1)h<x\leq ah}\left|F(x)-\sum_{i=0}^{a-1}f_i\right| \overset{\text{using }(2)}{=} \\ &= \sup_{a\in\mathbb{N}^*}\sup_{(a-1)h<x\leq ah}\left|F(x)-F\left(ah-\frac{h}{2}\right)\right|. \end{aligned} \tag{7}$$

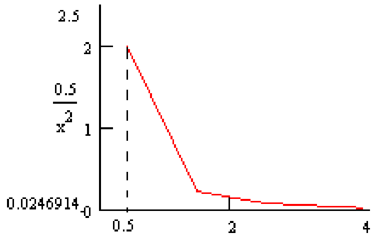
Let's consider two particular cases: Exponential and Pareto.

3.1.1 $X\sim Exponential\,(\theta)\,,\ \theta>0$

In this case, the density of X is $f(x)=\theta e^{-\theta x}, x>0$. Then the distance (7) reduces to the first term of (7), for $a=1$ (see the density's graph, first graph),

$$d(F,\hat{F})=1-e^{-\theta h/2}.$$





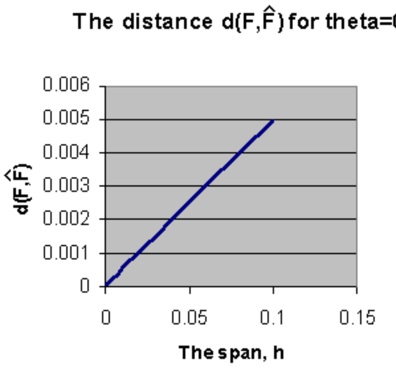
Remark. For fixed h, when $\theta \to \infty$, $d(F, \hat{F}) \to 1$, which is not good. For fixed θ, the smaller h is, the smaller $d(F, \hat{F})$ becomes (which is normal).

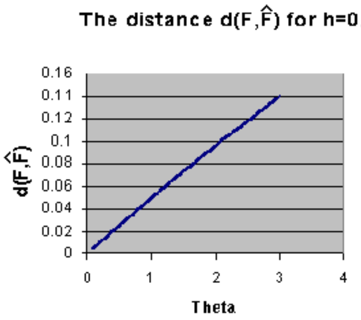
3.1.2 $X \sim Pareto\,(\alpha, \theta)\,,\ \alpha, \theta > 0$

The density of X is $f\,(x) = \dfrac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, x > \theta$. Based on the shape of the graph (second graph) of this density, the distance (7) gives

$$d(F, \hat{F}) = \begin{cases} \max\left\{ \left(\frac{\theta}{h}\right)^{\alpha} \left(\frac{1}{(i-0.5)^{\alpha}} - \frac{1}{i^{\alpha}}\right), 1 - \left(\frac{\theta}{h(i-0.5)}\right)^{\alpha} \right\}, \text{ if } \exists i \in \mathbb{N}, \\ \qquad (i-1)\,h < \theta \le (i-0.5)\,h \\ \\ \max\left\{ \left(\frac{\theta}{h}\right)^{\alpha} \left(\frac{1}{i^{\alpha}} - \frac{1}{(i+0.5)^{\alpha}}\right), 1 - \left(\frac{\theta}{hi}\right)^{\alpha} \right\}, \text{ if } \exists i \in \mathbb{N}, \\ \qquad (i-0.5)\,h < \theta \le ih \end{cases}.$$

Remark. The following two graphs show the evolution of the distance $d(F, \hat{F})$ in the Exponential case.





3.2 The bivariate case

Considering similar notations as in the univariate case, the distance (7) can be defined as

$$d(F,\hat{F}) = \sup_{x,y\in\mathbb{R}} \left| F(x,y) - \hat{F}(x,y) \right| =$$

$$= \sup_{a,b\in\mathbb{N}^*} \sup_{(a-1)h<x\leq ah,(b-1)k<y\leq bk} \left| F(x,y) - \sum_{i=0}^{a-1}\sum_{j=0}^{b-1} f_{ij} \right| =$$

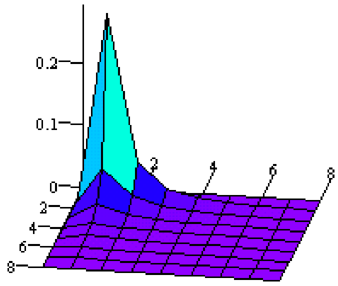
$$= \sup_{a,b\in\mathbb{N}^*} \sup_{(a-1)h<x\leq ah,(b-1)k<y\leq bk} \left| F(x,y) - F\left(ah-\frac{h}{2}, bk-\frac{k}{2}\right) \right|. \tag{8}$$

3.2.1 Particular case: $(X,Y) \sim Pareto\,(\theta_1,\theta_2,\alpha)\,, \theta_1>0, \theta_2>0, \alpha>0$

The bivariate Pareto distribution has the density function

$$f_{X,Y}(x,y) = a(a+1)\left(\theta_1\theta_2\right)^{a+1}\left(\theta_2 x+\theta_1 y-\theta_1\theta_2\right)^{-(a+2)}, x>\theta_1, y>\theta_2.$$

Then X and Y are univariate Pareto distributed (with parameters $\alpha_i\theta_i$, $i = 1, 2$) and for the arithmetized distribution see [4]. Unfortunately, due to the form of the graph of the density (see next figure), we don't have a simple form for the distance (8), which must be evaluated numerically.



References

[1] H.U. Gerber, *On the numerical evaluation of the distribution of aggregate claims and its stop-loss premiums*, *IME* **1**, 13–18, 1982.

[2] S.A. Klugman, H.H. Panjer & G.E. Willmot, *Loss models. From data to decisions*, Wiley, 1998.

[3] H.H. Panjer & B.W. Lutek, *Practical aspects of stop-loss calculations*, *IME* **2**, 159–177, 1983.

[4] R. Vernic, *Arithmetizing bivariate distributions, PAMM PC-139*, 189-194, 2002.

[5] J.F. Walhin & J. Paris, *On the use of equispaced distributions*, *AB* 28, 241–255,1998.

"Ovidius" University of Constanta,
124, Mamaia Blvd, 900527 Constanta, ROMANIA
E-mail: rvernic@univ-ovidius.ro