Comparison between two bivariate Poisson distributions through the phi-divergence.

Prévot C. Batsindila Nganga¹, Rufin Bidounga¹, Dominique Mizère¹,* and Célestin C. Kokonendji²,*

¹Marien Ngouabi University, BP 69, Brazzaville, Congo.
²Laboratoire de Mathématiques de Besançon, University Bourgogne Franche-Comté, Besançon, France.

Received October 29, 2017. Accepted November 25, 2017

Copyright © 2017, Afrika Statistika and The Statistics and Probability African Society (SPAS). All rights reserved

Abstract. In the literature, there are two probabilistic models of bivariate Poisson: the model according to Holgate and the model according to Berkhout and Plug. These two models express themselves by their probability mass function. The model of Holgate puts in evidence a strictly positive correlation, which is not always realistic. To remedy this problem, Berkhout and Plug proposed a bivariate Poisson distribution accepting the correlation as well negative, equal to zero, that positive. In this paper, we show that these models are nearly everywhere asymptotically equal. From this survey that the φ-divergence converges toward zero, both models are therefore nearly everywhere equal. Also, the model of Holgate converges toward the one of Berkhout and Plug. Some graphs will be presented for illustrating this comparison.

Key words: Correlation structure; distribution convergence.
AMS 2010 Mathematics Subject Classification : 62G20; G2H10; 40A05.
Résumé. Dans la littérature, il y a deux modèles probabilistes de Poisson bivariés: le modèle selon Holgate puis le modèle selon Berkhout et Plug. Ces deux modèles sont définis par leur fonction de masse. Le modèle de Holgate met en évidence une corrélation strictement positive, ce qui n’est pas toujours réaliste. Pour remédier à ce problème, Berkhout et Plug ont proposé une distribution de Poisson bivariée acceptant la corrélation aussi bien négative, nulle que positive. Dans ce papier, nous montrons que ces modèles sont presque partout asymptotiquement égaux car la φ−divergence de ces deux modèles converge vers zéro. Aussi, le modèle de Holgate converge vers celui de Berkhout et Plug. Les graphes seront présentés pour illustrer cette comparaison.

1. Introduction

The bivariate Poisson distribution has been introduced by Campbell (1934) who considered the limit of the distribution of a contingency table with two dimensions. Practically at the same time, Guldberg (1934) obtains the independent binomial distributions. The explicit form of the distribution is due a few years later in 1944 (Morin (1993)) to Aitken (1994). It is however necessary to await Holgate (1934) in 1964 to obtain a bivariate variables starting from three univariate variables of Poisson independent, that is with a matrix of variance-covariance not diagonal. Contrary to what is said in the paper of Elion et al. (2016), one cannot assign this find to Johnson and Kotz (1969) whose research are belated. Besides, the literature informs that several bivariate Poisson distributions have been put in evidence. The applied statistics makes use of the bivariate Poisson distribution more according to Holgate (1934) and a lot of scientist works (Kawamura (1973), Kocherlakota et al. (1992)) rest on this distribution. It permits to have a log-likelihood function with a difficult expression to manipulate of which the maximum of likelihood is estimated by means of the Hessian matrix. But, outside of the difficult matrix manipulations that it requires to estimate the parameters, the model of Holgate puts in evidence a strictly positive correlation. To remedy this problem, Berkhout et al. (2004) proposed a bivariate Poisson distribution accepting the correlation as well negative, equal to zero, that positive. Otherwise, in their paper, Berkhout and Plug reviewed the model of Holgate without comparing it to their model. (Elion et al. (2016)) passed in magazine the two models without saying a word on the relation that exists between the two. We wondered if these two models are identical or not. The model of Holgate (1934) and the one of Berkhout and Plug express themselves by their probability mass function (pmf).

Cuenin et al. (2016) proposed also a model of bivariate Poisson based on the distribution of Tweedie of which one knows that the pmf is not easy to calculate. In the literature, we raised an instrument permitting to compare two densities of probabilities to know: the φ−divergence (Toma (2009)).

We show, in this work, that the φ−divergence from model of Berkhout et al. (2004) to model of Holgate (1934) converge toward zero: therefore the model of Holgate is nearly everywhere asymptotically equal to the the model of Berkhout and Plug. We also shows that the model of Berkhout et al. (2004) is the asymptotic distribution of the model of Holgate (1934). Some graphs will be presented at the end of this paper to illustrate the comparison of these two models.
2. Generality

2.1. The bivariate Poisson distribution according to Holgate.

Let us consider three independent variables $V_1$, $V_2$ and $U$ which follow the univariate Poisson distribution of respective parameters $\lambda_1$, $\lambda_2$ and $\lambda_3$. With these three variables, one builds two new variables dependent $Y_1$ and $Y_2$ such as:

$$Y_j = V_j + U, \quad j = 1, 2.$$  (1)

Then the pmf of the bivariate Poisson distribution is written:

$$P(Y_1 = y_1, Y_2 = y_2) = e^{-\lambda_1 - \lambda_2 - \lambda_3} \sum_{l=0}^{\min(y_1, y_2)} \frac{\lambda_3^l}{l!} \frac{\lambda_1^{y_1-l}}{(y_1-l)!} \frac{\lambda_2^{y_2-l}}{(y_2-l)!}.$$  

The distribution of bivariate Poisson according to Holgate will be noted by $f_H(y_1, y_2; \lambda_1, \lambda_2, \lambda_3)$.

We have the following the following class of properties.

Class of Property (I).

(a) The covariance between $Y_1$ and $Y_2$ gives:

$$\text{cov}(Y_1, Y_2) = \text{cov}(V_1 + U, V_2 + U) = \lambda_3.$$  

(b) The correlation between $Y_1$ and $Y_2$ is equal to:

$$\text{cor}(Y_1, Y_2) = \frac{\lambda_3}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}}.$$  

(c) The marginal distributions follow:

$$P(Y_1 = y_1) = e^{-\lambda_1 - \lambda_3} \frac{(\lambda_1 + \lambda_3)^{y_1}}{y_1!} \quad \text{and} \quad P(Y_2 = y_2) = e^{-\lambda_2 - \lambda_3} \frac{(\lambda_2 + \lambda_3)^{y_2}}{y_2!}.$$  

(d) The conditional distribution is equal to:

$$P(Y_1 = y_1|Y_2 = y_2) = P(Y_1 = y_1, Y_2 = y_2) \times \left( \frac{(\lambda_2 + \lambda_3)^{y_2}}{y_2!} e^{-\lambda_2 - \lambda_3} \right)^{-1}$$

$$= e^{-\lambda_1} \sum_{l=0}^{\min(y_1, y_2)} \binom{y_1}{l} \left( \frac{\lambda_3}{\lambda_3 + \lambda_2} \right)^l \left( \frac{\lambda_2}{\lambda_3 + \lambda_2} \right)^{y_2-l} \frac{\lambda_3^{y_1-l}}{(y_1-l)!}.$$  

and has characteristics:

\[ E(Y_1|Y_2) = \lambda_1 + y_2 \cdot \frac{\lambda_3}{\lambda_3 + \lambda_2} \text{var}(Y_1|Y_2) \]
\[ = \lambda_1 + \frac{y_2\lambda_2\lambda_3}{(\lambda_3 + \lambda_2)^2}. \]

To finish, let us retain that the conditional law of \( Y_1 \) knowing \( Y_2 \) is the convolution of the variable of Poisson \( U \) of parameter \( \lambda_1 \) and of the variable binomial \( V \) of parameters \( y_2 \) and \( \lambda_3/ (\lambda_3 + \lambda_2) \).

### 2.2. The bivariate Poisson distribution according to Berkhout and Plug.

Let us consider a couple of dependent positive integer whole random variables \((Y_1, Y_2)\). According to theory of conditional probabilities, their joined density can be written like the product of a distribution marginal and of a conditional distribution, such as:

\[
P(Y_1 = y_1, Y_2 = y_2) = f(y_1, y_2) = g_1(y_1) \cdot g_2(y_2|y_1) \]
\[
P(Y_1 = y_1, Y_2 = y_2) = f(y_1, y_2) = g_1^\pi(y_1) \cdot g_2^\pi(y_1|y_2) \]

This decomposition of the joined density can take several forms compared to permutations of \( Y_1 \) and \( Y_2 \). In this bivariate case, we have two permutations:

That is to say \( \pi \) an indicator of permutation taking value 0 and 1; \( z_1 \) and \( z_2 \) can be written like permutations realized by \( y_1 \) and \( y_2 \), such as:

\[ z_1 = (1 - \pi)y_1 + \pi y_2 \quad \text{and} \quad z_2 = \pi y_1 + (1 - \pi)y_2 \]

ou \( \pi \in \{0, 1\} \).

Let us suppose \( g_1^\pi \) and \( g_2^\pi \) two Poisson distribution, for any joined distribution \( f^\pi \). For the choice of the permutation, the existing data enable us to fix the density \( f^\pi(z_1, z_2) \). Then the joined density is written:

\[
f(y_1, y_2) = f^\pi(z_1, z_2) = g_1^\pi(z_1) \cdot g_2^\pi(z_2|z_1) \quad \text{for} \quad \pi \in \{0, 1\}. \]

When the marginal one of \( z_1 \) is equal to:

\[
g_1^\pi(z_1) = \frac{\mu_1^{z_1}}{z_1!} e^{-\mu_1}, \tag{2} \]

and the conditional distribution \( z_2|z_1 \) which:

\[ g_2^\pi (z_2|z_1) = \frac{\mu_2^z}{z_2!} e^{-\mu_2} \]  

(3)

where \( \mu_1 = \exp(x'\beta_1) \) and \( \mu_2 = \exp(x'\beta_2 + \eta z_1) \), with \( \beta_1, \beta_2 \) and \( \eta \) are parameters of \( g_1^\pi \) and \( g_2^\pi \) and \( x' = (x_1, x_2, ..., x_p) \) the vector of the explanatory variables or factors. Then joined density of the permutations \( z^{\pi} \):

\[ f^{\pi} (z_1, z_2) = \frac{e^{z_1(x'\beta_1) - \exp(x'\beta_1) + z_2(x'\beta_2 + \eta z_1) - \exp(x'\beta_2 + \eta z_1)}}{z_1! z_2!} \]

is called pmf of the bivariate Poisson distribution according to Berkhout and Plug, and will be denoted \( f_{BP} (z_1, z_2; \mu_1, \mu_2) \).

Likely, it is difficult to handle such an expression is a direct way, but that is possible to see that by using the factorial moment \( (r, s)^{th} \) of the joint distribution, we get this second class of properties.

**Class of Property (II).**

(a) The factorial moment \( (r, s) \) of the joined distribution can be written:

\[ \mathbb{E} [Z_1 (Z_1 - 1) \cdots (Z_1 - r + 1) Z_2 (Z_2 - 1) \cdots (Z_2 - s + 1)] = \sum_{z_1=r}^{+\infty} \sum_{z_2=s}^{+\infty} e^{z_1(x'\beta_1) - \exp(x'\beta_1) + z_2(x'\beta_2 + \eta z_1) - \exp(x'\beta_2 + \eta z_1)} \]

\[ = e^{-\exp(x'\beta_1) + \exp(x'\beta_1 + \eta)) + r x' \beta_1 + s x' \beta_2 + r s \eta} \]

\[ = \mu_1 e^{h_1(\exp(\eta)) - 1} + s x' \beta_2 + r s \eta} \]

(4)

(b) From the expression (5), we can now calculate every moment of the distribution joined just by the choice of \( r \) and \( s \), one has what follows:

(b1) For \( s = 1 \) and \( r = 0 \),

\[ \mathbb{E} (Z_2) = \exp [x' \beta_2 + \mu_1 (\exp (\eta) - 1)] . \]

(b2) For \( s = 2 \) and \( r = 0 \), one has for the variable \( Z_2 \):

\[ \mathbb{E} (Z_2^2) = \mathbb{E} (Z_2) + e^{2x' \beta_2 + \mu_1 (\exp(2 \eta) - 1)} , \]

\[ \text{var} (Z_2) = \mathbb{E} (Z_2) + (\mathbb{E} (Z_2))^2 \left( e^{\mu_1 (\exp(\eta) - 1)^2} - 1 \right) . \]

We note that for all \( \eta \neq 0 \), \( \text{var} (Z_2) - \mathbb{E} (Z_2) > 0 \), which implies that \( Z_2 \) is overdispersed (Castillo et al. (1998)).
(c) Let us pose $s = 0$ and $r = 1$, $r = 2$, $r = 0$ successively, we have for the $Z_1$ variable:

$$E(Z_1) = \mu_1, \quad \text{var}(Z_1) = \mu_1, \quad E(Z_1Z_2) = \mu_1 e^{\mu_1(e^\eta - 1) + x' \beta_2 + \eta}.$$  

We have

$$E(Z_1) \times E(Z_2) = \mu_1 e^{\mu_1(e^\eta - 1) + x' \beta_2},$$

and then

$$\text{cov}(Z_1, Z_2) = E(Z_1Z_2) - E(Z_1)E(Z_2) = \mu_1 E(Z_2) (\exp(\eta) - 1).$$

(d) Thus the correlation, between $Z_1$ and $Z_2$, can be obtained as follows

$$\text{cor}(Z_1, Z_2) = \frac{\mu_1 \{E(Z_2) (\exp(\eta) - 1)\}}{\sqrt{\mu_1 \{E(Z_2) + E^2(Z_2)\} (\exp(\mu_1(\exp(\eta) - 1)) - 1)}}.$$  

We note that the correlation is positive (resp. negative) if $\eta > 0$ (resp. $\eta < 0$). However for $\eta = 0$, we have

$$E(Z_2) = \text{var}(Z_2) = \mu_2$$

and $\text{cov}(Z_1, Z_2) = 0$. Then, to respect with the interdependence of the variables $Z_1$ and $Z_2$, it is necessary that $\eta \neq 0$. $\diamond$

\textbf{Remark 1.} We will suppose in all that follows that the permutation is equal to $\pi = 0$, which provides that $z_1 = y_1$ and $z_2 = y_2$.

3. Comparison of the densities of probabilities $f_H$ and $f_{BP}$.

We are going to show here that the densities $f_H$ and $f_{BP}$ are nearly everywhere equal asymptotically.

3.1. functional relation between $f_H$ and $f_{BP}$.

\textbf{Proposition 1.} While taking $\mu_1 = \lambda_1 + \lambda_3$ and $\mu_2 = \lambda_2 + \lambda_3$ the pmf $f_H$ decompose himself in a product of factors:

$$f_H(y_1, y_2; \mu_1, \mu_2, \lambda_3) = \binom{\mu_1}{y_1} e^{-\mu_1} \binom{\mu_2}{y_2} e^{-\mu_2} \times b(y_1, y_2; \mu_1, \mu_2, \lambda_3)$$

with

$$b(y_1, y_2; \mu_1, \mu_2, \lambda_3) = e^{\lambda_3} \left(1 - \frac{\lambda_3}{\mu_1}\right)^{y_1} \left(1 - \frac{\lambda_3}{\mu_2}\right)^{y_2} \sum_{l=0}^{\min(y_1, y_2)} (-y_1)_{l} (-y_2)_{l} \frac{x^l}{l!}$$

and

$$z = \frac{\lambda_3}{(\mu_1 - \lambda_3)(\mu_2 - \lambda_3)}.$$


Proof of Proposition 1.
Let us put \( \mu_1 = \lambda_1 + \lambda_3, \mu_2 = \lambda_2 + \lambda_3 \), we have:

\[
\begin{align*}
    f_H(y_1, y_2; \mu_1, \mu_2, \lambda_3) &= e^{-\mu_1-\mu_2+\lambda_3} \sum_{l=0}^{\min(y_1,y_2)} \frac{(\mu_1 - \lambda_3)^{y_1-l} (\mu_2 - \lambda_3)^{y_2-l} \lambda_3^l}{(y_1-l)! (y_2-l)! l!} \\
    &= e^{-\mu_1-\mu_2} \sum_{l=0}^{\min(y_1,y_2)} \frac{(\mu_1 - \lambda_3)^{y_1} (\mu_2 - \lambda_3)^{y_2} \lambda_3^l}{(\mu_1 - \lambda_3)^l (\mu_2 - \lambda_3)^l l!} \frac{1}{(y_1-l)! (y_2-l)!} \\
    &= \frac{\mu_1^{y_1}}{(y_1)!} \frac{\mu_2^{y_2}}{(y_2)!} e^{-\mu_1-\mu_2} \lambda_3 \sum_{l=0}^{\min(y_1,y_2)} \frac{(\mu_1 - \lambda_3)^{y_1} (\mu_1 - \lambda_3)^{y_2} \lambda_3^l}{(y_1-l)! (y_2-l)! l!}
\end{align*}
\]

with

\[ z = \frac{\lambda_3}{(\mu_1 - \lambda_3)(\mu_2 - \lambda_3)}. \]

We have:

\[
\begin{align*}
    f_H(y_1, y_2; \mu_1, \mu_2, \lambda_3) &= \frac{\mu_1^{y_1}}{(y_1)!} \frac{\mu_2^{y_2}}{(y_2)!} e^{-\mu_1-\mu_2} \lambda_3 \sum_{l=0}^{\min(y_1,y_2)} \left(1 - \frac{\lambda_3}{\mu_1}\right)^{y_1} \left(1 - \frac{\lambda_3}{\mu_2}\right)^{y_2} \frac{(-1)^l y_1 (-1)^l y_2 z^l}{(y_1-l)! (y_2-l)! l!} \\
    f_H(y_1, y_2; \mu_1, \mu_2, \lambda_3) &= \left(e^{-\mu_1} \frac{\mu_1^{y_1}}{y_1!}\right) \left(e^{-\mu_2} \frac{\mu_2^{y_2}}{y_2!}\right) \left(1 - \frac{\lambda_3}{\mu_1}\right)^{y_1} \left(1 - \frac{\lambda_3}{\mu_2}\right)^{y_2} e^{\lambda_3} \sum_{l=0}^{\min(y_1,y_2)} (-y_1)_l (-y_2)_l \frac{z^l}{l!}
\end{align*}
\]

While putting

\[ b(y_1, y_2; \mu_1, \mu_2, \lambda_3) = \left(1 - \frac{\lambda_3}{\mu_1}\right)^{y_1} \left(1 - \frac{\lambda_3}{\mu_2}\right)^{y_2} e^{\lambda_3} \sum_{l=0}^{\min(y_1,y_2)} (-y_1)_l (-y_2)_l \frac{z^l}{l!} \]

with \((-y_1)_l\) represent Pochhammer’s symbol (Johnson et al. (1993)). The proof is finished. \( \square \)

The following results are easily derived for previous ones.

Corollary 1. In the expression (6), while taking (see expressions (2) and (3)):

\[ \frac{\mu_1^{y_1}}{(y_1)!} e^{-\mu_1} = \mathbb{P}[Y_1 = y_1], \]

the density marginal of \( Y_1 \) with \( \ln \mu_1 = x' \beta_1 \) and

\[ \frac{\mu_2^{y_2}}{(y_2)!} e^{-\mu_2} = \mathbb{P}[Y_2 = y_2/Y_1 = y_1], \]

the conditional distribution of \( Y_2 \) when one considers \( Y_1 = y_1 \), with \( \ln \mu_2 = x' \beta_2 + \eta y_1 \).
Indeed, we have
\[ P[Y_1 = y_1, Y_2 = y_2] = P[Y_1 = y_1] P[Y_2 = y_2 \mid Y_1 = y_1] = f_{BP}(y_1, y_2; \mu_1, \mu_2), \]
and therefore:
\[ f_H(y_1, y_2; \mu_1, \mu_2, \lambda_3) = f_{BP}(y_1, y_2; \mu_1, \mu_2) \times b(y_1, y_2; \mu_1, \mu_2, \lambda_3). \] (9)

**Proposition 2.** Let us consider two generalized linear models \( \ln \mu_1 = x' \beta_1 \) and \( \ln \mu_2 = x' \beta_2 + \eta y_1 \). The response variables \( Y_1 \) and \( Y_2 \) which follows the densities of Poisson of parameters \( \mu_1 \) and \( \mu_2 \).

Let us put \( \lambda_3 = \frac{1}{n} \) with \( n \in \mathbb{N}^* \); we construct a family of pmf \( f_{H,n}/n \in \mathbb{N}^* \) of Holgate, with \( f_{H,n}(y_1, y_2; \mu_1, \mu_2) = f_H(y_1, y_2; \mu_1, \mu_2, \frac{1}{n}) \).

Then we have the following result:
\[ \lim_{n \to +\infty} b\left(y_1, y_2; \mu_1, \mu_2, \frac{1}{n}\right) = 1. \] (10)

**Proof of Proposition 2.**

According to the expressions (7) and (8), we have:
\[ b\left(y_1, y_2; \mu_1, \mu_2, \frac{1}{n}\right) = e^{\frac{1}{n}} \left(1 - \frac{1}{n} \mu_1\right)^{y_1} \left(1 - \frac{1}{n} \mu_2\right)^{y_2} \sum_{l=0}^{\min(y_1, y_2)} (-y_1)_l (-y_2)_l \frac{z^l}{l!} \]
with
\[ z = \left(\mu_1 - \frac{1}{n} \right) \left(\mu_2 - \frac{1}{n}\right). \]

Let us \( z_n = z \). We have
\[ b\left(y_1, y_2; \mu_1, \mu_2, \frac{1}{n}\right) = e^{\frac{1}{n}} \left(1 - \frac{1}{n} \mu_1\right)^{y_1} \left(1 - \frac{1}{n} \mu_2\right)^{y_2} \left[(-y_1)_0 (-y_2)_0 + \sum_{l=1}^{\min(y_1, y_2)} (-y_1)_l (-y_2)_l \frac{(z_n)^l}{l!}\right] \]
\[ = e^{\frac{1}{n}} \left(1 - \frac{1}{n} \mu_1\right)^{y_1} \left(1 - \frac{1}{n} \mu_2\right)^{y_2} \left[1 + \sum_{l=1}^{\min(y_1, y_2)} (-y_1)_l (-y_2)_l \frac{(z_n)^l}{l!}\right] \]

as \( \lim_{n \to +\infty} z_n = 0 \), then
\[ \lim_{n \to +\infty} b\left(y_1, y_2; \mu_1, \mu_2, \frac{1}{n}\right) = 1. \]

The proof is complete. □

The two following corollaries are easy to derive from the above results.

Corollary 2. While taking into account the expressions (9) and (10), we have the result
\[ \lim_{n \to +\infty} f_{H,n}(y_1, y_2; \mu_1, \mu_2) = f_{BP}(y_1, y_2; \mu_1, \mu_2). \]

Corollary 3. Let us consider \( F_{H,n} \) and \( F_{BP} \) the cumulative distributions functions associated to \( f_{H,n} \) and \( f_{BP} \). We have the following result.
\[ \forall (x_1, x_2) \in \mathbb{R}^2, \lim_{n \to +\infty} F_{H,n}(x_1, x_2) = F_{BP}(x_1, x_2). \]

3.2. \( \phi \)-divergence from \( f_{BP} \) to \( f_H \)

We have

Proposition 3. For the distances of the Kullback-Leibler, \( \chi^2 \), Hellinger and the variational distance (\( L^1 \)-distance) not belonging to the class introduced by Cressie and Read (Csiszár (1967)),
\[ \lim_{n \to +\infty} D_\phi (f_{H,n}, f_{BP}) = 0. \]

Therefore, the pmf \( f_{H,n} \) and \( f_{BP} \) are nearly everywhere equal asymptotically (Csiszár (1967)).

Proof Proposition 3.

By definition, while considering the distance associated to the divergence of Kullback-Leibler (Csiszár (1967)),
\[ D_\phi (f_{H,n}, f_{BP}) = \sum_{y_1=0}^{+\infty} \sum_{y_2=0}^{+\infty} f_{H,n}(y_1, y_2) \ln \left( \frac{f_{H,n}(y_1, y_2)}{f_{BP}(y_1, y_2)} \right) \]
\[ = \sum_{y_1, y_2} \left( \frac{\mu_1^{y_1}}{(y_1)!} e^{-\mu_1} \right) \left( \frac{\mu_2^{y_2}}{(y_2)!} e^{-\mu_2} \right) \times b \left( y_1, y_2; \mu_1, \mu_2, \frac{1}{n} \right) \times \ln b \left( y_1, y_2; \mu_1, \mu_2, \frac{1}{n} \right). \]

Knowing that \( \lim_{n \to +\infty} b \left( y_1, y_2; \mu_1, \mu_2, \frac{1}{n} \right) \times \ln b \left( y_1, y_2; \mu_1, \mu_2, \frac{1}{n} \right) = 0 \). We are assured of the answer.

Besides, while considering the distance of the \( \chi^2 \) (Csiszár (1967)),
\[ D_\phi (f_{H,n}, f_{BP}) = \sum_{y_1=0}^{+\infty} \sum_{y_2=0}^{+\infty} \left( \frac{f_{H,n}(y_1, y_2)}{f_{BP}(y_1, y_2)} - 1 \right)^2 f_{BP}. \]

However, we have
\[
\lim_{n \to +\infty} \frac{f_{H,n}}{f_{BP}} = \lim_{n \to +\infty} b\left(y_1, y_2; \mu_1, \mu_2, \frac{1}{n}\right) = 1,
\]
and therefore
\[
\lim_{n \to +\infty} D_{\phi}(f_{H,n}, f_{BP}) = 0.
\]

Otherwise, while considering the distance of Hellinger (Csiszár (1967)), we have
\[
D_{\phi}(f_{H,n}, f_{BP}) = \sum_{y_1=0}^{+\infty} \sum_{y_2=0}^{+\infty} \left(\sqrt{\frac{f_{H,n}}{f_{BP}}} - 1\right)^2 f_{BP}.
\]
Therefore, the result is proved.

Finally, as considering the \(L^1\) distance, we have:
\[
D_{\phi}(f_{H,n}, f_{BP}) = \sum_{y_1=0}^{+\infty} \sum_{y_2=0}^{+\infty} f_{BP} \left|\frac{f_{H,n}}{f_{BP}} - 1\right|.
\]
It follow that:
\[
\lim_{n \to +\infty} D_{\phi}(f_{H,n}, f_{BP}) = 0.
\]

While being based on the distance of Kolmogorov-Smirnov, we propose a new divergence that we call divergence of Kolmogorov-Smirnov below:

**Definition 1.**
\[
D_{KS}(F_{H,n}, F_{BP}) = \sum_{y_1=0}^{+\infty} \sum_{y_2=0}^{+\infty} \sup_{y_1 \leq x_1, y_2 \leq x_2} |F_{H,n}(x_1, x_2) - F_{BP}(x_1, x_2)|, \quad \forall (x_1, x_2) \in \mathbb{R}^2.
\]
With \(y_{1i}\) (\(y_{2i}\)) the \(i^{th}\) realization of \(Y_1\) (\(Y_2\)).

Here is the result based on that divergence measure.

**Proposition 4.** We have the following results:
\[
\lim_{n \to +\infty} D_{KS}(F_{H,n}, F_{BP}) = 0
\]
and
\[
D_{KS}(F_{H,n}, F_{BP}) = D_{L^1}(f_{H,n}, f_{BP}).
\]
Proof of Proposition 4.

The first point of the proposition is obvious and then omitted.

As to the second, we have, by definition: \( \forall (x_1, x_2) \in \mathbb{R}^2, F_{H,n} (x_1, x_2) = \sum_{y_1 \leq x_1, y_2 \leq x_2} f_{H,n} (y_1, y_2), \)

and:

\[
\left| F_{H,n} (x_1, x_2) - F_{BP} (x_1, x_2) \right| = \left| \sum_{y_1 \leq x_1, y_2 \leq x_2} f_{H,n} (y_1, y_2) - f_{BP} (y_1, y_2) \right| \\
\leq \sum_{y_1 \leq x_1, y_2 \leq x_2} f_{H,n} (y_1, y_2) - f_{BP} (y_1, y_2),
\]

and therefore,

\[
\sup_{y_1 \leq x_1, y_2 \leq x_2} \left| F_{H,n} (y_1, y_2) - F_{BP} (y_1, y_2) \right| = \left| f_{H,n} (y_1, y_2) - f_{BP} (y_1, y_2) \right|.
\]

Finally, we have

\[
D_{KS} (F_{H,n}, F_{BP}) = \sum_{y_1 = 0}^{+\infty} \sum_{y_2 = 0}^{+\infty} \left| F_{H,n} (y_1, y_2) - f_{BP} (y_1, y_2) \right| \\
= \sum_{y_1 = 0}^{+\infty} \sum_{y_2 = 0}^{+\infty} f_{BP} (y_1, y_2) \left| \frac{f_{H,n} (y_1, y_2)}{f_{BP} (y_1, y_2)} - 1 \right| \\
= D_{L_1} (f_{H,n}, f_{BP}).
\]

This closes the proof. \( \square \)

We present the following graph to illustrate this comparison (see figures 1 and 2).

Concluding remarks

If we note \( D_H, D_{KL}, D_2 \) and \( D_{L_1} \) the divergences according to the respective distances of Hellinger, Kullback-Leibler, Khi-2 and \( L_1 \); it takes out again of the following figures that \( D_H \leq D_{KL} \leq D_2 \leq D_{L_1} \). The graphs of the function (or sequences) \( n \mapsto D_\phi (f_{H,n}, f_{BP}), \phi \in I, \) with \( I = \{ H, KL, \chi^2, L_1 \} \) are monotonic decreasing to the lower bound zero:

\( \forall \varepsilon > 0, \exists N_\phi (\varepsilon) \in \mathbb{N}, \) such as \( \forall n \geq N_\phi (\varepsilon) \Rightarrow D_\phi (f_{H,n}, f_{BP}) \leq \varepsilon. \) The integers \( N_\phi (\varepsilon), \phi \in I, \) inform us on the speed of convergence of the sequences.

---


For $\varepsilon$ (arbitrarily small) given, a graphic resolution allows us to determine the $N_\phi (\varepsilon), \phi \in I$. We found, for $\varepsilon = 10^{-4}$, $N_H (\varepsilon) = 100$, $N_{KL} (\varepsilon) = 150$, $N_{\chi^2} (\varepsilon) = 200$ et $N_{L^1} (\varepsilon) > 2000$. For the smallest integer $N_H (\varepsilon) = 100$, we cannot say that $D_H$ converge very quickly toward zero. On the other hand, one can compare the speeds of convergence as follows: $D_H$ converge more quickly than $D_{KL}$ toward zero ; $D_{KL}$ converges more quickly than $D_{\chi^2}$ toward zero and that $D_{\chi^2}$ converges more quickly than $D_{L^1}$ toward zero. The divergences tend to zero. It attests that the two distributions are nearly everywhere equal asymptotically.

The pmf of the bivariate Poisson distribution according to Berkhout and Plug is the asymptotic density of the couple $(Y_1, Y_2)$, while the pmf of the bivariate distribution according to Holgate is her exact density.

References


Fig. 1. \( \mu_1 = 0.4, \mu_2 = 0.6, y_1 = y_2 \in [0, 32] \) and \( n \in [80, 500] \)

Fig. 2. \( \mu_1 = 0.4, \mu_2 = 0.6, y_1 = y_2 \in [0, 32] \) and \( n \in [80, 2000] \)