On a simple construction of bivariate probability functions with fixed marginals [1]

Djilali AIT AOUDIA[a], Éric MARCHAND[b,2]

a Université du Québec à Montréal, Département de mathématiques, 201, Ave Président-Kennedy Montréal, QC, CANADA, H2X 3Y7

b Université de Sherbrooke, Département de mathématiques, Sherbrooke, QC, CANADA, J1K 2R1

SUMMARY. We introduce a family of bivariate discrete distributions whose members are generated by a decreasing mass function p, and with margins given by p. Several properties and examples are obtained, including a family of seemingly novel "Bivariate Poisson" distributions.

AMS 2010 subject classifications: 60E05, 62E99

Keywords and phrases: Bivariate distributions, covariance, discrete distributions, mixtures of Poisson, Poisson, probability generating function.

1. Introduction

Univariate and multivariate count data abound in many disciplines and contexts (economic and actuarial science, disease prevalence, growth models, reliability, sports scores, etc). Strategies for constructing bivariate (or multivariate) discrete models, and an understanding of their properties, are thus important for stochastic modeling and statistical inference. Techniques that have given rise to such strategies include convolution, random summation, mixing, and copula based (e.g.,

[1] October 30, 2013

[2] Corresponding author: eric.marchand@usherbrooke.ca

Lai, 2004 and the references therein; Nelsen, 1987; Genest and Nešlehová, 2007; Piperigou, 2009; Panagiotelis, Czado and Joe, 2012). This note provides a simple and unified construction of a bivariate discrete distribution with fixed margins which have a non-increasing probability mass function on $\mathbb{N}$. New distributions, including a Poisson bivariate model for rare events, are obtained and several properties that arise are expressed in terms of the common marginal distribution. The applicability of the technique, as well as extensions, are discussed.

2. A class of bivariate discrete distributions

Consider p as a non-increasing mass function on $\mathbb{N}$. We begin with the following observation which forms the basis for the distributions which we will study.

Lemma 1. *The function $\gamma(x_1, x_2) = p(x_1 + x_2) - p(x_1 + x_2 + 1)$ is a probability mass function on $\mathbb{N}^2$ with marginals given by $p(\cdot)$.*

Proof. The non-increasing property of p implies that γ is nonnegative and we have the telescopic $\sum_{x_2} \gamma(x_1, x_2) = \sum_{x_2 \geq 0} \{p(x_1 + x_2) - p(x_1 + x_2 + 1)\} = p(x_1)$, which establishes the result. □

Remark 1. *Lemma 1, as well as the developments that follow in this paper, are presented under the assumption that the basis distribution p is non-increasing on $\mathbb{N}$. However, Lemma 1 can be extended to any mass function p^* on $\mathbb{N}$ with the help of a surrogate or a dual version p of p^*, which is constructed by ordering the values $p^*(i)$; $i \in \mathbb{N}$; from largest to smallest. A way to do this is as follows.*

(I) Start with p^* a probability mass function on $\mathbb{N} = \{0, 1, 2, \ldots\}$.

(II) Rearrange or permute the support $\mathbb{N}$ as $S = \{\pi(0), \pi(1), \pi(2), \ldots\}$ in such a way that $p(i) =$

$p^*(\pi(i))$ is non-increasing in i; $i \in \mathbb{N}$. As an illustration, we obtain $S = \{1, 2, 3, 0, 4, 5, \ldots\}$ for a Poisson p^* with mean parameter equal to 2. Note that S is not necessarily unique but it does not matter in such cases which choice is made.

(III) Apply Lemma 1 to p and obtain $\gamma(x_1, x_2) = p(x_1 + x_2) - p(x_1 + x_2 + 1) = p^*(\pi(x_1 + x_2)) - p^*(\pi(x_1 + x_2 + 1))$. This has margins given by p.

(IV) Define $\gamma^*(x_1, x_2) = \gamma(\pi^{-1}(x_1), \pi^{-1}(x_2))$. By construction, the bivariate pmf γ^* has margins given by p^*. Indeed, illustrating for X_1, we have $\sum_{x_2} \gamma^*(x_1, x_2) = \sum_{x_2} \gamma(\pi^{-1}(x_1), \pi^{-1}(x_2)) = \sum_{x_2} \gamma(\pi^{-1}(x_1), x_2) = p(\pi^{-1}(x_1)) = p^*(x_1)$.

Although Lemma 1's scheme is particularly simple and restricts $\gamma(x_1, x_2)$ to be constant on sets where $x_1 + x_2$ is constant, it generates several attractive properties and examples linking p and γ.[3] As a first example, we consider the Poisson case.

Example 1. *For $p \sim$ Poisson(α), with $\alpha \leq 1$ so that p be non-increasing on $\mathbb{N}$, we have*

$$\gamma(x_1, x_2) \;=\; \frac{e^{-\alpha}\alpha^{x_1+x_2}}{(x_1 + x_2 + 1)!}(x_1 + x_2 + 1 - \alpha)\,.$$

This represents a bivariate Poisson distribution at least in the sense that the marginals are Poisson distributed. Of course, many such distributions are known (e.g., Kocherlakota and Kocherlakota, 1992; Johnson, Kotz and Balakrishnan, 1997) but the above has a simple enough form so that it possibly has arisen in previous work, but we cannot identify such a source. The conditional distributions turn out to be given as

$$p(x_1|x_2) = \frac{\gamma(x_1, x_2)}{p(x_2)} = x_2!\,\frac{\alpha^{x_1}\;(x_1 + x_2 + 1 - \alpha)}{(x_1 + x_2 + 1\,)!}\;;$$

[3]A multivariate generalization, which is not pursued here, for a joint distribution for $(X_1, \ldots X_n)$ generated from a non-increasing p, with univariate marginals given by p, as above is of the form $\gamma(x_1, \ldots, x_n) = \sum_{j=0}^{n-1} \binom{n-1}{j} (-1)^j \, p(\sum_i x_i + j)\,.$

and these, with parameters x_2 and α, form a family of univariate probability functions on their own. Finally, distribution γ possesses at least two other interesting properties: **(i)** *$X_1 + X_2$ is distributed as the sum of a Bernoulli(α) and a Poisson(α) independently distributed random variables (see Theorem 1), and* **(ii)** *the Pearson correlation coefficient between X_1 and X_2 is equal to $-\alpha/2$ (see Example 3).*

Here is small catalog of further bivariate distributions generated by Lemma 1.

Distribution	$p(x)$	Parameters	$\gamma(x_1, x_2) = h(t)$
Poisson	$\frac{e^{-\alpha}\,\alpha^x}{x!}$	$\alpha \leq 1$	$\frac{e^{-\alpha}\,\alpha^t (t+1-\alpha)}{(t+1)\,!}$
Uniform	$(N+1)^{-1}\,\mathbb{I}_{\{0,1,\ldots,N\}}(x)$	$N \in \{1, 2, \ldots\}$	$(N+1)^{-1}\,\mathbb{I}_{\{N\}}(t)$
Geometric	$p\,(1-p)^x$	$0 < p \leq 1$	$p(x_1)\,p(x_2)$
Mixture of Geometric	$\int_{(0,1]} p(1-p)^x\,dH(p)$	cdf H on $(0,1]$	$\int_{(0,1]} p^2(1-p)^t\,dH(p)$
Beta-Geometric mixture	$\frac{\beta(a+1,b+x)}{\beta(a,b)}$	$a, b\ >\ 0$	$\frac{\beta(a+2,t+b)}{\beta(a,b)}$
Negative Binomial	$\binom{x+r-1}{r-1}\,p^r\,(1-p)^x$	$r(1-p) < 1, r \in \mathbb{N}$	$p(t)\,(\frac{(t+r)p+1-r}{t+1})$
Logarithmic	$\frac{-\theta^{x+1}}{(x+1)\,\log(1-\theta)}$	$(0 < \theta < 1)$	$p(t)\,(1-\theta+\frac{\theta}{t+2})$
Binomial	$\binom{n}{x}\,p^x\,(1-p)^{n-x}\,\mathbb{I}_{\{0,1,\ldots,n\}}(x)$	$(n+1)\,p \leq 1$	$p(t)\,(1-\frac{(n-t)\,p}{(1-p)\,(t+1)})$

In the above table, the name Beta-Geometric refers to a specific subclass of mixtures of Geometric distributions, where H is the cumulative distribution function of a Beta(a, b) random variable with density $\frac{1}{\beta(a,b)}\,p^{a-1}(1-p)^{b-1}$, β being the Beta function. Yule distributions arises here as Beta-Geometric mixtures with $b = 1$. Among these examples, we also highlight that independence arises for the Geometric family and this can be shown to characterize the Geometric as seen by showing that the functional equation $p(x_1 + x_2) - p(x_1)p(x_2) = p(x_1 + x_2 + 1)$, for all $x_1, x_2 \in \mathbb{N}$, implies that the sequence $p(x), x \in \mathbb{N}$, must necessarily be a geometric sequence. In contrast to this independence, we point out that the uniform example which represents a case of perfect dependence between X_1 and X_2 since $P(X_1 + X_2 = N) = 1$. We now turn to probability generating functions.

Lemma 2. *Suppose the probability generating function associated with mass function p exists and is given by* $\psi(s) = \sum_k s^k p(k)$. *Let* $(X_1, X_2) \sim \gamma$ *as above.*

(a) *The probability generating function of* $T = X_1 + X_2$ *is given by* $\psi_T(s) = (s-1)\psi'(s) + \psi(s)$*;*

(b) *The joint probability generating function of* (X_1, X_2) *is given by*

$$\psi_{X_1,X_2}(s_1, s_2) = E({s_1}^{X_1}\,{s_2}^{X_2}) \;=\; \frac{\psi(s_1)(s_1-1) \;-\; \psi(s_2)(s_2-1)}{s_1 - s_2}\,, \tag{1}$$

for $s_1 \neq s_2$*, and by* $\psi_T(s_1)$ *if* $s_1 = s_2$*.*

Proof. Part **(a)** follows from **(b)** by evaluating $\lim_{s_1 \to s} \psi_{X_1,X_2}(s_1, s)$ with l'Hopital's rule [4]. In part **(b)**, we have for $s_1 \neq s_2$

$$\begin{aligned}
E({s_1}^{X_1}\,{s_2}^{X_2}) &= \sum_{x_1,x_2} p(x_1+x_2){s_1}^{x_1}\,{s_2}^{x_2} \;-\; \sum_{x_1,x_2} p(x_1+x_2+1){s_1}^{x_1}\,{s_2}^{x_2} \\
&= \sum_{t\geq 0}\sum_{x_1=0}^{t} p(t)\,(\frac{s_1}{s_2})^{x_1}\,{s_2}^{t} \;-\; \sum_{t\geq 0}\sum_{x_1=0}^{t} p(t+1)\,(\frac{s_1}{s_2})^{x_1}\,{s_2}^{t} \\
&= \frac{1}{1-\frac{s_1}{s_2}}\left(\sum_{t\geq 0} p(t)\,{s_2}^{t}(1-(\frac{s_1}{s_2})^{t+1}) - \sum_{t\geq 0} p(t+1)\,{s_2}^{t}(1-(\frac{s_1}{s_2})^{t+1})\right) \\
&= \frac{1}{1-\frac{s_1}{s_2}}\left(\psi(s_2) - \frac{s_1}{s_2}\,\psi(s_1) - (\frac{1}{s_2}(\psi(s_2)-p(0)) - \frac{1}{s_2}(\psi(s_1)-p(0))\right),
\end{aligned}$$

[4] or by a direct evaluation

$$\begin{aligned}
\psi_T(s) &= \sum_{x_1,x_2}\{p(x_1+x_2) - p(x_1+x_2+1)\}\,s^{x_1+x_2} \\
&= \sum_{t\geq 0}(t+1)p(t)s^t - \sum_{t\geq 0}(t+1)p(t+1)s^t \\
&= \frac{d}{ds}(s\,\psi(s)) - \psi'(s) \\
&= (s-1)\psi'(s) + \psi(s)\,.
\end{aligned}$$

which yields the result. $\square$

Example 2. *(Example 1 continued) For the Poisson case with $\alpha \leq 1$, the joint probability generating function associated with γ given in Example 1 simplifies to*

$$\psi_{X_1,X_2}(s_1, s_2) \;=\; \sum_{k,l\geq 0} \alpha^{k+l}\,\frac{(s_1-1)^k\,(s_2-1)^l}{(k+l\,)!}\,.$$

This follows from part **(b)** *of Lemma 2 by substituting $\psi(s_i) = e^{\alpha(s_i-1)}$ and verifying that $(s_1-1)\,e^{\alpha(s_1-1)} - (s_2-1)\,e^{\alpha(s_2-1)} \;=\; ((s_1-1)-(s_2-1))\,\sum_{k,l\geq 0}\alpha^{k+l}\,\frac{(s_1-1)^k\,(s_2-1)^l}{(k+l\,)!}$. Further more general relationships concerning ψ_{X_1,X_2} and binomial moments are presented in the Appendix.*

In the Poisson case, a simple representation for the distribution of the sum T arises.

Theorem 1. *For $(X_1, X_2) \sim \gamma$ with $p \sim$ Poisson(α), $\alpha \leq 1$, the distribution of $T = X_1 + X_2$ matches the distribution of $Y_1 + Y_2$, where Y_1 and Y_2 are independently distributed as Poisson(α) and Bernoulli(α) respectively.*

Proof. Apply part **(a)** of Lemma 2 with $\psi(s) = e^{\alpha(s-1)}$ yielding indeed $\psi_T(s) = (\alpha s + 1 - \alpha)\,e^{\alpha(s-1)} = E(s^{Y_1})\,E(s^{Y_2})$. $\square$

The following is an alternate representation in terms of the Binomial moments $E\binom{Z}{r}$, defined for a random variable Z on $\mathbb{N}$, $r \in \mathbb{N}$, and with $\binom{z}{r} = 0$ for $r > z$.

Lemma 3. *Under the conditions of Lemma 2, we have $E\binom{T}{k} = (k+1)\,E\binom{X_1}{k}$.*

Proof. Denoting $\psi_T^{(k)}$ and $\psi^{(k)}$ as the k-th derivatives of ψ_T and ψ, it is readily verified from part **(a)** Lemma 2 that

$$\psi_T^{(k)}(s) = (s-1)\,\psi^{(k+1)}(s) \;+\; (k+1)\,\psi^{(k)}(s)\,.$$

With such derivatives evaluated at $s = 1$ yielding the factorial moments, we have indeed

$$k!\,E\binom{T}{k} = \psi_T^{(k)}(1) = (k+1)\,\psi^{(k)}(1) = (k+1)!\,E\binom{X_1}{k}. \quad \square$$

Corollary 1. *Under the conditions of Lemma 2, we have* $Var(T) = 3Var(X_1) - E(X_1) - (E(X_1))^2$ *and* $Cov(X_1, X_2) = \frac{1}{2}(Var(X_1) - E(X_1) - (E(X_1))^2)$.

Proof. It suffices to apply Lemma 3 for $k = 1, 2$ to obtain $E(T^2) = 3E(X_1^2) - E(X_1)$, as well as the above expressions for $Var(T)$ and $Cox(X_1, X_2) = \frac{Var(T)}{2} - Var(X_1)$. $\square$

We conclude with further illustrations for Poisson mixtures which are a natural model for Poisson counts in presence of an heterogeneous environment (e.g., Johnson, Kemp and Kotz, 2005, section 8.3; Karlis and Xekalaki, 2005). As seen as well in the following example, the covariance between X_1 and X_2, whenever it exists, can be negative, positive or equal to 0.

Example 3. *Consider Poisson mixtures*

$$p(x) = \frac{1}{x!} \int_0^1 e^{-\alpha} \, \alpha^x \, dG(\alpha)$$

with the mixing distribution G *supported on* $[0, 1]$ *guaranteeing*[5] *that* $p(\cdot)$ *decrease on* $\mathbb{N}$, *and with the developments above applying. Several further Poisson mixture representations arise, such as for the joint probability function*

$$\gamma(x_1, x_2) = \int_0^1 \frac{e^{-\alpha}\alpha^{x_1+x_2}}{(x_1 + x_2 + 1)!}(x_1 + x_2 + 1 - \alpha) \, dG(\alpha) \, ,$$

as in Example 1. With $E_p(X_1) = E_G(\alpha)$ *and* $Var_p(X_1) = Var_G(\alpha) + E_G(\alpha)$, *the covariance given in Corollary 1, as well as the Pearson correlation coefficient, specialize to*

$$\begin{aligned} Cov_G(X_1, X_2) &= \frac{1}{2}\left(Var_G(\alpha) - (E_G(\alpha))^2\right) \, , \\ \text{and } \rho_G(X_1, X_2) &= \frac{1}{2} \frac{Var_G(\alpha) - (E_G(\alpha))^2}{Var_G(\alpha) + E_G(\alpha)} \, . \end{aligned}$$

[5] This is not a necessary condition. For instance, the Negative Binomial distributions presented earlier in the Table are Poisson mixtures with Gamma distributed α.

Observe that X_1 and X_2 are positively correlated if and only if the coefficient of variation of the mixing parameter $\alpha \sim G$ is greater than one. We also point out, as further observations or examples, that:

(i) For the degenerate case, i.e., $p \sim Poisson(\alpha)$, we obtain immediately $Cov(X_1, X_2) = -\frac{\alpha^2}{2}$ and $\rho(X_1, X_2) = -\frac{\alpha}{2}$ (as mentioned in Example 1);

(ii) For $\alpha \sim Beta(a, b)$, we obtain $Cov(X_1, X_2) = \frac{a^2}{2(a+b)^2}\left(\frac{b}{a(a+b+1)} - 1\right)$ and $\rho(X_1, X_2) = \frac{1}{2}\frac{b-a(a+b+1)}{(a+b)(a+b+1)+b}$, with for instance $Cov(X_1, X_2) = -\frac{1}{12}$ and $\rho(X_1, X_2) = -\frac{1}{7}$ for $\alpha \sim U(0, 1)$.

(iii) It is apparent from the above expression for ρ_G that $\rho_G(X_1, X_2) \geq -1/2$ with equality iff α is degenerate at 1, and we can also show that $\sup_G \rho_G(X_1, X_2) = 1/4$, where the sup is taken over all distributions G supported on $(0, 1]$. This can be seen as follows. For fixed $r = E_G(\alpha)$, we have that[6] $\sup_G Var_G(\alpha) = r(1-r)$ and ρ_G increases in $Var_G(\alpha)$, from which we infer that

$$\sup_G \rho_G(X_1, X_2) = \sup_{r\in(0,1]} \frac{1}{2}\frac{r(1-r) - r^2}{r(1-r) + r} = \frac{1}{4}\,.$$

Concluding Remarks

We have introduced a simple technique for generating a bivariate discrete mass function γ from an univariate non-increasing mass function p on $\mathbb{N}$. We have derived and illustrated various properties of γ which are conveniently expressed in a unified way in terms of p. Among the examples presented, the resulting Poisson bivariate distribution is particularly simple and seemingly novel. Despite the fact that the technique of construction is amenable to both general mass functions on $\mathbb{N}$ (Remark 1) and to multivariate settings, limitations in terms of flexibility are present (i.e., $\gamma(x_1, x_2)$ depends on $x_1 + x_2$ only, many values of Pearson correlation coefficient cannot be attained). Nevertheless,

[6]By taking a sequence of distributions with mean r converging in distribution to a Bernoulli(r).

the paper contributes to the understanding of bivariate discrete models and may well serve as a stepping stone to generate a wider, more flexible class of models.

Acknowledgments

The research work of Éric Marchand is partially supported by NSERC of Canada.

References

[1] Genest, C. and Nešlehová, J. (2007). A Primer on Copulas for Count Data. *The Astin Bulletin*, **37**, 475-515.

[2] Johnson, N.L., Kotz, S., and Balakrishnan, N. (2005). *Discrete multivariate distributions*, John Wiley & Sons.

[3] Johnson, N.L., Kemp, A.W., and Kotz, S. (2005). *Univariate discrete distributions*, John Wiley & Sons, third edition.

[4] Karlis, D. and Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review*, **73**, 35-58.

[5] Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate discrete distributions*, Marcel Dekker.

[6] Lai, C.D. (2006). Construction of discrete bivariate distributions. In *Advances in distribution theory, order statistics, and inference*, Birkhäuser, Boston, pp. 29-58.

[7] Nelsen, R.B. (1999). Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics: Simulation and Computation*, **16**, 199-208.

[8] Panagiotelis, A., Czado, C., and Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, **107**, 1063-1072.

[9] Piperigou, V.E. (2009). Discrete distributions in the extended FGM family: The p.g.f. approach. *Journal of Statistical Planning and Inference*, **139**, 3891-3899.

Appendix : Binomial moment representations

Complementing the Poisson example (Example 2) and part **(b)** of Lemma 2, we expand here on an alternate representation of the joint probability generating function ψ_{X_1,X_2}. We make use of the mixed Binomial moments $E\binom{X_1}{k}\binom{X_2}{l}$, $k,l \geq 0$, with $\binom{z}{r}$ taken to be equal to 0 for $r > z$. Binomial expansions tell us that

$$\begin{aligned}\psi_{X_1,X_2}(s_1,s_2) &= E\left((s_1-1+1)^{X_1}(s_2-1+1)^{X_2}\right) \\ &= \sum_{k,l\geq 0} E\binom{X_1}{k}\binom{X_2}{l}(s_1-1)^k(s_2-1)^l\,, \end{aligned} \tag{2}$$

for all s_1, s_2 for which $\psi_{X_1,X_2}(s_1,s_2)$ exists. Now, as an illustration of a general phenomenon which is the objective of this remark, observe in the Poisson case (Example 2) that $\frac{\alpha^{k+l}}{(k+l)!} = E\binom{X_1}{k}\binom{X_2}{l} = E\binom{X_1}{k+l}$ by a direct evaluation of $E\binom{X_1}{k+l}$ in the Poisson case and given the uniqueness of the coefficients in power series representation (2).

Theorem 2. *Under the conditions of Lemma 2, the probability generating function of* $(X_1, X_2) \sim \gamma$ *is given by*

$$\psi_{X_2,X_2}(s_1,s_2) = \sum_{k,l\geq 0} E\binom{X_1}{k+l}(s_1-1)^k(s_2-1)^l\,. \tag{3}$$

Equivalently, we have

$$E\binom{X_1}{k}\binom{X_2}{l} = E\binom{X_1}{k+l}, \quad \textit{for all } k,l \geq 0\,. \tag{4}$$

Proof. Begin by observing that ψ satisfies the equation

$$\psi_{X_1,X_2}(s_1,s_2)(s_1-s_2) \;=\; (s_1-1)\,\psi_{X_1,X_2}(s_1,1) \;-\; (s_2-1)\,\psi_{X_1,X_2}(1,s_2)\,, \tag{5}$$

for all (s_1,s_2) for which $\psi(s_1,s_2)$ is defined. With the series representation (2) and the uniqueness of the coefficients, (3) will follow once we establish (4). Now, equation (5) implies, for all (s_1,s_2), with $c_{k,l} \;=\; E\binom{X_1}{k}\binom{X_2}{l}$,

$$\begin{aligned}
\sum_{k,l\geq 0} c_{k,l}\,(s_1-1)^{k+1}\,(s_2-1)^{l} \;-\; \sum_{k,l\geq 0} c_{k,l}\,(s_1-1)^{k}\,(s_2-1)^{l+1} \;&=\; \sum_{k\geq 0} c_{k,0}(s_1-1)^{k+1} \;-\; \sum_{k\geq 0} c_{0,l}(s_2-1)^{l+1} \\
\Longrightarrow \sum_{k,l\geq 1} c_{k-1,l}\,(s_1-1)^{k}\,(s_2-1)^{l} \;&=\; \sum_{k,l\geq 1} c_{k,l-1}\,(s_1-1)^{k}\,(s_2-1)^{l}\,.
\end{aligned}$$

But the above is equivalent to $c_{k-1,l}=c_{k,l-1}$ for all $k,l\geq 1$, which implies $c_{0,k+l-1}=\cdots=c_{k-1,l}=c_{k,l-1}=\cdots=c_{k+l-1,0}$ for all $k,l\geq 1$, which is (4). $\square$