HOMENAJE AL PROFESOR SIXTO RIOS
TRABAJOS DE ESTADISTICA Y DE INVESTIGACION OPERATIVA
Vol. 36, Núm. 3, 1985, pp. 101 a 111

A NOTE ON POISSON APPROXIMATION

Paul Deheuvels

Université Paris VI . 7 Avenue du Chateau, 92340 Bourg-la-Reine, France .

We obtain in this note evaluations of the total variation distance and of the Kolmogorov-Smirnov distance between the sum of n random variables with non identical Bernoulli distributions and a Poisson distribution. Some of our results precise bounds obtained by Le Cam, Serfling, Barbour and Hall.

It is shown, among other results, that if $p_1=P(X_1=1),\dots,p_n=P(X_n=1)$ satisfy some appropriate conditions, such that $p = 1/n\ \Sigma_i p_i \to 0$, $np \to \infty$ $np^2 \to 0$, then, the total variation distance between $X_1+\dots+X_n$ and a Poisson distribution with expectacion np is $p\,(2\Pi e)^{-1/2}(1+o(1))$.

Key words: Poisson approximation.

AMS Classification:(1980): Primary, 60F05; Secondary, 60K99.

Sobre Aproximaciones de la Ley de Poisson

En este trabajo, consideramos evaluaciones de la distancia en variación entre leyes de Poisson, Binomial y de sumas de variables de Bernoulli independientes.

Palabras Clave: Aproximación Poisson.

Clasificación AMS (1980): Primaria, 60F05; Secundaria, 60K99.

1. INTRODUCTION AND RESULTS.

Let $X_1,\dots,X_n$ be Bernoulli variables, and let $S_n = X_1+\dots+X_n$, $p_i = P(X_i=1)$, and $\lambda = \Sigma_i p_i = np$, $i=1,\dots,n$.

It is well known that, under appropriate conditions on $p_1,\dots,p_n$, the distribution $L(S_n)$ of S_n can be closely approximated by a Poisson distribution P_λ with expectation λ. The adequacy of such an approximation can be measured by the total variation distance d(.,.) and by the Kolmogorov distance $d_K(.,.)$:

$$d(\mu,\upsilon) = \sup_{A \subset Z} | \mu(A) - \upsilon(A) |, \quad d_K(\mu,\upsilon) = \sup_x | \mu((-\infty, x)) - \upsilon(-\infty, x)) |.$$

The following upper bounds for $d(L(S_n), P_\lambda)$ are known:

$$d(L(S_n), P_\lambda) \leq \lambda^{-1} (1- e^{-\lambda}) \Sigma_i \ p_i^2 , \ i=1,...,n \ \text{(Barbour and Hall, 1984)}, \quad (1)$$

$$d(L(S_n), P_\lambda) \leq \Sigma_i p_i(1- \exp(- p_i)) \leq \Sigma_i p_i^2, \quad i=1,...,n \ \text{(Le Cam, 1960)}. \quad (2)$$

In the case where $p= p_1=...= p_n$, it can be verified that (1) gives always a sharper bound that (2), since we get

$$d(L(S_n), P_\lambda) \leq p(1- e^{-np}) = p(1- \{1- (1- e^{-p})\}^n) \leq np(1- e^{-p}) . \quad (3)$$

Barbour and Hall's result precise also the upper bound of Romanowska (1979) who showed that when $p= p_1=...= p_n$, we have

$$d(L(S_n), P_\lambda) \leq p (1- p)^{-1/2}; \quad (4)$$

by (1), we get in general the bound

$$d(L(S_n), P_\lambda) \leq \max_{1 \leq i \leq n} p_i . \quad (5)$$

Even though (1) is sharp when $\lambda \to 0$ (it gives even the best result for n=1), it does not come near the best possible evaluation when $\lambda \to 0$. This follows from the results of Deheuvels and Pfeifer (1984) which we cite in Theorem A.

Theorem A. Let $p= p(n)= p_1=...= p_n$ be such that $np \to \alpha$ as $n \to \infty$. We have the following results.

1) If $\alpha = 0$, then

$$d(L(S_n), P_\lambda) = np^2 (1+ o(1)) . \quad (6)$$

2) If $0 < \alpha < \infty$, and if

$$R = [\alpha + 2^{-1} - (\alpha + 4^{-1})^{1/2}] \ \text{and} \ S = [\alpha + 2^{-1} + (\alpha + 4^{-1})^{1/2}], \ \text{then}$$

$$d(L(S_n), P_\lambda) = 2^{-1}np^2\{(\alpha^{S-1}(S- \alpha)) /S! - (\alpha^{R-1}(R- \alpha)) /R!\}e^{-\alpha} (1+ o(1)). \quad (7)$$

3) If $\alpha \to \infty$, and if, in addition, $np^2 \to 0$ as $n \to \infty$, then

$$d(L(S_n), P_\lambda) = p(2\Pi e)^{-1/2} (1+ o(1)). \tag{8}$$

Note that in (7) we can replace np by α, and that, when $\alpha \to 0$, $R(\alpha) \to 0$ and $S(\alpha) \to 1$, which yields (6) as a limiting case.

We shall extend in the sequel Theorem A to the non-identically distributed case. This will be achieved through the following result.

Theorem 1. Let $X_1,...,X_n$ be independent Bernoulli random variables. Let $S_n = X_1+...+X_n$ and $p_i = P(X_i=1)$. Let S_n^* denote a binomial $B(n,p)$ random variable, where $p = \lambda/n = n^{-1}\Sigma_i p_i$, $i=1,...,n$. Then

$$d(L(S_n),L(S_n^*)) \leq \{1 - n^{-1} \Sigma_i p_i + n^{-1}(\Sigma_i p_i(1- p_i)^{-1}) (\Pi_i (1- p_i))^{1/n} \}^n - 1 . \tag{9}$$

Before getting further, it is worthwhile to obtain a simpler expression for (9).

Corollary 1. Under the assumptions of Theorem 1, if we assume further that

$$\max_{1<i<n} p_i \to 0 \text{ and } \Sigma_i p_i^2 - n^{-1}(\Sigma_i p_i^2) \to 0, \quad i=1,...,n ; \tag{10}$$

then

$$d(L(S_n), L(S_n^*)) \leq \{ \Sigma_i p_i^2 - n^{-1} (\Sigma_i p_i)^2 \} (1+ o(1)) \quad i=1,...,n . \tag{11}$$

Corollary 2. Under the assumptions of Theorem 1, if we assume in addition that

$$\lim_{n\to\infty} n\{ (\Sigma_i p_i^2/ \Sigma_i p_i) - n^{-1}\Sigma_i p_i \} = 0 , \quad i=1,...,n ; \tag{12}$$

then the results (6), (7) and (8) of Theorem A can be extended to the case where p is replaced by $n^{-1}\Sigma_i p_i$, $i=1,...,n$.

Corollary 2 follows from Corollary 1 and the triangle inequality

$$d(L(S_n), P_\lambda) \leq d(L(S_n^*), P_\lambda) + d(L(S_n), L(S_n^*)) . \tag{13}$$

We could use reversely this bound, starting from (1) and (8) to show for instance that, if $\lambda \to 0$ and if $\lambda^2 n^{-1} \log \lambda \to 0$, then we must have

$$d(L(S_n), L(S_n^*)) \leq \lambda^{-1}\{ \Sigma_i p_i^2 - (1 / n(2\Pi e)^{1/2}) (\Sigma_i p_i)^2 \}, \quad i=1,\dots,n . \qquad (14)$$

It can be seen that (14) may be on some circumstances better than (11).

We shall also precise in this paper the bounds obtained by Serfling (1978) for the following different choice of λ. Let in general $\lambda = \Sigma_i \lambda_i$, $i=1,\dots,n$. We have, up to now, considered the case where $\lambda_i = p_i$. However, some different choices may be of interest as follows from Theorem B due to Serfling (1978).

Theorem B. Let $\gamma_i = -\log(1-p_i)$ and let $\gamma = \Sigma_i \gamma_i$, $i=1,\dots,n$, then

$$d(L(S_n), P_\gamma) \leq \Sigma_i (1-(\gamma_i + 1)\exp(-\gamma_i)) \leq 2^{-1}\Sigma_i \gamma_i^2, \quad i=1,\dots,n . \qquad (15)$$

We shall prove in the sequel Theorems 2 and 3:

Theorem 2. We have

$$d(L(S_n), P_\lambda) \leq 1- \Pi_i (1-p_i(1-\exp(-p_i))) \leq \Sigma_i p_i(1-\exp(-p_i)) \leq \Sigma_i p_i^2, \qquad (16)$$
$$i=1,\dots,n$$

and

$$d(L(S_n), P_\gamma) \leq - \Pi_i \{(\gamma_i+1)\exp(-\gamma_i)\} \leq \Sigma_i (1-(\gamma_i+1)\exp(-\gamma_i)) \leq 2^{-1}\Sigma_i \gamma_i^2,$$
$$i=1,\dots,n .$$

Theorem 3. We have

$$d_K(L(S_n), P_\lambda) \leq 1- \Pi_i (1-(\exp(-p_i)-1+p_i)) \leq \Sigma_i (\exp(-p_i)-1+p_i) \leq 2^{-1}\Sigma_i p_i^2 \qquad (17)$$
$i=1,\dots,n$, and

$$d_K(L(S_n), P_\gamma) \leq 1- \Pi_i \{(\gamma_i+1)\exp(-\gamma_i)\} \leq \Sigma_i (1-(\gamma_i+1)\exp(-\gamma_i)) \leq 2^{-1}\Sigma_i \gamma_i^2,$$
$$i=1,\dots,n .$$

We remark that the first inequality in (16) is mentioned by Le Cam (1960), while the second inequality of (16), and the inequalities of (17) precise results of Serfing (1978).

2. PROOF OF THE THEOREMS.

We start with a simple lemma.

Lemma 1. We have the inequality

$$\Sigma_i p_i \leq (\Sigma_i p_i (1-p_i)^{-1}) \{\Pi_i (1- p_i) \}^{1/n} , \quad i=1,...,n . \tag{18}$$

Proof. Put $q_i = p_i(1- p_i)^{-1}$ (18) is equivalent tothe inequality

$$\Sigma_i q_i \geq (\Sigma_i q_i(1+ q_i)^{-1}) \{\Pi_i (1+ q_i) \}^{1/n}, \quad i=1,...,n .$$

But $\{\Pi_i (1+ q_i)\}^{1/n} \leq 1+ n^{-1} \Sigma_i q_i$, $i=1,...,n$. (see e.g. Hardy-Littlewood-Pólya, p.17), and hence, it suffices to prove that

$$(\Sigma_i q_i(1+ q_i)^{-1}) \{1+ n^{-1} \Sigma_i q_i\} \leq \Sigma_i q_i = \Sigma_i q_i (q_i+1)(1+q_i)^{-1}, \quad i=1,...,n.$$

This in turn is equivalent to show that

$$\Sigma_i q_i (1+q_i)^{-1} \{q_i - n^{-1} \Sigma_i q_i\} = - \Sigma_i (1+q_i)^{-1} \{q_i - n^{-1} \Sigma_i q_i\} \geq 0.$$

This last result follows from the fact that, if $\mu = n^{-1} \Sigma_i q_i$, then

$$(q_i - \mu) (q_i +1)^{-1} \leq (q_i - \mu) (\mu+1)^{-1}, \text{ and hence}$$

$$\Sigma_i (q_i - \mu) (q_i + 1)^{-1} \leq \Sigma_i (q_i - \mu) (\mu +1)^{-1} = 0, \quad i=1,...,n .$$

Lemma 2. We have the inequality

$$P(S_n= k) \leq C(n,k) (1- n^{-1} \Sigma_i p_i)^{n-k} ((n^{-1} \Sigma_i p_i(1- p_i)^{-1}) \{\Pi_i(1- p_i)\}^{1/n})^k, \tag{19}$$
$k=0,1,...,n$.

Where $C(n,k) = n! / (n - k)! k!$

Proof. Consider the expansion

$$(x+a_1)...(x+a_n) = x^n+ C(n,1)R_1x^{n-1}+ C(n,2)R_2x^{n-2}+...+R_n .$$

The following inequality is due to Mac Laurin (see e.g. Hardy - Littlewood - Pólya, p.51):

$$R_1 \geq R_2^{1/2} \geq \ldots \geq R_n^{1/n}.$$

We have evidently

$$P(S_n = k) = \prod_i (1-p_i)\, C(n,k) R_k, \text{with } a_i = p_i(1-p_i)^{-1} \text{ and } R_1 = n^{-1} \Sigma_i p_i (1-p_i)^{-1}.$$
$i=1,\ldots,n$

It follows that

$$P(S_n = k) \leq \{\prod_i (1-p_i)\}\, C(n,k)\, \{n^{-1}\Sigma_i p_i(1-p_i)^{-1}\}^k \leq \qquad (20)$$

$$A_k = C(n,k)\,\{1- n^{-1}\Sigma_i p_i\}^{n-k}\{(n^{-1}\Sigma_i p_i(1-p_i)^{-1})\,\{\prod_i (1-p_i)\}^{1/n}\}^k, \quad i=1,\ldots,n.$$

Here, we have used again Cauchy's inequality (see e.g. Hardy-Littlewood - Pólya, p.17) by which $a_1 \ldots a_n \leq (n^{-1}\Sigma_i\, a_i)^n$.

We now compare (20) to $P(S_n{*} = k) = C(n,k)\,\{1- n^{-1}\Sigma_i\, p_i\}^{n-k}\,\{n^{-1}\Sigma_i p_i\}^k$.

By Lemma 1, $A_k \geq P(S_n = k)$, and hence,

$$d(L(S_n),L(S_n{*})) = \Sigma_i \max\{0, P(S_n = k) - P(S_n{*} = k)\} \leq \Sigma_i \{A_k - P(S_n{*} = k)\} = \Sigma_i A_k - 1,$$
$i=1,\ldots,n$

which proves easily Theorem1.

The proofs of Corollary 1 and Corollary 2 follow by straightforward expansions. The proof of Theorem 2 is based on a different technique, which will be described in the sequel. We consider first the case n=1, and put $x = x_1$, $p = p_1$ and define Y to be a random variable with a Poisson P_λ distribution, where λ is arbitrary. We get easily:

$$d(L(X),L(Y)) = 2^{-1}\{\,|1-p-e^{-\lambda}| + |p-\lambda e^{-\lambda}| + (1-(1+\lambda)\,e^{-\lambda})\,\}.$$

Lemma 3. Let $\gamma = -\log(1-p)$. Then, we have the following results:

1) $0 < p < \gamma$.

2) The equation in λ: $\lambda e^{-\lambda} = p$ has:

(i) Two roots $0 < \gamma < \lambda_a(p) \leq 1 \leq \lambda_b(b) < \infty$ when $0 < p < e^{-1}$;
(ii) One root $0 < \gamma < \lambda_a(p) = \lambda_b(p) = 1$ when $p = e^{-1}$;
(iii) No root otherwise.

We have $\lambda e^{-\lambda} \leq p$ whenever:

(i) $p \geq e^{-1}$;
(ii) $p < e^{-1}$ and either $\lambda \leq \lambda_a(p)$ or $\lambda \geq \lambda_b(p)$.

In any other case, $\lambda e^{-\lambda} > p$.

Lemma 4.

1) If $0 < \lambda \leq \gamma = -\log(1-p)$, then

$$d(L(Y),L(Y)) = p - \lambda e^{-\lambda} \geq p - \gamma e^{-\gamma} = \Sigma_j p^j (j(j-1))^{-1} =$$
$$1 - (\gamma + 1) e^{-\gamma} = \Sigma_j ((j-1)/j!)\, \gamma^j (-1)^j , \quad j = 2,\ldots,\infty .$$

In particular, for $\lambda = p$, we have

$$d(L(X),L(Y)) = p\,(1 - e^{-p}) = \Sigma_j\, (p^j / (j-1)!)\, (-1)^j \leq p^2.$$

2) If $\gamma < \lambda \leq \lambda_a(p)$, or $\lambda \geq \lambda_b(p)$, or $p \geq e^{-1}$ and $\lambda > \gamma$, then

$$d(L(X),L(Y)) = 1 - (\lambda + 1)e^{-\lambda} > 1 - (\gamma + 1)e^{-\gamma}.$$

3) If $\lambda_a(p) \leq \lambda \leq \lambda_b(p)$ and $0 < p < e^{-1}$, then

$$d(L(X),L(Y)) = 1 - p - e^{-\lambda} \geq 1 - (\lambda + 1)e^{-\lambda} > 1 - (\gamma + 1)e^{-\gamma}.$$

4) In all cases,

$$\inf_\lambda d(L(X),L(Y)) = 1 - (\gamma + 1)e^{-\gamma} = p - (p - 1)\log(1 - p) = \Sigma_j\, p^j (j(j-1))^{-1} < 2^{-1}\gamma^2 .$$

The result of Lemma 4 is due to Serfling (1978) for a general λ. The case λ=p has been treated by Le Cam (1960). We may compute likewise $d_K(L(X),L(Y))$:

$$d_K(L(X),L(Y)) = \max\{ |1- p - e^{-\lambda}|, 1- (1 + \lambda)e^{-\lambda} \} .$$

Lemma 5. Let $\gamma = -\log(1- p)$ and let $\lambda_a(p)$ be defined as Lemma 3 (for $0<p<e^{-1}$). Let also $\lambda_c(p)$ be defined as the unique positive root of the equation $(\lambda + 2)e^{-\lambda}= 2- p$. Then :

1) If $\gamma < \lambda < \lambda_a(p)$, or $\lambda > \lambda_b(p)$, or $p \geq e^{-1}$ and $\lambda > \gamma$, then

$$d_K(L(X),L(Y)) = 1- (\lambda+1)e^{-\lambda} > 1- (\gamma + 1)e^{-\gamma}.$$

2) If $\lambda_a(p) \leq \lambda < \lambda_b(p)$ and $0 < p < e^{-1}$, then

$$d_K(L(X),L(Y)) = 1- p - e^{-\lambda} \geq 1 - (\lambda+1)e^{-\lambda}.$$

3) If $\lambda_c(p) \leq \lambda \leq \gamma$, then

$$d_K(L(X)),L(Y)) = 1- (\lambda + 1)e^{-\lambda}.$$

4) If $0 < \lambda < \lambda_c(p)$, then

$$d_K(L(X),L(Y)) = e^{-\lambda} - 1 + p .$$

This covers the case $\lambda = p$, for which we have $d_K(L(X),L(Y)) = e^{-p} - 1- p$.

5) In all cases

$$\inf_\lambda d_K(L(X),L(Y)) = 1- (\lambda_c(p) + 1)\exp(- \lambda_c(p)) \leq e^{-p} - 1+ p < 2^{-1} p^2 .$$

The result of Lemma 5 was obtained for $\lambda = p$ by Daley (1975, see e.g. Serfling (1978)).

We shall now specialize in the two following cases:

A) $\lambda= p$. In this case, we note that

$$P(X=0)= 1- p < p(Y=0)= e^{-p} \text{ and } P(X=1)= p > P(Y=1)= pe^{-p} .$$

It follows that one can easily construct X and Y on the same probability space, by setting: $Y = XZ$, where Z is independent of X and such that

$$P(Z=0) = (P(Y=0) - P(X=0)) / P(X=1) = (e^{-p} - 1 + p) / p,$$
$$P(Z=k) = P(Y=k) / P(X=1) = (p^{k-1} / k!)\, e^{-p} \quad (k=1,2,...).$$

For such a construction, we have

$$P(X \neq Y) = P(X=1)\, P(Z \neq 1) = p(1 - e^{-p}) = d(L(X),L(Y)).$$

We see that the coupling betweem X and Y is then maximal in the sense that the upper bound $d(L(X), L(Y)) \leq P(X \neq Y)$ is reached.

For this same coupling, we have

$$\max\{P(X < Y), P(X > Y)\} = \max\{P(X=1)\, P(Z \geq 2), P(X=1)\, P(Z=0)\} =$$
$$\max\{e^{-p} - 1 + p,\ 1 - (p+1)e^{-p}\} = e^{-p} - 1 + p = d_K(L(X),L(Y)).$$

We sum up these results in

Lemma 6. For the contruction above, we have $d(L(X),L(Y)) = P(X \neq Y)$, and $d_K(L(X),L(Y)) = \max\{P(X<Y),P(X>Y)\} = P(X<Y)$.

B) $\lambda = \gamma = -\log(1-p)$. In this case, we have

$$P(X=0) = 1-p = P(Y=0) = e^{-\gamma}, \text{ and } P(X=1) = p > P(Y=1) = \gamma e^{-\gamma}.$$

It follows that one can easily construct X and Y on the same probabitility space by setting: $Y = XZ$, where Z is independent of X and such that $P(Z=0) = 0$ and

$$P(Z=k) = P(Y=k) / p(X=1) = (\gamma^k / pk!)\, e^{-\gamma} \quad (k=1,2,...).$$

For such a construction, we have

$$P(X \neq Y) = P(X=1)\, P(Z \neq 1) = p(1 - (\gamma/p)\, e^{-\gamma}) =$$
$$p - \gamma e^{-\gamma} = 1 - (\gamma+1)\, e^{-\gamma} = d(L(X),L(Y)).$$

We also have here $P(X<Y) = 1$, and hence

$$P(X<Y) = P(X\neq Y) = 1-(\gamma+1)\,e^{-\gamma} = d_K(L(X),L(Y)), \quad P(X>Y)=0.$$

We sum up these results in:

Lemma 7. For the construction above, we have $d(L(X),L(Y)) = P(X\neq Y)$, and

$$d_K(L(X),L(Y)) = \max\{P(X<Y),\ P(X>Y)\} = P(X<Y)\ .$$

Proof of Theorem 2. Using any of the constructions obtained in Lemmas 6 and 7, we define X_i and Y_i jointly for $i=1,\ldots,n$, and let $T_n = Y_1+\ldots+Y_n$. We have then

$$d(L(S_n),L(T_n)) \leq P(S_n\neq T_n) \leq 1-\prod_i P(X_i=Y_i),$$

which proves Theorem 2.

Proof of Theorem 3. We have likewise

$$d_K(L(S_n),L(T_n)) \leq \max\{P(S_n<T_n),\ P(S_n>T_n)\}\ .$$

Taking any one of these probabilities (for instance $P(S_n<T_n)$), we have

$$P(S_n<T_n) = 1 - P(S_n\geq T_n) \leq 1-\prod_i (1 - P(X_i<Y_i)).$$

REMARK

By using a direct method, the result in Theorem A obtained originally in Deheuvels and Pfeifer (1984) for $p = p_1 = \ldots = p_n$ has been extended to cover the non i.i.d. case under weaker conditions as those of Corollary 2.

AKNOWLEDGEMENT

It is a great pleasure for me to dedicate this paper to Professor Sixto Rios whose great scientific achievements have been stimulating to a generation of probabilists and statisticians.

REFERENCES

BARBOUR, A.D. and HALL, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Phil. Soc.* **95**, 473-480

DEHEUVELS, P. and PFEIFER, D. (1984). A semi-group approach to Poisson approximation, Preprint nº 20, LSTA, Université Paris VI, to appear in Annals of Probability.

HARDY, G.H., LITTLEWOOD, J.E. and POLYA, G. (1952). Inequalities, second edition, Cambridge University Press, Cambridge.

LE CAM, J.L.(1960). An approximation theorem for the Poisson binomial distribution, *Pacific J. Math.* **10**, 1181-1197.

ROMANOWSKA, M. (1979). A note on the upper bound for the distance in total variation between binomial and the Poisson distribution, *Statistica Neerlandica,* 127-130.

SEFLING, R.J. (1978). Some elementary results on Poisson approximation in a sequence of Bernoulli trials, *SIAM Review*, **20**, Nº 3, 567-579.