STATISTICA, anno LXXVIII, n. 4, 2018

DISCRETE POWER DISTRIBUTIONS AND INFERENCE USING LIKELIHOOD

Andrea Pallini[1]
Dipartimento di Economia e Management, Università di Pisa, Pisa, Italia

1. Introduction

In nonparametric testing problems, Lehmann (1953) introduces a class of alternatives, that are defined as $F(x)^\alpha$, where $F(x)$ is a continuous distribution function, and α is a positive integer, for all $-\infty < x < +\infty$. Durrans (1992), without knowing Lehmann (1953), extends the distribution of the nth largest order statistic, $F(x)^n$ say, to $F(x)^\alpha$, where $\alpha > 0$ is a real number, for all $-\infty < x < +\infty$. More generally, for a continuous distribution function $F(x)$, with density $f(x)$, the power distribution function $H(x;\alpha)$ can be defined as $H(x;\alpha) = F(x)^\alpha$, with density $h(x;\alpha) = \alpha F(x)^{\alpha-1} f(x)$, where $\alpha > 0$, for all $-\infty < x < +\infty$.

Generalizing Lehmann (1953), Miura and Tsukahara (1993) study different classes of continuous alternatives. Continuous power distributions have recently been considered in Gupta and Gupta (2008), Pewsey *et al.* (2012) and Gómez and Bolfarine (2015). In particular, Pewsey *et al.* (2012) and Gómez and Bolfarine (2015) propose and study the basic theory for the likelihood-based inference in continuous power distributions.

Jones (2004) is relevant to further extensions and similarities, since important continuous distributions are obtained from the distributions of order statistics. Nadarajah and Kotz (2006) can be considered for understanding the fact that continuous power distributions are also studied as exponentiated distributions, with interesting achievements.

In this paper, a discrete counterpart of the continuous power distributions introduced in Lehmann (1953) and Durrans (1992) is studied. Let $F(x;\theta)$ be a discrete distribution function, that can be regarded as the original distribution, where θ is a model parameter, with values θ in a space Θ, that is $\theta \in \Theta$. The corresponding discrete power distribution function $H(x;\theta,\alpha)$ can be obtained as $F(x;\theta)^\alpha$, by defining convenient positive jumps $F(x_i;\theta)^\alpha - F_{-}(x_i;\theta)^\alpha$, on the discontinuities $\{x_i\}$, where $\alpha > 0$, for all $-\infty < x < +\infty$. Inequalities in moments and distribution functions, based on a specific application of the Jensen's inequality, for the original and power distributions, are

[1] Corresponding Author. E-mail: andrea.pallini@unipi.it

studied. Such inequalities also allow the definition of the discrete intermediate distributions $G(x;\theta,\alpha)$, that lie between an original distribution and a power distribution, for all $-\infty < x < +\infty$.

Power and intermediate discrete distributions $H(\theta,\alpha)$ and $G(\theta,\alpha)$ are studied theoretically in detail. In particular, the uniform, the binomial, the Poisson, the negative binomial, the hypergeometric distributions are examined, with the corresponding new power and intermediate distributions. Power and intermediate distributions $H(\theta,\alpha)$ and $G(\theta,\alpha)$ are flexble and suitable for analysing and fitting discrete data with various degrees of variance, namely overdispersion and underdispersion, skewness and kurtosis, as α varies, along $\alpha > 0$.

Problems of estimation for the power and intermediate distributions $H(\theta,\alpha)$ and $G(\theta,\alpha)$ are considered using likelihood methods, with specific attention to maximum likelihood estimation, information and asymptotics. Classic numerical optimization tools for maximum likelihood estimation are also explored, by performing simulation experiments.

Similar power approaches for the exponentiated geometric distribution can be found in Chakraborty and Gupta (2015) and Nadarajah and Bakar (2016). Further conclusions about the intermediate distributions can be obtained by following the results, from the Conway-Maxwell Poisson and binomial distributions, in Shmueli *et al.* (2005), Daly and Gaunt (2016), and Kadane (2016).

The theory is referred to Balakrishnan and Nevzorov (2003) and Johnson *et al.* (2005), for the basic definitions and general results in discrete distributions. See also Hardy *et al.* (1951), Piskunov (1979), Spivak (1994), Shorack (2000), Pawitan (2001), and R Core Team (2017), for other discussions and results.

2. Discrete power distributions

Let X be a discrete random variable (r.v.), with distribution function (d.f.) $F(\theta)$, where $F(x;\theta) = P_\theta(X \leq x)$, for all $-\infty < x < +\infty$. Then, $F(\theta)$ is nondecreasing and right continuous, $F(x;\theta) = F_+(x;\theta)$, with $F(-\infty;\theta) = 0$ and $F(+\infty;\theta) = 1$. We then denote by

$$\Delta F(x;\theta) = F(x;\theta) - F_-(x;\theta), \tag{1}$$

the probability mass of $F(\theta)$ at x. Let $\{x_i\}$ be the set of all discontinuities of $F(\theta)$, that define positive jumps $p_i(\theta) = \Delta F(x_i;\theta)$, such that $\sum_i p_i(\theta) = 1$. In particular, we have the step function $F(\theta) = \sum_i p_i(\theta) 1_{[x_i,\infty)}$. The sth moment about zero, of the r.v. X, is $\mu_s = E(X^s) = \sum_i x_i^s p_i(\theta)$.

We call $\{p_i(\theta)\}$, the probability distribution on the values $\{x_i\}$ of the r.v. X, with d.f. $F(\theta)$, the original distribution. Following Lehmann (1953) and Durrans (1992), we consider a parameter α for power functions, where $\alpha > 0$, that can be used for determining, from the d.f. $F(\theta)$, the power distribution $\{r_i(\theta,\alpha)\}$, of a discrete r.v. Z, with values $\{x_i\}$, d.f. $H(\theta,\alpha) = \sum_i r_i(\theta,\alpha) 1_{[x_i,\infty)}$, and sth moment $\omega_s = \sum_i x_i^s r_i(\theta,\alpha)$, and the intermediate distribution $\{q_i(\theta,\alpha)\}$, of a discrete r.v. Y, with values $\{x_i\}$, d.f. $G(\theta,\alpha)=\sum_i q_i(\theta,\alpha)1_{[x_i,\infty)}$, and sth moment $\nu_s=\sum_i x_i^s q_i(\theta,\alpha)$.

For simplicity, we suppose that $\{x_i\}$ only contains nonnegative values, and, conventionally, we consider x_i so that $x_{i-1}<x_i$.

We define by (1), the positive jumps

$$\Delta F(x_i;\theta,\alpha)=F(x_i;\theta)^\alpha - F_{\_}(x_i;\theta)^\alpha . \tag{2}$$

From (2), we can determine the parametric power distribution $\{r_i(\theta,\alpha)\}$, with d.f. equal to $H(\theta,\alpha)=\sum_i r_i(\theta,\alpha)1_{[x_i,\infty)}$, as

$$r_i(\theta,\alpha)=\Delta F(x_i;\theta,\alpha). \tag{3}$$

so that $\sum_i r_i(\theta,\alpha)=1$.

Whereas $\alpha=1$, the power distribution $\{r_i(\theta,\alpha)\}$, with d.f. $H(\theta,\alpha)$, coincides with the original distribution $\{p_i(\theta)\}$, with d.f. $F(\theta)$. We distinguish between the convex case, $\alpha>1$, and the concave case, $0<\alpha<1$, for the power distribution $\{r_i(\theta,\alpha)\}$, with d.f. $H(\theta,\alpha)$.

The moments ω_s, for the power d.f. $H(\theta,\alpha)$, cannot be expressed in closed form, and must be calculated or approximated, as explicit sums.

2.1. Power uniform distribution

The original uniform distribution $\{p_i\}$ is

$$p_i=m^{-1}, \tag{4}$$

for $\{x_i\}=\{0,1,\dots,m-1\}$.

The power uniform distribution $\{r_i(\alpha)\}$ can be obtained from (3) and (4), for a given $\alpha>0$, as

$$r_i(\alpha)=(m^{-1}i)^\alpha-(m^{-1}(i-1))^\alpha , \tag{5}$$

where $r_1(\alpha)=(m^{-1})^\alpha$ and $i=2,\dots,m$.

In Figure 1, we consider an original uniform distribution $\{p_i\}$, given by (4), with $m-1=20$, and the power uniform distributions $\{r_i(\alpha)\}$, given by (5), for $\alpha=4.35$ and $\alpha=0.45$. In Figure 2, we show the values for the mean, the variance, the skewness, and the kurtosis of the corresponding power uniform distributions $\{r_i(\alpha)\}$, along $\alpha>0$. Most importantly, the skewness changes sign between the convex case, $\alpha>1$, and the concave case, $0<\alpha<1$.

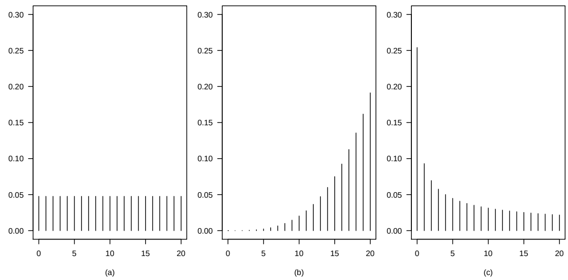

Figure 1 – An original uniform distribution $\{p_i\}$, in panel (a), where $m-1 = 20$, and power uniform distributions $\{r_i(\alpha)\}$, for $m-1=20$ and $\alpha=4.35$, in panel (b), and for $m-1=20$ and $\alpha=0.45$, in panel (c).

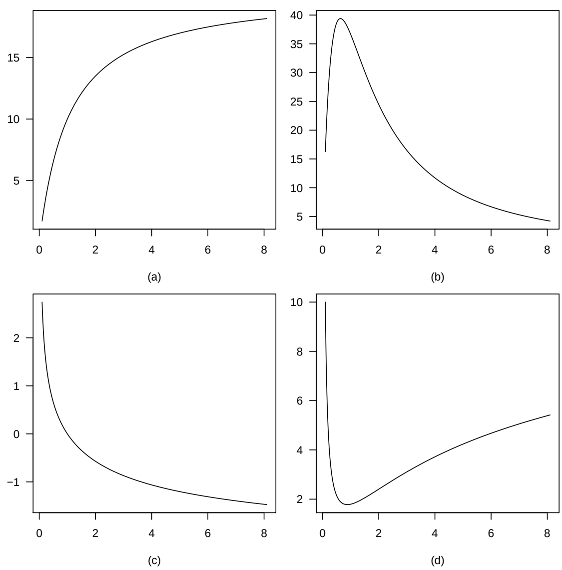

Figure 2 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the power uniform distributions $\{r_i(\alpha)\}$, for $m-1=20$ and $\alpha>0$.

2.2. *Power binomial distribution*

The original binomial distribution $\{p_i(\theta)\}$, of the number of successes in m Bernoulli trials, is

$$p_i(\theta) = \binom{m}{x_i}\theta^{x_i}(1-\theta)^{m-x_i}, \tag{6}$$

for $\{x_i\} = \{0, 1, \ldots, m\}$, where $0 < \theta < 1$.

The power binomial distribution $\{r_i(\theta, \alpha)\}$, can be obtained from (3) and (6), for a given $\alpha > 0$, as

$$r_i(\theta,\alpha) = \left(\sum_{j=1}^{i}\binom{m}{x_j}\theta^{x_j}(1-\theta)^{m-x_j}\right)^{\alpha} - \left(\sum_{j=1}^{i-1}\binom{m}{x_j}\theta^{x_j}(1-\theta)^{m-x_j}\right)^{\alpha}, \tag{7}$$

where $r_1(\theta,\alpha) = \left(\binom{m}{x_1}\theta^{x_1}(1-\theta)^{m-x_1}\right)^{\alpha}$ and $i = 2, 3, \ldots, m+1$.

In Figure 3, we consider an original binomial distribution $\{p_i(\theta)\}$, given by (6), with $m = 20$ and $\theta = 0.75$, and the power binomial distributions $\{r_i(\theta, \alpha)\}$, given by (7), for $\alpha = 7.8$ and $\alpha = 0.25$.

In Figure 4, we show the values for the mean, the variance, the skewness, and the kurtosis of the corresponding power binomial distributions $\{r_i(\theta, \alpha)\}$, along $\alpha > 0$. The power binomial distribution is a flexible distribution for situations characterized by overdispersion, and also by underdispersion.

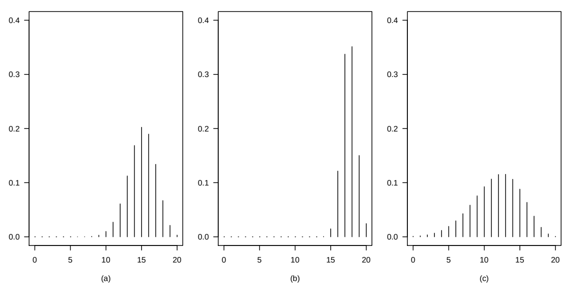

Figure 3 – An original binomial distribution $\{p_i(\theta)\}$, in panel (a), where $m = 20$ and $\theta = 0.75$, and power binomial distributions $\{r_i(\theta, \alpha)\}$, for $m = 20$, $\theta = 0.75$, and $\alpha = 7.8$, in panel (b), and for $m = 20$, $\theta = 0.75$, and $\alpha = 0.25$, in panel (c).

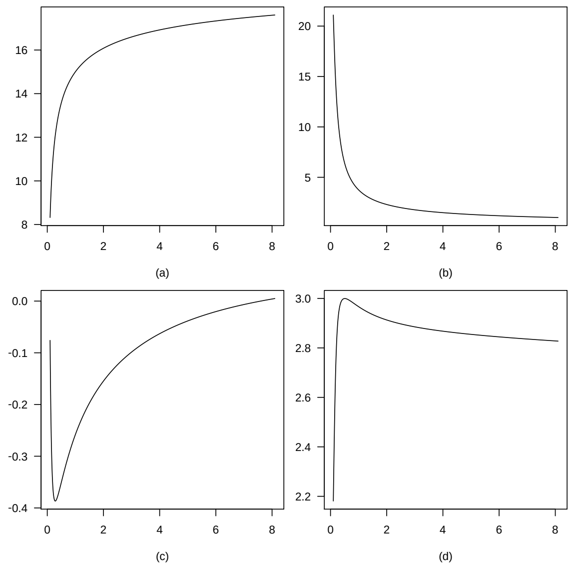

Figure 4 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the power binomial distributions $\{r_i(\theta,\alpha)\}$, for $m = 20$, $\theta = 0.75$, and $\alpha > 0$.

2.3. Power Poisson distribution

The original Poisson distribution $\{p_i(\theta)\}$, used as a limiting distribution, and for the occurrence of rare events, is

$$p_i(\theta) = \frac{e^{-\theta}\theta^{x_i}}{x_i!}, \tag{8}$$

for $\{x_i\} = \{0, 1, \ldots\}$, where $\theta > 0$.

The power Poisson distribution $\{r_i(\theta,\alpha)\}$ can be obtained from (3) and (8), for a given $\alpha > 0$, as

$$r_i(\theta,\alpha) = \left(\sum_{j=1}^{i} \frac{e^{-\theta}\theta^{x_j}}{x_j!}\right)^{\alpha} - \left(\sum_{j=1}^{i-1} \frac{e^{-\theta}\theta^{x_j}}{x_j!}\right)^{\alpha}, \tag{9}$$

where $r_1(\theta,\alpha) = (x_1!)^{-\alpha}\big(e^{-\theta}\theta^{x_1}\big)^{\alpha}$ and $i = 2, 3, \ldots$.

In Figure 5, we consider an original Poisson distribution $\{p_i(\theta)\}$, given by (8), with $\theta = 7.75$, and the power Poisson distributions $\{r_i(\theta,\alpha)\}$, given by (9), for $\alpha = 6.8$ and $\alpha = 0.37$. In Figure 6, we show the values for the mean, the variance, the skewness, and the kurtosis of all the corresponding power Poisson distributions $\{r_i(\theta,\alpha)\}$, along $\alpha > 0$. The power Poisson distribution is a flexible distribution for situations characterized by overdispersion, and also by underdispersion.

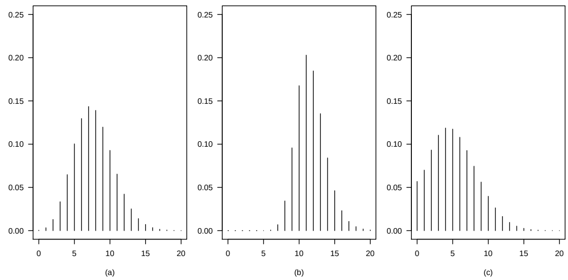

Figure 5 – An original Poisson distribution $\{p_i(\theta)\}$, in panel (a), where $\theta = 6.5.75$, and power Poisson distributions $\{r_i(\theta,\alpha)\}$, for $\theta = 7.75$ and $\alpha = 6.8$, in panel (b), and for $\theta = 7.75$ and $\alpha = 0.37$, in panel (c).

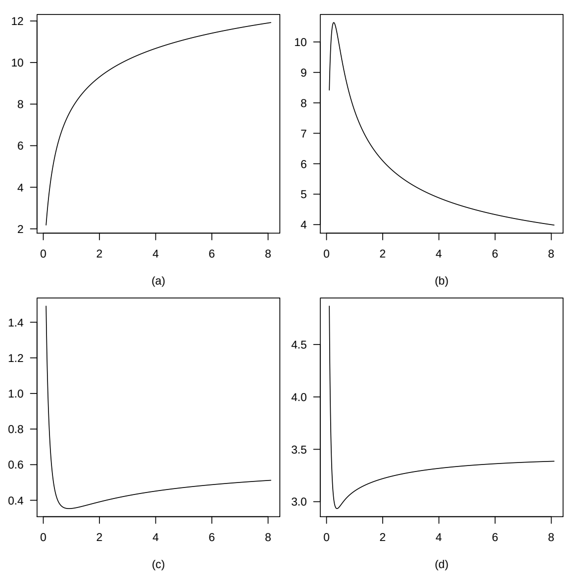

Figure 6 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the power Poisson distributions $\{r_i(\theta,\alpha)\}$, for $\theta = 7.75$ and $\alpha > 0$.

2.4. *Power negative binomial distribution*

The original negative binomial distribution $\{p_i(\theta)\}$, of the number of failures which occur in a sequence of Bernoulli trials, with probability of success θ, before a target number of successes η is reached, is

$$p_i(\theta) = \binom{\eta + x_i - 1}{x_i} (1-\theta)^{\eta} \theta^{x_i}, \qquad (10)$$

for $\{x_i\} = \{0, 1, \ldots, \}$, where $\eta > 0$ may be a real value and $0 < \theta < 1$.

The power negative binomial distribution $\{r_i(\theta, \alpha)\}$ can be obtained from (3) and (10), for a given $\alpha > 0$, as

$$r_i(\theta, \alpha) = \left(\sum_{j=1}^{i} \binom{\eta + x_j - 1}{x_j} (1-\theta)^{\eta} \theta^{x_j} \right)^{\alpha} - \left(\sum_{j=1}^{i-1} \binom{\eta + x_j - 1}{x_j} (1-\theta)^{\eta} \theta^{x_j} \right)^{\alpha}, \qquad (11)$$

where $r_1(\theta, \alpha) = \left(\binom{\eta + x_1 - 1}{x_1} (1-\theta)^{\eta} \theta^{x_1} \right)^{\alpha}$ and $i = 2, 3, \ldots$.

In particular, the power Pascal distribution $\{r_i(\theta, \alpha)\}$ and the power geometric distribution $\{r_i(\theta, \alpha)\}$ can be deduced from (11), by taking an integer η and the integer $\eta = 1$, respectively. Interesting properties for the power geometric distribution may be deduced from Chakraborty and Gupta (2015) and Nadarajah and Bakar (2016). In particular, Chakraborty and Gupta (2015) studied the probability mass function, moments and an index of dispersion, quantiles and the median, and reliability characteristics. Nadarajah and Bakar (2016) study specific expansions, shape properties, the probability generating function, the moment generating function, and order statistics.

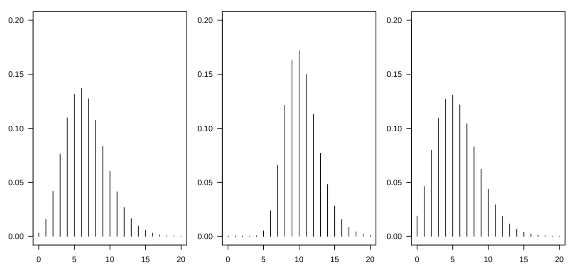

Figure 7 – An original negative binomial distribution $\{p_i(\theta)\}$, in panel (a), where $\eta = 6.67$ and $\theta = 0.75$, and power negative binomial distributions $\{r_i(\theta, \alpha)\}$, for $\eta = 6.67$, $\theta = 0.75$, and $\alpha = 5.32$, in panel (b), and for $\eta = 6.67$, $\theta = 0.75$, and $\alpha = 0.69$, in panel (c).

In Figure 7, we consider an original negative binomial distribution $\{p_i(\theta)\}$, given by (10), with $\eta = 6.67$ and $\theta = 0.75$, and the power negative binomial distributions $\{r_i(\theta, \alpha)\}$, given by (11), for $\alpha = 5.32$ and $\alpha = 0.69$. In Figure 8, we show the values for

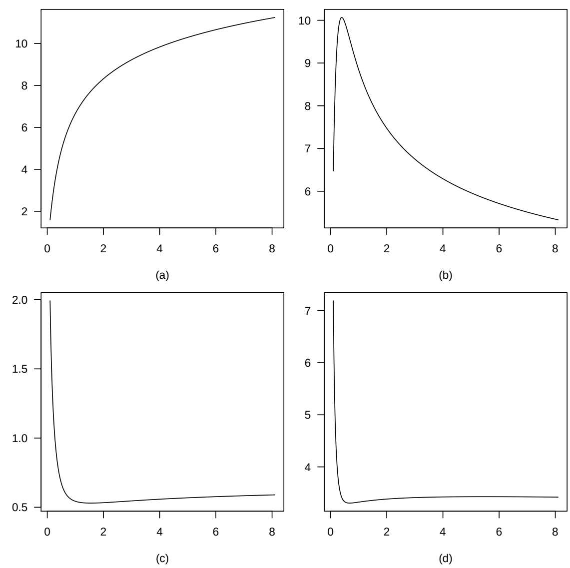

Figure 8 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the power negative binomial distributions $\{r_i(\theta,\alpha)\}$, for $\eta = 6.67$, $\theta = 0.75$, and $\alpha > 0$.

the mean, the variance, the skewness, and the kurtosis of all the corresponding power negative binomial distributions $\{r_i(\theta,\alpha)\}$, along $\alpha > 0$.

2.5. *Power hypergeometric distribution*

The original hypergeometric distribution $\{p_i(\theta)\}$ of the number of white balls, in a sample of m balls, without replacement, from a population of M balls, $M\theta$ of which are white and $M - M\theta$ are black, is

$$p_i(\theta) = \frac{\binom{M\theta}{x_i}\binom{M-M\theta}{m-x_i}}{\binom{M}{M\theta}}, \tag{12}$$

for $\{x_i\}$ ranging in $\max\big(0, m - M + M\theta\big) \le x_i \le \min\big(m, M\theta\big)$. where $m = 1,2,\ldots$ and $0 < \theta < 1$.

The power hypergeometric distribution $\{r_i(\theta,\alpha)\}$ can be obtained from (3) and (12),

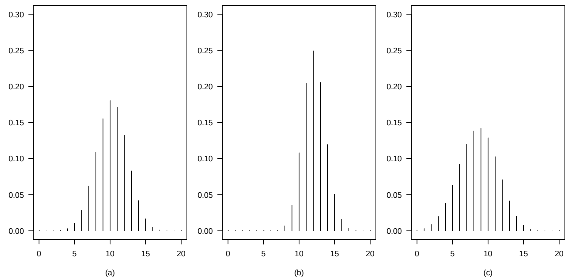

Figure 9 – An original hypergeometric distribution $\{p_i(\theta)\}$, in panel (a), where $M = 350$, $m = 20$, and $\theta = 0.51$, and power hypergeometric distributions $\{r_i(\theta, \alpha)\}$, for $M = 350$, $m = 20$, $\theta = 0.51$, and $\alpha = 3.15$, in panel (b), and for $M = 350$, $m = 20$, $\theta = 0.51$, and $\alpha = 0.47$, in panel (c).

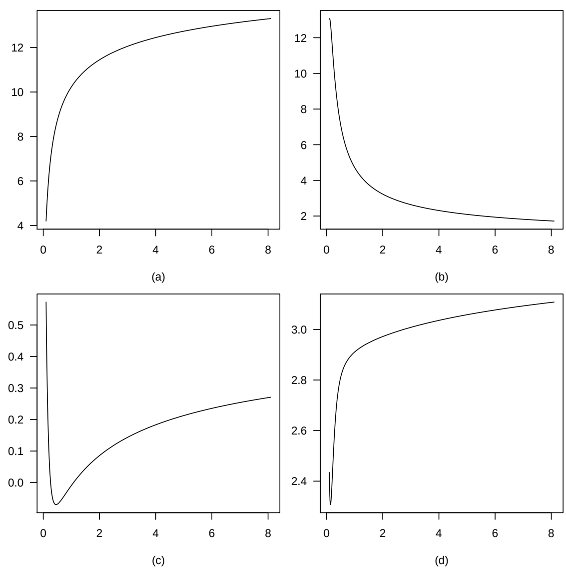

Figure 10 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the power hypergeometric distributions $\{r_i(\theta, \alpha)\}$, for $M = 350$, $m = 20$, $\theta = 0.51$, and $\alpha > 0$.

for a given $\alpha > 0$, as

$$r_i(\theta,\alpha) = \frac{\left(\sum_{j=1}^{i}\binom{M\theta}{x_j}\binom{M-M\theta}{m-x_j}\right)^{\alpha}}{\binom{M}{M\theta}^{\alpha}} - \frac{\left(\sum_{j=1}^{i-1}\binom{M\theta}{x_j}\binom{M-M\theta}{m-x_j}\right)^{\alpha}}{\binom{M}{M\theta}^{\alpha}}, \tag{13}$$

where $r_1(\theta,\alpha) = \binom{M}{M\theta}^{-\alpha}\left(\binom{M\theta}{x_1}\binom{M-M\theta}{m-x_1}\right)^{\alpha}$ and $i = 2,3,\ldots,m+1$.

In Figure 9, we consider an original hypergeometric distribution $\{p_i(\theta)\}$, given by (12), with $M = 350$, $m = 20$, and $\theta = 0.51$, and the power hypergeometric distributions $\{r_i(\theta,\alpha)\}$, given by (13), for $\alpha = 3.15$ and $\alpha = 0.47$. In Figure 10, we show the values for the mean, the variance, the skewness, and the kurtosis of all the corresponding power hypergeometric distributions $\{r_i(\theta,\alpha)\}$, along $\alpha > 0$.

3. Inequalities in moments

We study inequalities in moments by a specific application of the system of inequalities introduced in Jensen (1906). More precisely, we apply the well known Jensen's inequality to what is commonly thought of as weights in a mean of values, in the convex case, $\alpha > 1$, and the concave case, $0 < \alpha < 1$.

3.1. *Convex case*

We define $B_s = \sum_i x_i^s$ and we suppose that $B_s > 0$. We have that $B_s \min_i p_i(\theta) \leq \mu_s \leq B_s \max_i p_i(\theta)$ and $\min_i p_i(\theta) \leq B_s^{-1}\mu_s \leq \max_i p_i(\theta)$. Hence, we can choose a quantity $A(\theta) \geq (\min_i p_i(\theta))^{-1}$, so that $(B_s^{-1}A(\theta))\mu_s \geq 1$.

We introduce the sth order quantity $\tau_s = \sum_i x_i^s(\Delta F(x_i;\theta))^{\alpha} = \sum_i x_i^s p_i(\theta)^{\alpha}$.

In the convex case, $\alpha > 1$, we have that $(B_s^{-1}A(\theta))\mu_s \leq ((B_s^{-1}A(\theta))\mu_s)^{\alpha}$. The Jensen's inequality then determines the inequalities in moments

$$\mu_s \leq A(\theta)^{\alpha-1}\tau_s \leq A(\theta)^{\alpha-1}2^{\alpha-1}\omega_s. \tag{14}$$

When $(B_s^{-1}A(\theta))\mu_s = 1$, the quantity $A(\theta)^{\alpha-1}$ in (14) is the least upper bound. Of course, $\tau_s \leq 2^{\alpha-1}\omega_s$. See Appendix A.

Considering α^{-1} in place of α, where $\alpha > 1$, we have the inequalities in moments in the concave case, below.

3.2. *Concave case*

In the concave case, $0 < \alpha < 1$, we have that $(B_s^{-1}A(\theta))\mu_s \geq ((B_s^{-1}A(\theta))\mu_s)^{\alpha}$, where $A(\theta) \geq (\min_i p_i(\theta))^{-1}$. The Jensen's inequality then determines the inequalities in moments

$$\mu_s \geq A(\theta)^{\alpha-1}\tau_s \geq A(\theta)^{\alpha-1}2^{\alpha-1}\omega_s\,. \tag{15}$$

When $(B_s^{-1}A(\theta))\mu_s = 1$, the quantity $A(\theta)^{\alpha-1}$ in (15) is the greatest lower bound. Of course, $\tau_s \geq 2^{\alpha-1}\omega_s$. See Appendix A.

Considering α^{-1} in place of α, where $0 < \alpha < 1$, we have the inequalities in moments in the convex case.

3.3. *Distribution functions*

We define the step function $B = \sum_i 1_{[x_i,\infty)}$. We have that $B \min_i p_i(\theta) \leq F(\theta) \leq B \max_i p_i(\theta)$. Hence, we can choose a nondecreasing function $A(x;\theta)$, where $-\infty < x < +\infty$, so that $A(x_i;\theta) \geq (\min_i p_i(\theta))^{-1}$ and $(B(x_i)^{-1}A(x_i;\theta))F(x_i;\theta) \geq 1$, for all $\{x_i\}$.

We put the step function $K(\theta,\alpha) = \sum_i (\Delta F(x_i;\theta))^\alpha 1_{[x_i,\infty)} = \sum_i p_i(\theta)^\alpha 1_{[x_i,\infty)}$.

In the convex case, $\alpha > 1$, since

$$\frac{A(x_i;\theta)F(x_i;\theta)}{B(x_i)} \leq \left(\frac{A(x_i;\theta)F(x_i;\theta)}{B(x_i)}\right)^\alpha, \tag{16}$$

for all $\{x_i\}$, the Jensen's inequality determines the inequalities in d.f.'s

$$F(x;\theta) \leq A(x;\theta)^{\alpha-1}K(x;\theta,\alpha) \leq A(x;\theta)^{\alpha-1}2^{\alpha-1}H(x;\theta,\alpha), \tag{17}$$

where $K(x;\theta,\alpha) \leq 2^{\alpha-1}H(x;\theta,\alpha)$, for all $-\infty < x < +\infty$. See Appendix B.

In the concave case, $0 < \alpha < 1$, since

$$\frac{A(x_i;\theta)F(x_i;\theta)}{B(x_i)} \geq \left(\frac{A(x_i;\theta)F(x_i;\theta)}{B(x_i)}\right)^\alpha, \tag{18}$$

for all $\{x_i\}$, the Jensen's inequality determines the inequalities in d.f.'s

$$F(x;\theta) \geq A(x;\theta)^{\alpha-1}K(x;\theta,\alpha) \geq A(x;\theta)^{\alpha-1}2^{\alpha-1}H(x;\theta,\alpha), \tag{19}$$

where $K(x;\theta,\alpha) \geq 2^{\alpha-1}H(x;\theta,\alpha)$, for all $-\infty < x < +\infty$. See Appendix B.

When $(B(x_i)^{-1}A(x_i;\theta))F(x_i;\theta) = 1$, for all $\{x_i\}$, the values $A(x;\theta)^{\alpha-1}$ in (17), where $\alpha > 1$, are the least upper bounds and the values $A(x;\theta)^{\alpha-1}$ in (19), where $0 < \alpha < 1$, are the greatest lower bounds, for all $-\infty < x < +\infty$.

Considering α^{-1} in place of α, where $\alpha > 1$, we have the inequalities in d.f.'s in the concave case, and considering α^{-1} in place of α, where $0 < \alpha < 1$, we have the inequalities in d.f.'s in the convex case.

3.4. *Intermediate distributions*

The concept of intermediate distributions is based on the fact that these distributions lie, in some sense, between an original distribution and a power distribution.

From inequalities in moments (14) and (15), and inequalities in d.f.'s (17) and (19), the parametric intermediate distribution $\{q_i(\theta,\alpha)\}$, with d.f. $G(\theta,\alpha)=\sum_i q_i(\theta,\alpha)1_{[x_i,\infty)}$ and sth moment $\nu_s=\sum_i x_i^s q_i(\theta,\alpha)$, can be defined as

$$q_i(\theta,\alpha)=\frac{(\Delta F(x_i;\theta))^\alpha}{\sum_j(\Delta F(x_j;\theta))^\alpha}, \tag{20}$$

where the jump $\Delta F(x_i;\theta)$ is according to (1).
It simply follows that $q_i(\theta,\alpha)=\left(\sum_j p_j(\theta)^\alpha\right)^{-1}p_i(\theta)^\alpha$ and $\sum_i q_i(\theta,\alpha)=1$. In inequalities (14) and (15), we may note that $\nu_s=\left(\sum_j(\Delta F(x_j;\theta))^\alpha\right)^{-1}\tau_s=\left(\sum_j p_j(\theta)^\alpha\right)^{-1}\tau_s$. Similarly, in inequalities (17) and (19), we have that

$$G(\theta,\alpha)=\Big(\sum_j(\Delta F(x_j;\theta))^\alpha\Big)^{-1}K(\theta,\alpha)=\Big(\sum_j p_j(\theta)^\alpha\Big)^{-1}K(\theta,\alpha),$$

where $\alpha>0$.

Whereas $\alpha=1$, the intermediate distribution $\{q_i(\theta,\alpha)\}$, with d.f. $G(\theta,\alpha)$, coincide with the original distributions $\{p_i(\theta)\}$, with d.f. $F(\theta)$, since $\sum_i p_i(\theta)=1$. It is important to distinguish between the convex case, $\alpha>1$, and the concave case, $0<\alpha<1$.

The moments ν_s, for the intermediate d.f. $G(\theta,\alpha)$, cannot be expressed in closed form, and must be calculated or approximated, as explicit sums.

Recalling also the situation of Figure 1, we may observe that an original uniform distribution coincides with all intermediate uniform distributions $q_i\{\alpha\}$, for all $\alpha>0$.

The intermediate binomial distribution $\{q_i(\theta,\alpha)\}$ can be obtained from (6) and (20), for a given $\alpha>0$, as

$$q_i(\theta,\alpha)=\frac{\left(\binom{m}{x_i}\theta^{x_i}(1-\theta)^{m-x_i}\right)^\alpha}{\sum_{j=1}^{m+1}\left(\binom{m}{x_j}\theta^{x_j}(1-\theta)^{m-x_j}\right)^\alpha}, \tag{21}$$

where $i=1,2,\ldots,m+1$.

In Figure 11, we consider the original binomial distribution $\{p_i(\theta)\}$, given by (6), with $m=20$ and $\theta=0.75$, and the intermediate binomial distributions $\{q_i(\theta,\alpha)\}$, given by (21), for $m=20$, $\theta=0.75$, $\alpha=4.15$, and $\alpha=0.35$. In Figure 12, we show the values for the mean, the variance, the skewness, and the kurtosis of all the corresponding intermediate binomial distributions $\{q_i(\theta,\alpha)\}$, along $\alpha>0$.

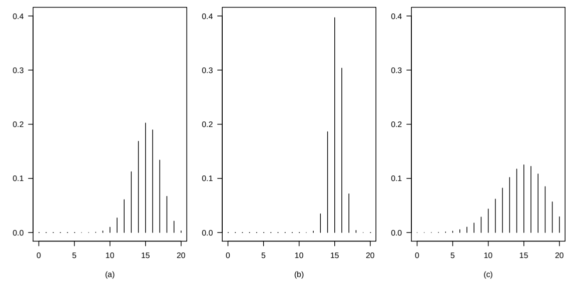

Figure 11 – An original binomial distribution $\{p_i(\theta)\}$, in panel (a), where $m = 20$ and $\theta = 0.75$, and intermediate binomial distributions $\{q_i(\theta, \alpha)\}$, for $m = 20$, $\theta = 0.75$, and $\alpha = 4.15$, in panel (b), and for $m = 20$, $\theta = 0.75$, and $\alpha = 0.35$, in panel (c).

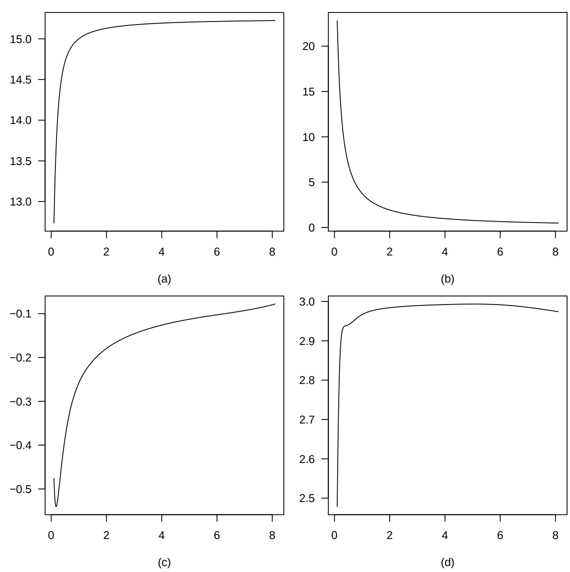

Figure 12 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the intermediate binomial distributions $\{q_i(\theta, \alpha)\}$, for $m = 20$, $\theta = 0.75$, and $\alpha > 0$.

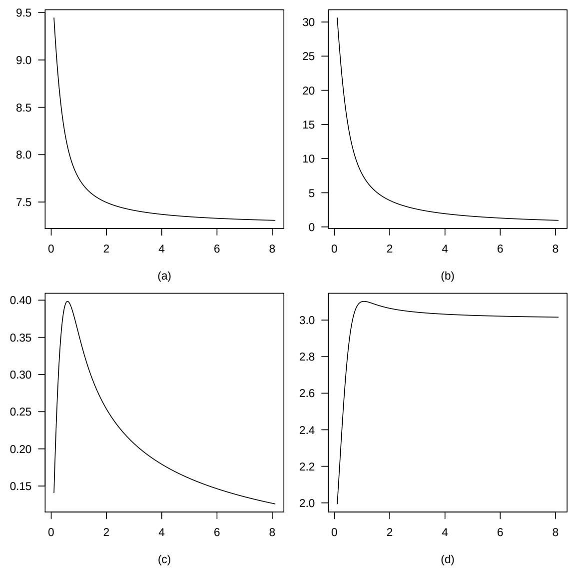

Figure 13 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the intermediate Poisson distributions $\{q_i(\theta, \alpha)\}$, for $\theta = 7.75$, and $\alpha > 0$.

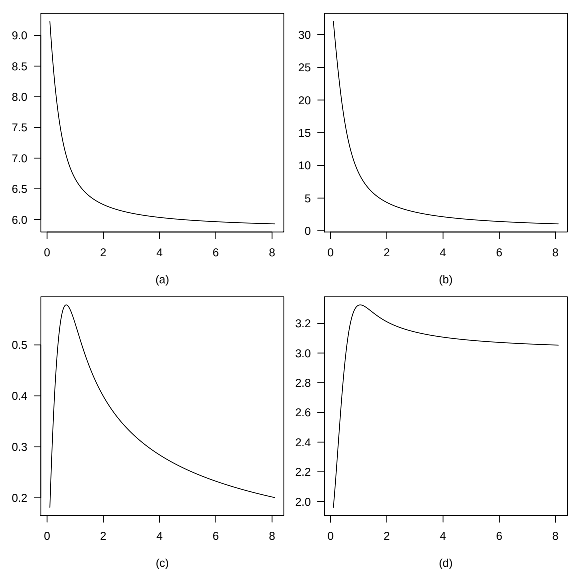

Figure 14 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the intermediate negative binomial distributions $\{q_i(\theta, \alpha)\}$, for $\eta = 6.67$, $\theta = 0.75$, and $\alpha > 0$.

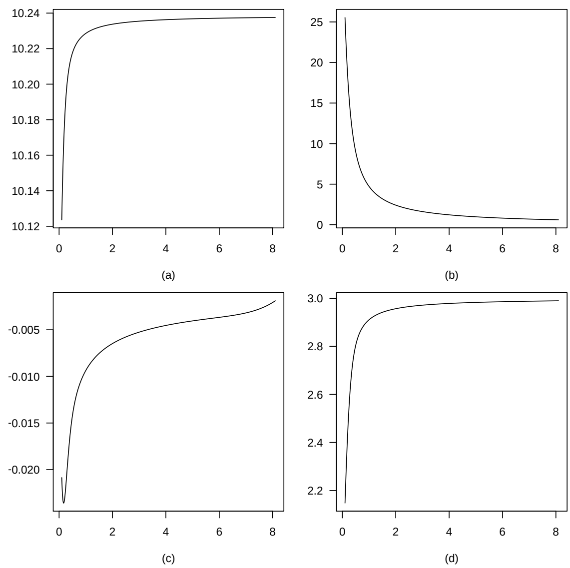

Figure 15 – Mean, in panel (a), variance, in panel (b), skewness, in panel (c), kurtosis, in panel (d), of the intermediate hypergeometric distributions $\{q_i(\theta,\alpha)\}$, for $M = 350$, $m = 20$, $\theta = 0.51$, and $\alpha > 0$.

Similarly, the intermediate Poisson, the intermediate negative binomial, the intermediate hypergeometric distributions $\{q_i(\theta,\alpha)\}$, can be obtained from (20), for a given $\alpha > 0$. The intermediate binomial and Poisson distributions are proportional to the corresponding Conway-Maxwell binomial and Poisson distributions.

In Figures 13, 14, and 15, we show the values for the mean, the variance, the skewness, and the kurtosis of intermediate Poisson, intermediate negative binomial, and intermediate hypergeometric distributions $\{q_i(\theta,\alpha)\}$, along $\alpha > 0$.

4. STOCHASTIC ORDERS

We refer to Müller and Stoyan (2002, chapter 1), and Belzunce *et al.* (2016, chapter 2), for the basic theory on univariate stochastic orders.

Given an original d.f. $F(\theta)$, we simply have that both the power d.f. $H(\theta,\alpha)$ and the intermediate d.f. $G(\theta,\alpha)$, where $\alpha > 0$, satisfy the usual stochastic order, as $H(\theta,\alpha_1) \geq H(\theta,\alpha_2)$ and $G(\theta,\alpha_1) \geq G(\theta,\alpha_2)$, if $\alpha_1 < \alpha_2$, respectively.

In particular, the power d.f. $H\theta,(\alpha)$, and the intermediate d.f. $G(\theta,\alpha)$ increase with respect to d.f. $F(\theta)$, for the concave case, $0 < \alpha < 1$, as α goes to 0, and decrease with respect to d.f. $F(\theta)$, as α increases along the convex case, $\alpha > 1$. The analytic behaviour is faster for the power d.f. $H(\theta,\alpha)$, for all $\alpha > 0$.

Since the values in $\{x_i\}$ are nonnegative, the sth moments ω_s and ν_s, of the power d.f. $H(\theta,\alpha)$ and the intermediate d.f. $G(\theta,\alpha)$, increase, as α increases, along $\alpha > 0$.

5. UNIMODALITY

We refer to Dharmadhikari and Joag-Dev (1988, chapter 4), for the basic theory on unimodality of discrete distributions.

Considering the original distribution $\{p_i(\theta)\}$, we say that $\{p_i(\theta)\}$ is k-unimodal, about a mode k, if there exist at least one integer k, such that $p_i(\theta) \geq p_{i-1}(\theta)$, for $i \leq k$, and $p_{i+1}(\theta) \leq p_i(\theta)$, for $i \geq k$. A distribution $\{p_i(\theta)\}$ is strongly unimodal if and only if the sequence $\{p_i(\theta)\}$ is log-concave, that is $p_i(\theta)^2 \geq p_{i+1}(\theta)p_{i-1}(\theta)$, for all i, namely $\log p_i(\theta) \geq 2^{-1}(\log p_{i+1}(\theta) + \log p_{i-1}(\theta))$, for all i.

Unimodality of the original distribution $\{p_i(\theta)\}$ implies unimodality for the power distribution $\{r_i(\theta,\alpha)\}$, given by (3), and the intermediate distribution $\{q_i(\alpha)\}$, given by (20), where $\alpha > 0$. If the power distribution $\{r_i(\theta,\alpha)\}$ and the intermediate distribution $\{q_i(\theta,\alpha)\}$ are unimodal, then strong unimodality of the original distribution $\{p_i(\theta)\}$ implies strong unimodality for $\{r_i(\theta,\alpha)\}$ and $\{q_i(\theta,\alpha)\}$, respectively, where $\alpha > 0$.

6. INFERENCE USING LIKELIHOOD

6.1. *Power distributions*

Let $\{z_k\}$ be a sample of n i.i.d. observations from the r.v. Z, with power distribution $\{r_i(\theta,\alpha)\}$, given by (3), on the values $\{x_i\}$. Since a sample value z_k is drawn by choosing a value from $\{x_i\}$, we may write $z_k = x_{i(k)}$. The power log-likelihood $l(\theta,\alpha) = \log L(\theta,\alpha)$ then is

$$l(\theta,\alpha) = \sum_k \log r_{i(k)}(\theta,\alpha). \tag{22}$$

The score function $S(\theta,\alpha)$ is the gradient vector $S(\theta,\alpha) = (S(\theta,\alpha)_1, S(\theta,\alpha)_2)$, with components $S(\theta,\alpha)_1 = (\partial/\partial\theta)l(\theta,\alpha)$ and $S(\theta,\alpha)_2 = (\partial/\partial\alpha)l(\theta,\alpha)$. In particular, for the power score function, we have that

$$S(\theta,\alpha)_1 = \sum_k \left(\frac{1}{r_{i(k)}(\theta,\alpha)} \frac{\partial r_{i(k)}(\theta,\alpha)}{\partial\theta} \right), \tag{23}$$

$$S(\theta,\alpha)_2 = \sum_k \left(\frac{1}{r_{i(k)}(\theta,\alpha)} \frac{\partial r_{i(k)}(\theta,\alpha)}{\partial\alpha} \right). \tag{24}$$

6.2. *Intermediate distributions*

Let $\{y_k\}$ be a sample of n i.i.d. observations from the r.v. Y, with intermediate distribution $\{q_i(\theta,\alpha)\}$, given by (20), on the values $\{x_i\}$, where $y_k = x_{i(k)}$. The intermediate log-likelihood $l(\theta,\alpha)=\log L(\theta,\alpha)$ then is

$$l(\theta,\alpha)=\sum_k \log q_{i(k)}(\theta,\alpha). \tag{25}$$

The intermediate score function $S(\theta,\alpha)=(S(\theta,\alpha)_1, S(\theta,\alpha)_2)$ can be obtained by substituting $r_{i(k)}(\theta,\alpha)$ with $q_{i(k)}(\theta,\alpha)$, in the components (23) and (24).

6.3. *Information*

For the power log-likelihood $l(\theta,\alpha)$, given by (22), the expected information matrix $\mathscr{I}(\theta,\alpha)$ can be obtained, from minus the Hessian of $l(\theta,\alpha)$, as

$$\mathscr{I}(\theta,\alpha)=\begin{pmatrix}\mathscr{I}(\theta,\alpha)_{11} & \mathscr{I}(\theta,\alpha)_{12}\\ \mathscr{I}(\theta,\alpha)_{21} & \mathscr{I}(\theta,\alpha)_{22}\end{pmatrix}, \tag{26}$$

where

$$\begin{aligned}\mathscr{I}(\theta,\alpha)_{11} &= E_{(\theta,\alpha)}\Big(-\frac{\partial S(\theta,\alpha)_1}{\partial\theta}\Big)\\ &= -n\sum_i\Big(\frac{1}{r_i(\theta,\alpha)}\Big(\frac{\partial r_i(\theta,\alpha)}{\partial\theta}\Big)^2-\frac{\partial^2 r_i(\theta,\alpha)}{\partial\theta^2}\Big), \qquad (27)\\ \mathscr{I}(\theta,\alpha)_{22} &= E_{(\theta,\alpha)}\Big(-\frac{\partial S(\theta,\alpha)_2}{\partial\alpha}\Big)\\ &= -n\sum_i\Big(\frac{1}{r_i(\theta,\alpha)}\Big(\frac{\partial r_i(\theta,\alpha)}{\partial\alpha}\Big)^2-\frac{\partial^2 r_i(\theta,\alpha)}{\partial\alpha^2}\Big), \qquad (28)\\ \mathscr{I}(\theta,\alpha)_{21} &= E_{(\theta,\alpha)}\Big(-\frac{\partial S(\theta,\alpha)_2}{\partial\theta}\Big)\\ &= -n\sum_i\Big(\frac{1}{r_i(\theta,\alpha)}\frac{\partial r_i(\theta,\alpha)}{\partial\theta}\frac{\partial r_i(\theta,\alpha)}{\partial\alpha}-\frac{\partial^2 r_i(\theta,\alpha)}{\partial\theta\partial\alpha}\Big), \qquad (29)\\ \mathscr{I}(\theta,\alpha)_{12} &= E_{(\theta,\alpha)}\Big(-\frac{\partial S(\theta,\alpha)_1}{\partial\alpha}\Big)\\ &= \mathscr{I}(\theta,\alpha)_{21}. \qquad (30)\end{aligned}$$

For the intermediate log-likelihood $l(\theta,\alpha)$, given by (25), the expected information matrix $\mathscr{I}(\theta,\alpha)$ can be obtained from (26), by substituting $r_i(\theta,\alpha)$ with $q_i(\theta,\alpha)$ in the elements (27), (28), (29), and (30).

6.4. *Asymptotics*

Applying Wald (1949), under some regularity conditions, and by using the strong law of large numbers, the convergence, with probability 1, of the m.l.e.'s $(\hat{\theta},\hat{\alpha})$ to (θ_0,α_0), as $n \to \infty$, can be shown.

Following Lehmann and Casella (1998, chapter 6), we can see that the third derivatives of the power and intermediate log-likelihoods $l(\theta,\alpha)$, given by (22) and (25), exist and can be bounded, in absolute value, by specific functions with finite expected values. The information matrices $\mathscr{I}(\theta,\alpha)$, defined as (26), for the log-likelihoods (22) and (25), have finite elements (27), (28), (29), and (30), and are positive definite. We also have that $n^{1/2}((\hat{\theta},\hat{\alpha})-(\theta_0,\alpha_0))$ is asymptotically normal with mean $(0,0)$ and covariance matrices $\mathscr{I}(\theta_0,\alpha_0)^{-1}$, as $n \to \infty$. Furthermore, we have that $\hat{\alpha}$ and $\hat{\theta}$ in $(\hat{\theta},\hat{\alpha})$ are asymptotically efficient, in the sense that $n^{1/2}(\hat{\theta}-\theta_0)$ and $n^{1/2}(\hat{\alpha}-\alpha_0)$ have asymptotic variances $\mathscr{I}(\theta_0,\alpha_0)^{-1}_{11}$ and $\mathscr{I}(\theta_0,\alpha_0)^{-1}_{22}$, respectively, as $n \to \infty$.

7. Simulation experiments

We performed simulation experiments to study the bias and the mean square error of the m.l.e'.s $(\hat{\theta},\hat{\alpha})$ in the power distributions $\{r_i(\theta,\alpha)\}$, given by (3), and the intermediate distributions $\{q_i(\theta,\alpha)\}$, given by (20). We always simulated 10000 replications of the same experiment that consists in drawing a sample of n i.i.d. observations, from a distribution $\{r_i(\alpha)\}$, $\{r_i(\theta,\alpha)\}$ or $\{q_i(\theta,\alpha)\}$, where $n = 5, 10, 20, 50, 100$ and $\alpha > 0$. In all the simulations we obtained, we have a smaller mean square error for the m.l.e $\hat{\theta}$ in convex cases, $\alpha > 1$, and a smaller mean square error for the m.l.e $\hat{\alpha}$ in concave cases, $0 < \alpha < 1$.

We used the computational environment for statistics R, by R Core Team (2017). In particular, in the R "optim", we considered the algorithm of Brent (1973, chapter 5), for the univariate optimization problems $\min_\alpha(-l(\alpha))$, with a log-likelihood of the form $l(\alpha)$, and the algorithm of Nelder and Mead (1965), for the optimization problems $\min_{(\theta,\alpha)}(-l(\theta,\alpha))$, with a log-likelihood of the form $l(\theta,\alpha)$. The numerical algorithms of Brent (1973, chapter 5), and Nelder and Mead (1965) do not require the derivative and the gradient, respectively, of the corresponding log-likelihoods $l(\alpha)$ and $l(\theta,\alpha)$.

In Table 1, we provide the simulation results about the m.l.e. $\hat{\alpha}$ of α, in the power uniform distribution $\{r_i(\alpha)\}$, given by (5), with $m-1=10$, $\alpha = 4.35$ and $\alpha = 0.45$. The performance of $\hat{\alpha}$ improves, as n increases, without a significant effect due to $m-1$.

TABLE 1

Bias and mean square error of the m.l.e. $\hat{\alpha}$, for the power uniform distribution $\{r_i(\alpha)\}$, with $m-1=10$, $\alpha = 4.35$, and $\alpha = 0.45$.

n	$b(\hat{\alpha})$	$mse(\hat{\alpha})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	0.1074	0.0730	-0.1408	0.0254
10	0.0996	0.0712	-0.1370	0.0243
20	0.1004	0.0702	-0.1314	0.0226
50	0.0965	0.0690	-0.1225	0.0201
100	0.0925	0.0691	-0.1149	0.0179

In Tables 2 and 3, we consider the simulation results about the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$ of (θ, α), in the power binomial distribution $\{r_i(\theta, \alpha)\}$, given by (7), with $\theta = 0.75$, $m = 10$, $\alpha = 7.80$, and $\alpha = 0.25$. The performance of $\hat{\theta}$ improves, as n increases, with a significant effect due to m. The behaviour of $\hat{\alpha}$ shows a positive bias.

TABLE 2

Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the power binomial distribution $\{r_i(\theta, \alpha)\}$, with $\theta = 0.75$, $m = 20$, and $\alpha = 7.80$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	0.0006	0.0007	0.0749	0.1096
10	-0.0020	0.0005	0.0869	0.1195
20	-0.0027	0.0003	0.0913	0.1224
50	-0.0016	0.0002	0.1158	0.1341
100	-0.0005	0.0001	0.1313	0.1425

TABLE 3

Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the power binomial distribution $\{r_i(\theta, \alpha)\}$, with $\theta = 0.75$, $m = 20$, and $\alpha = 0.25$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	-0.0051	0.0023	0.0069	0.0025
10	-0.0041	0.0019	0.0048	0.0024
20	-0.0030	0.0016	0.0036	0.0022
50	-0.0018	0.0012	0.0039	0.0020
100	-0.0008	0.0010	0.0040	0.0018

In Tables 4 and 5, we provide the simulation results about the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$ of (θ, α), in the power Poisson distribution $\{r_i(\theta, \alpha)\}$, given by (9), with $\theta = 7.75$, $\alpha = 6.80$, and $\alpha = 0.37$. The performance of $\hat{\theta}$ and $\hat{\alpha}$ improve, as n increases, and their behaviour shows bias.

TABLE 4

Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the power Poisson distribution $\{r_i(\theta, \alpha)\}$, with $\theta = 7.75$ and $\alpha = 6.80$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	0.0422	0.2397	0.2171	0.3834
10	0.0399	0.2081	0.2010	0.3734
20	0.0143	0.1660	0.2197	0.3596
50	-0.0088	0.1237	0.2151	0.3377
100	-0.0389	0.0979	0.2310	0.3301

TABLE 5
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the power Poisson distribution $\{r_i(\theta, \alpha)\}$, with $\theta = 7.75$ and $\alpha = 0.37$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	0.0914	0.2304	0.1324	0.0786
10	0.1421	0.2295	0.0595	0.0373
20	0.1746	0.2351	0.0192	0.0171
50	0.1873	0.2476	-0.0037	0.0076
100	0.1950	0.2327	-0.0112	0.0051

In Tables 6 and 7, we provide the simulation results about the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$ of (θ, α), in the power negative binomial distribution $\{r_i(\theta, \alpha)\}$, given by (11), with $\eta = 6.67$, $\theta = 0.75$, $\alpha = 5.32$, and $\alpha = 0.69$. The performance of $\hat{\theta}$ improves, as n increases. The behaviour of $\hat{\alpha}$ shows a negative bias.

TABLE 6
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the power negative binomial distribution $\{r_i(\theta, \alpha)\}$, with $\eta = 6.67$, $\theta = 0.75$, and $\alpha = 5.32$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	0.0087	0.0005	0.0567	0.1694
10	0.0062	0.0003	0.0343	0.1717
20	0.0040	0.0002	-0.0028	0.1742
50	0.0014	0.0001	-0.0693	0.1607
100	-0.00001	0.00009	-0.1279	0.1501

TABLE 7
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the power negative binomial distribution $\{r_i(\theta, \alpha)\}$, with $\eta = 6.67$, $\theta = 0.75$, and $\alpha = 0.69$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	-0.0141	0.0015	-0.1169	0.0175
10	-0.0118	0.0010	-0.1201	0.0181
20	-0.0112	0.0007	-0.1239	0.0188
50	-0.0129	0.0005	-0.1300	0.0197
100	-0.0149	0.0005	-0.1336	0.0201

In Tables 8 and 9, we provide the simulation results about the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$ of (θ, α), in the power hypergeometric distribution $\{r_i(\theta, \alpha)\}$, given by (13), with $\theta = 0.75$, $M = 350$, $\alpha = 3.15$, and $\alpha = 0.47$. The performance of $\hat{\theta}$ improves, as n increases. The behaviour of $\hat{\alpha}$ shows a negative bias.

TABLE 8
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the power hypergeometric distribution $\{r_i(\theta, \alpha)\}$, with $\theta = 0.75$, $M = 350$, and $\alpha = 3.15$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	0.0135	0.0064	0.0135	0.0418
10	0.0053	0.0018	0.0087	0.0525
20	0.0041	0.0006	0.0160	0.0661
50	0.0014	0.0001	0.0056	0.0843
100	0.0009	0.00004	-0.0072	0.0881

TABLE 9
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the power hypergeometric distribution $\{r_i(\theta, \alpha)\}$, with $\theta = 0.75$, $M = 350$, and $\alpha = 0.47$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	-0-0841	0.0081	-0.0848	0.0083
10	-0.0726	0.0063	-0.0857	0.0085
20	-0.0491	0.0039	-0.0917	0.0101
50	0.0125	0.0009	-0.1123	0.0160
100	0.0148	0.0004	-0.1103	0.0157

In Tables 10 and 11, we provide the simulation results, about the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$ of (θ, α), in the intermediate binomial distribution $\{q_i(\theta, \alpha)\}$, given by (21), with $\theta = 0.5$, $m = 20$, $\alpha = 4.15$, and $\alpha = 0.35$. The performances of $\hat{\theta}$ and $\hat{\alpha}$ improve, as n increases, showing the bias of $\hat{\alpha}$.

TABLE 10
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the intermediate binomial distribution $\{q_i(\theta, \alpha)\}$, with $\theta = 0.5$, $m = 20$, and $\alpha = 4.15$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	-0.0001	0.0006	0.0249	0.0850
10	-0.0001	0.0003	0.0081	0.0826
20	-0.0004	0.0002	-0.0181	0.0790
50	-0.0003	0.0001	-0.0568	0.0823
100	-0.0003	0.0001	-0.0864	0.0856

TABLE 11
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the intermediate binomial distribution $\{q_i(\theta,\alpha)\}$, with $\theta = 0.5$, $m = 20$, and $\alpha = 0.35$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	-0.0817	0.0075	-0.0845	0.0081
10	-0.0775	0.0068	-0.0868	0.0085
20	-0.0725	0.0061	-0.0899	0.0090
50	-0.0640	0.0049	-0.0917	0.0094
100	-0.0556	0.0039	-0.0941	0.0099

7.1. *Stochastic approximation*

Sometimes, intermediate distributions $\{q_i(\theta,\alpha)\}$ must be estimated, by approximating their denominator, that is a normalizing constant. An intermediate distribution $\{q_i(\theta,\alpha)\}$ may be estimated as

$$\hat{q}_i(\theta,\alpha)) = \frac{n\, p_i(\theta)^\alpha}{\sum_k p_{i(k)}(\theta)^{\alpha-1}}\,. \tag{31}$$

Monte Carlo integration shows that, in (31), the approximation of the normalizing constant satisfies $E_\theta\Big(n^{-1}\sum_k p_{i(k)}(\theta)^{-1} p_{i(k)}(\theta)^\alpha\Big) = \sum_i p_i(\theta)^\alpha$, with a variance that decreases to 0, as $n \to \infty$. See Ross (2013, chapter 9).

In Tables 12 and 13, we provide the simulation results about the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$ of (θ,α), in the intermediate binomial distribution $\{\hat{q}_i(\theta,\alpha)\}$, given by (21), with the approximation (31) of the normalizing constant, $\theta = 0.5$, $m = 20$, $\alpha = 4.15$, and $\alpha = 0.35$.

TABLE 12
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the intermediate binomial distribution $\{\hat{q}_i(\theta,\alpha)\}$, with an approximation of the normalizing constant, $\theta = 0.5$, $m = 20$, and $\alpha = 4.15$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	0.0001	0.0006	-0.0076	0.0873
10	0.0001	0.0004	-0.0252	0.0873
20	-0.0001	0.0002	-0.0582	0.0870
50	-0.00001	0.0001	-0.1153	0.1035
100	-0.0002	0.0001	-0.1882	0.1441

TABLE 13
Bias and mean square error of the m.l.e.'s $(\hat{\theta}, \hat{\alpha})$, for the intermediate binomial distribution $\{\hat{q}_i(\theta, \alpha)\}$, with an approximation of the normalizing constant, $\theta = 0.5$, $m = 20$, and $\alpha = 0.35$.

n	$b(\hat{\theta})$	$mse(\hat{\theta})$	$b(\hat{\alpha})$	$mse(\hat{\alpha})$
5	-0.0782	0.0070	-0.0893	0.0091
10	-0.0739	0.0064	-0.0904	0.0094
20	-0.0712	0.0060	-0.0903	0.0093
50	-0.0672	0.0055	-0.0900	0.0094
100	-0.0624	0.0049	-0.0893	0.0094

8. AN APPLICATION

We considered a data set in Kadane (2016), about the number of nice plants $\{x_i\} = \{0, 1, 2, 3, 4, 5, 6\}$, with the number of observed pots $\{0, 2, 2, 5, 5, 3, 3\}$.

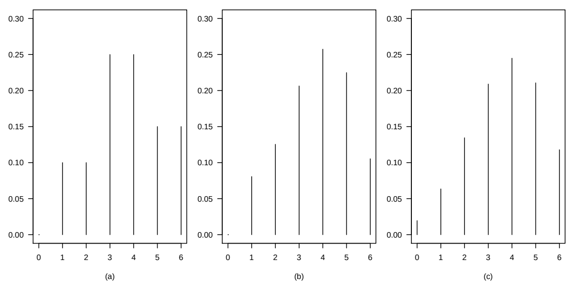

Figure 16 – Empirical distribution from the data set, in panel (a), fitted power binomial distribution $\{r_i(\hat{\theta}, \hat{\alpha})\}$, in panel (b), and fitted intermediate binomial distribution $\{q_i(\hat{\theta}, \hat{\alpha})\}$, in panel (c).

The dataset is interesting, because there was a situation of dependence for the Bernoulli r.v.'s, that ought to define a binomial r.v.. In particular, the use of a parameter α that determines the power and intermediate binomial distributions $\{r_i(\theta, \alpha)\}$ and $\{q_i(\theta, \alpha)\}$, given by (7) and (21), respectively, could be applied for an effective fitting.

In Figure 16, we consider the empirical distribution from the data set, and the fitted power binomial distribution $\{r_i(\hat{\theta}, \hat{\alpha})\}$, where the m.l.e.'s were $\hat{\theta} = 0.7870$ and $\hat{\alpha} = 0.4103$, and the fitted intermediate binomial distribution $\{q_i(\hat{\theta}, \hat{\alpha})\}$, where the m.l.e.'s were $\hat{\theta} = 0.6479$ and $\hat{\alpha} = 0.4913$. The m.l.e.'s $\hat{\theta}$ and $\hat{\alpha}$ were all obtained by the R "optim", with the algorithm of Nelder and Mead (1965). The values for $\hat{\alpha}$ were comparable, but they induced slightly different values for $\hat{\theta}$. In any case, the resulting shape of both fitted distributions show a good approximation to the observed distribution. We may observe that the fitted power binomial distribution $\{r_i(\hat{\theta}, \hat{\alpha})\}$ may be preferable, for small values in $\{x_i\}$, while the fitted intermediate binomial distribution

$\{q_i(\hat{\theta},\hat{\alpha})\}$ performs better, for large values in $\{x_i\}$. However, power and intermediate binomial distributions, at least for this example, show a similar performance.

ACKNOWLEDGEMENTS

The author thanks the Associate Editor and the Reviewer for their helpful and insightful comments.

APPENDIX

A. PROOF

For $\alpha > 1$, the Jensen's inequality determines

$$\begin{aligned}\frac{A(\theta)}{B_s}\mu_s &\le \left(\frac{A(\theta)}{B_s}\mu_s\right)^{\alpha}\\ &= A(\theta)^{\alpha}\left(\frac{\sum_i x_i^s p_i(\theta)}{B_s}\right)^{\alpha}\\ &\le \frac{A(\theta)^{\alpha}}{B_s}\sum_i x_i^s p_i(\theta)^{\alpha}\\ &= \frac{A(\theta)^{\alpha}}{B_s}2^{\alpha}\sum_i x_i^s\left(\frac{(1)\sum_{j=1}^{i} p_j(\theta) - (1)\sum_{j=1}^{i-1} p_j(\theta)}{2}\right)^{\alpha}\\ &\le \frac{A(\theta)^{\alpha}}{B_s}2^{\alpha}\frac{\sum_i x_i^s\left(\left(\sum_{j=1}^{i} p_j(\theta)\right)^{\alpha} - \left(\sum_{j=1}^{i-1} p_j(\theta)\right)^{\alpha}\right)}{2},\end{aligned}$$

and then (14).

For $0 < \alpha < 1$, the Jensen's inequality determines

$$\frac{A(\theta)}{B_s}\mu_s \ge \frac{A(\theta)^{\alpha}}{B_s}2^{\alpha}\frac{\sum_i x_i^s\left(\left(\sum_{j=1}^{i} p_j(\theta)\right)^{\alpha} - \left(\sum_{j=1}^{i-1} p_j(\theta)\right)^{\alpha}\right)}{2},$$

and then (15). □

B. Proof

For $\alpha > 1$, the Jensen's inequality determines

$$\begin{aligned}
\frac{A(x_i;\theta)}{B(x_i)}F(x_i;\theta) &\leq \left(\frac{A(x_i;\theta)}{B(x_i)}F(x_i;\theta)\right)^{\alpha} \\
&= A(x_i;\theta)^{\alpha}\left(\frac{\sum_{j=1}^{i} 1_{[x_j,\infty)}\Delta F(x_j;\theta)}{B(x_i)}\right)^{\alpha} \\
&\leq \frac{A(x_i;\theta)^{\alpha}}{B(x_i)} \sum_{j=1}^{i} 1_{[x_j,\infty)}(\Delta F(x_j;\theta))^{\alpha} \\
&= \frac{A(x_i;\theta)^{\alpha}}{B(x_i)} 2^{\alpha} \sum_{j=1}^{i} 1_{[x_j,\infty)}\left(\frac{(1)F(x_j;\theta)-(1)F_(x_j;\theta)}{2}\right)^{\alpha} \\
&\leq \frac{A(x_i;\theta)^{\alpha}}{B(x_i)} 2^{\alpha} \frac{\sum_{j=1}^{i} 1_{[x_j,\infty)}\Delta F(x_j;\theta,\alpha)}{2},
\end{aligned}$$

and then (17).

For $0 < \alpha < 1$, the Jensen's inequality determines

$$\frac{A(x_i;\theta)}{B(x_i)}F(x_i;\theta) \geq \frac{A(x_i;\theta)^{\alpha}}{B(x_i)} 2^{\alpha} \frac{\sum_{j=1}^{i} 1_{[x_j,\infty)}\Delta F(x_j;\theta,\alpha)}{2},$$

and then (19). □

References

N. Balakrishnan, V. B. Nevzorov (2003). *A Primer on Statistical Distributions*. John Wiley & Sons, Hoboken, New Jersey.

F. Belzunce, C. Martinez-Riquelme, J. Mulero (2016). *An Introduction to Stochastic Orders*. Academic Press, San Diego, California.

R. P. Brent (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, New Jersey.

S. Chakraborty, R. D. Gupta (2015). *Exponentiated geometric distribution: Another generalization of geometric distribution*. Communications in Statistics - Theory and Methods, 44, pp. 1143–1157.

F. Daly, R. E. Gaunt (2016). *The Conway-Maxwell-Poisson distribution: Distributional theory and approximation*. Latin American Journal of Probability and Mathematical Statistics, 13, pp. 635–658.

S. DHARMADHIKARI, K. JOAG-DEV (1988). *Unimodality, Convexity, and Applications*. Academic Press, San Diego, California.

S. R. DURRANS (1992). *Distributions of fractional order statistics in hydrology*. Water Resources Research, 28, pp. 1649–1655.

Y. M. GÓMEZ, H. BOLFARINE (2015). *Likelihood-based inference for the power half-normal distribution*. Journal of Statistical Theory and Applications, 14, pp. 383–398.

R. D. GUPTA, R. C. GUPTA (2008). *Analyzing skewed data by power normal model*. Test, 17, pp. 197–210.

G. HARDY, J. E. LITTLEWOOD, G. PÓLYA (1951). *Inequalities*. Second Edition, Cambridge University Press, Cambridge.

J. L. W. V. JENSEN (1906). *Sur les fonctions convexes et les inégalités entre les valeurs moyenne*. Acta Mathematica, 30, pp. 175–193.

N. L. JOHNSON, A. W. KEMP, S. KOTZ (2005). *Univariate Discrete Distributions*. Third Edition, John Wiley & Sons, Hoboken, New Jersey.

M. C. JONES (2004). *Families of distributions arising from distributions of order statistics (with discussion)*. Test, 13, pp. 1–43.

J. B. KADANE (2016). *Sums of possibly associated Bernoulli variables: The Conway-Maxwell-binomial distribution*. Bayesian Analysis, 1, pp. 403–420.

E. L. LEHMANN (1953). *The power of rank tests*. The Annals of Mathematical Statistics, 24, pp. 23–43.

E. L. LEHMANN, G. CASELLA (1998). *Theory of Point Estimation*. Second Edition, Springer, New York.

R. MIURA, H. TSUKAHARA (1993). *One-sample estimation for generalized Lehmann's alternative models*. Statistica Sinica, 3, pp. 83–101.

A. MÜLLER, D. STOYAN (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Chichester, England.

S. NADARAJAH, S. A. A. BAKAR (2016). *An exponentiated geometric distribution*. Applied Mathematical Modelling, 40, pp. 6775–6784.

S. NADARAJAH, S. KOTZ (2006). *The exponentiated type distributions*. Acta Applicandae Mathematicae, 92, pp. 97–111.

J. A. NELDER, R. MEAD (1965). *A simplex method for function minimization*. The Computer Journal, 7, pp. 308–313. Correction: 8, p. 27.

Y. PAWITAN (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Clarendon Press, Oxford.

A. PEWSEY, H. W. GÓMEZ, H. BOLFARINE (2012). *Likelihood-based inference for power distributions*. Test, 21, pp. 775–789.

N. S. PISKUNOV (1979). *Calcolo Differenziale e Integrale*. Volumi 1 e 2, Seconda Edizione, Editori Riuniti, Edizioni MIR, Roma, Mosca.

R CORE TEAM (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

S. M. ROSS (2013). *Simulation*. Fifth Edition, Academic Press, San Diego, California.

G. SHMUELI, T. P. MINKA, J. B. KADANE, S. BORLE, P. BOATWRIGHT (2005). *A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution*. Journal of the Royal Statistical Society. Series C, 54, pp. 127–142.

G. R. SHORACK (2000). *Probability for Statisticians*. Springer-Verlag, New York.

M. SPIVAK (1994). *Calculus*. Third Edition, Cambridge University Press, Cambridge.

A. WALD (1949). *Note on the consistency of the maximum likelihood estimates*. The Annals of Mathematical Statistics, 20, pp. 595–601.

SUMMARY

Discrete power distributions are proposed and studied, by considering the positive jumps on the discontinuities of an original discrete distribution function. Inequalities in moments and distribution functions are studied, allowing the definition of discrete intermediate distributions that lie between an original distribution and a power distribution. Original uniform, binomial, Poisson, negative binomial, and hypergeometric distributions are considered, to propose new power and intermediate distributions. Stochastic orders and unimodality are discussed. Estimation problems using likelihood are investigated. Simulation experiments are performed, to evaluate the bias and the mean square error of the maximum likelihood estimates, that are numerically calculated, with classic tools for numerical optimization.

Keywords: Asymptotics; Inequalities; Information; Intermediate distributions; Maximum likelihood estimation; Power distributions; Stochastic orders; Unimodality.