POISSON APPROXIMATION FOR INDEPENDENT BERNOULLI SUMMANDS USING THE SEMI-GROUP OPERATORS THEORY

A. R. Carrera, M. Martinez and J. M. González

Universidad del Pais Vasco — Euskal Herriko Universitatea Departamento de Matemática Aplicada. Escuela Técnica Superior de Ingenieros Industriales y de Ingenieros de Telecomunicación de Bilbao
Alda. de Urquijo s/n. 48013 Bilbao, Spain

Received August 11, 1988

1. Introduction

The Semi-Group of Operators Theory is especially useful in the treatment of problems concerning the approximation of some probability measures by others.

The particular instance of Poisson approximation for independent summands, can be treated in a suitable operator semigroup framework, allowing at the same time precise evaluations for the norm of the difference between the associated probability distributions.

In this paper we study the concrete case in which the independent summands are Bernoulli random variables with different probabilities of success and the considered norm induces the total variation distance.

The space of all absolutely summable sequences is considered and also its subset of all probability measures with support contained in the non-negative integers. The function that maps the sum of the absolute value terms to each sequence, verifies the axioms of a complete norm.

An operator from this space to itself is defined via the convolution. Hence, we may determine the infinitesimal generator of the Poisson convolution semigroup. Under these assumptions, we can formulate the approximation problem for independent Bernoulli summands, by the semigroup mentioned before and with the total variation distance, as a norms evaluation problem in this Banach space.

The first question we will solve in this work is to establish a first estimate of the norm term commented above, when the Poisson approximating operator is a general one. Finally, we will find the Poisson random variable which optimizes the approximation in the sense that it minimizes the expression of the total variation distance obtained. To achieve this, we will use some auxiliary results from the semigroup approach theory, and other previous techniques given by different authors, such as Le Cam, Chen, Barbour y Hall.

2. The semigroup approach for the Poisson approximation problem

Let the Banach space l^1 of all absolutely summable sequences be

$$l^1 = \{\{x_n\}_{n \in N} / \sum_{n=1}^{\infty} |x_n| < \infty\} \tag{2.1}$$

where for each sequence $\{x_n\}_{n \in N}$ its norm will be:

$$||\{x_n\}_{n \in N}|| = \sum_{n=1}^{\infty} |x_n|.$$

Let M be the subset of all probability measures with support contained in the non-negative integers:

$$M = \{m \colon Z^+ \to [0, 1] / m \text{ is a probability function}\}. \tag{2.2}$$

For two elements $\{f(n)\}_{n \in N}$, $\{g(n)\}_{n \in N} \in l^1$, we define the convolution as:

$$(f * g)(n) = \sum_{k=0}^{n} f(k) g(n-k) \quad n \geq 0 \tag{2.3}$$

that also belongs to l^1 and its norm verifies:

$$||f * g|| \leq ||f|| \, ||g|| \tag{2.4}$$

Let us define the operators B^k on l^1 by:

$B^k \colon l^1 \to l^1$

$f \to E_k * f \colon N \to R$

$n \mapsto E_k * f(n)$

with E_k denoting the unit mass at point $k \in N$. Then

$$B^k f(n) = (E_k * f)(n) = \sum_{m=0}^{n} E_k(m) f(n-m) =$$

$$= \begin{cases} f(n-k) & \text{if} \quad k \leq n \\ 0 & \text{if} \quad k > n \end{cases} \tag{2.5}$$

Therefore, any measure $m \in M$ can be interpreted as an operator on l^1, such as:

$$mf(n) = m * f(n) = \sum_{k=0}^{n} m(k) f(n-k) = \sum_{k=0}^{\infty} m(k) B^k f(n). \tag{2.6}$$

Let I be the identity mapping from l^1 to l^1, and $A = B - I$ that is the infinitesimal generator of the convolution semigroup:

$$\{e^{tA} / 0 \leq t < \infty\}.$$

But

$$e^{tAf} = e^{-tI}e^{tBf} = \sum_{k=0}^{\infty} e^{-t}t^k/k\,!B^{kf} = \sum_{k=0}^{\infty} e^{-t}t^k/k\,!E_k * f = P_0(t) * f \quad (2.7)$$

where $P_0(t) = \sum_{k=0}^{\infty} e^{-t}\, t_k/k\,!\; E_0$ denotes the Poisson distribution with mean t. Then A is the infinitesimal generator of Poisson convolution semigroup.

On the other hand, given two probability measures $m_i \in M$ $i = 1, 2$ we can interpret the total variation distance between them as an 1^1 operator norm, since:

$$d(m_1, m_2) = (1/2) \sum_{k=0}^{\infty} |m_1(k) - m_2(k)| = (1/2)||(m_1 - m_2) * E_0|| \quad (2.8)$$

with E_0 the unit mass at point 0.

Hence, we can formulate the approximation problem with the total variation metric for independent Bernoulli summands by a Poisson distribution, as a norms evaluation problem in this Banach space.

To achieve this, let $X_1, \ldots, X_n$ be an independent sequence of Bernoulli random variables with successful probabilities $p_1, \ldots, p_n$, and probability distributions $m_1, \ldots, m_n$. Let $S_n = \sum_{i=1}^{n} X_i$ with probability distribution be the convolution of $m_1, \ldots, m_n$, that will be denoted by $\prod_{i=1}^{n} m_i$. Let $Y_1, \ldots, Y_n$ be an independent sequence of Poisson random variables with expectation $\mu_1, \ldots, \mu_n$ and probability distributions $P_0(\mu_1), \ldots, P_0(\mu_n)$. Its sum $T_n(\mu)$ is another Poisson random variable with mean $\mu = \sum_{i=1}^{n} \mu_i$ and probability distribution:

$$P_0(\mu) = P_0(\mu_1) * P_0(\mu_2) * \ldots * P_0(\mu_n) = \prod_{i=1}^{n} P_0(\mu_i)$$

For (2.8), (2.7) and (2.4) we can write:

$$d(S_n, T_n(\mu)) = 1/2||\,(\prod_{i=1}^{n} e^{\mu_i A} - \prod_{i=1}^{n} m_i) * E_0|| \leq 1/2||\prod_{i=1}^{n} e^{\mu_i A} - \prod_{i=1}^{n} m_i||. \quad (2.9)$$

Using the results concerning the convolution of n operators $T_1, \ldots, T_n$ and $G_1, \ldots, G_n$, that establishes:

$$||T_1 * T_2 * \ldots * T_n - G_1 * G_2 * \ldots * G_n|| \leq \sum_{i=1}^{n} ||T_i - G_i||. \quad (2.10)$$

the inequality (2.9) can be transformed as

$$d(S_n, T_n(\mu)) = 1/2||\prod_{i=1}^{n} e^{\mu_i A} - \prod_{i=1}^{n} m_i\,|| \leq 1/2 \sum_{i=1}^{n} ||e^{\mu_i A} - m_i||. \quad (2.11)$$

As m_i is the Bernoulli distribution with $p\{X_i = 1\} = p_i \; i = 1, \ldots, n$, then by (2.6) we have:

$$m_i f = m_i(0) B^0 f + m_i(1) B f = [(1 - p_i) I + p_i B] f = (I + p_i A) f \;\; \forall i = 1, \ldots, n. \tag{2.12}$$

Consequently, (2.11) can be written like:

$$d(S_n, T_n(\mu)) \leq 1/2 \sum_{i=1}^{n} ||e^{\mu_i A} - (I + p_i A)|| \tag{2.13}$$

3. Evaluation of the total variation distance following the semi-group theory

In this paragraph, we will calculate the values of the total variation distance between the Binomial-Poisson random variable S_n defined above and a general Poisson distribution $T_n(\mu)$ with mean μ, using the operators defined in the previous section. We consider again the Poisson random variable $T_n(\mu)$ as a sum of "n" Poisson random variables $Y_1, \ldots, Y_n$ with expectations $\mu_1, \ldots, \mu_n$, that is $\mu = \sum_{i=1}^{n} \mu_i$.

Theorem 1. Under these hypotheses the value of the total variation distance between the random variables S_n and $T_n(\mu)$ is given by the expression:

$$d(S_n, T_n(\mu)) = 1/2 || \sum_{i=1}^{n} (\mu_i - p_i) \exp\left(\left(\sum_{i=1}^{n} p_i\right) A\right) A \, E_0 +$$

$$+ 1/2 \sum_{i=1}^{n} p_i^2 \exp\left(\left(\sum_{i=1}^{n} p_i\right) A\right) A^2 E_0 || + r_n(p) + S_n(p, \mu) \tag{3.1}$$

with

$$r_n(p) \leq 2 \sum_{i=1}^{n} p_i^3 + 2 \left\{\sum_{i=1}^{n} p_i^2\right\} \max \{p_1, \ldots, p_n\} \tag{3.2}$$

and

$$S_n(p, \mu) \leq 1/4 \left\{\sum_{i=1}^{n} (\mu_i - p_i)\right\}^2 \max \left\{ ||\exp\left(\left(\sum_{i=1}^{n} \mu_i\right) A\right) A^2 E_0||, \right.$$

$$\left. ||\exp\left(\left(\sum_{i=1}^{n} p_i\right) A\right) A^2 E_0|| \right\} \leq$$

$$\leq \max \left\{ \min\left(1, 1/\left(\sum_{i=1}^{n} \mu_i\right)\right), \min\left(1, 1/\left(\sum_{i=1}^{n} p_i\right)\right) \right\} \left\{\sum_{i=1}^{n} (\mu_i - p_i)\right\}^2. \tag{3.3}$$

We do not expose here the proof of this result because it was completely proved in another article of these authors entitled "Application of Semi-group Theory to the norms evaluation of probability measures" [5]. In that work we established another theorem which gives a precise evaluation of the norm terms in the distance expression obtained before, that appear in the relation (3.1) and (3.3).

Theorem 2. For $t > 0$, $\gamma \in R$, we get:

$$||e^{tA} A\, E_0|| = 2e^{-t} \frac{t^{[t]}}{[t]\,!} \sim \frac{2}{2\pi t} \quad \text{for} \quad t \to \infty \tag{3.18}$$

where $[t]$ denotes the integer part of t.

$$||e^{tA} A^2\, E_0|| = 2 \left\{ \frac{t^{\alpha-1}(\alpha - t)}{\alpha\,!} - \frac{t^{\beta-1}(\beta - t)}{\beta\,!} \right\} e^{-t} \sim \frac{4}{t\sqrt{2\pi e}} \quad \text{for} \quad t \to \infty \tag{3.19}$$

where:

$$\alpha = [t + 1/2 + (t + 1/4)^{1/2}] \qquad \text{and} \qquad \beta = [t + 1/2 - (t + 1/4)^{1/2}] \tag{3.20}$$

$$||\gamma t^{-1/2} e^{tA} A E_0 + e^{tA} A^2 E_0|| = 2 \left\{ \frac{t^{\delta-1}(\delta - t + \gamma\sqrt{t})}{\delta\,!} - \frac{t^{\eta-1}(\eta - t + \gamma\sqrt{t})}{\eta\,!} \right\} e^{-t}$$

$$(2/t\sqrt{2\pi})\{\zeta e^{-(1/2)\zeta^{-2}} + 1/\zeta e^{-(1/2)\zeta^2}\} \geq 4/(t\sqrt{2\pi e}) \quad \text{for} \quad t \to \infty \tag{3.21}$$

where:

$$\delta = [t - \varrho + (t + \varrho^2)^{1/2}] \quad \eta = [t - \varrho - (t + \varrho^2)^{1/2}]$$

with

$$\varrho = 1/2(\gamma\sqrt{t} - 1) \qquad \text{and} \qquad \zeta = \gamma/2 + (1 + \gamma^2/4)^{1/2}.$$

4. The optimal approximating Poisson random variable

In this section we will finally find the Poisson random variable which optimizes the approximation in the sense that it minimizes the expression of the total variation distance obtained in the third paragraph.

We will first observe that the asymptotic optimality may be reached for any $p = \sum_{i=1}^{n} p_i$. In fact, the value $\inf_{\mu} d\big(S_n, T_n(\mu)\big)$ is actually attained by the continuity of the $d\big(m, P_0(t)\big)$ in $t \geq 0$, for any measure $m \in M$.

Now, we will try to find which is the parameter μ asymptotically optimal in each case. This question will depend on the S_n random variable mean's size, as we are going to establish in the following theorems:

Theorem 3. Let $p_1, \ldots, p_n$ and $\lambda_i = -\log(1 - p_i)$ for $i = 1, \ldots, n$. If $0 < \lambda = \sum_{i=1}^{n} \lambda_i \leq 1$ then for all choice of μ_i, $i = 1, \ldots, n$ we have:

$$d(S_n, T_n) \geq \{\sum_{i=1}^{n} (e^{\lambda_i} - 1 - \lambda_i)\} \exp\,(-\sum_{i=1}^{n} \lambda_i) \geq 1/2\{\sum_{i=1}^{n} \lambda_1^2\} \exp\,(-\sum_{i=1}^{n} \lambda_i). \tag{4.1}$$

Consequently, if $p_1, \ldots, p_n$ depend on n such $\sum_{i=1}^{n} p_i \to 0$ for $n \to \infty$ then uniformly in n, we have:

$$\inf_{\mu} d\big(S_n, T_n(\mu)\big) \sim d\big(S_n, T_n(\lambda)\big) \sim 1/2 \sum_{i=1}^{n} p_1^2 \tag{4.2}$$

while

$$d\big(S_n, T_n(p)\big) \sim \sum_{i=1}^{n} p_i \quad \text{only} \tag{4.3}$$

where for $\mu = (\mu_1, \ldots, \mu_n)$, $T_n(\mu)$ denotes a Poisson random variable with expectation $\sum_{i=1}^{n} \mu_i$, and the inf. is taken over all allowable values of μ.

Proof: The expression of the total variation distance

$$d\big(S_n, T_n(\mu)\big) = 1/2 \sum_{k=0}^{\infty} |p(S_n = k) - p(T_n(\mu) = k)|$$

can be bounded by:

$$2d\big(S_n, T_n(\mu)\big) = |p(S_n = 0) - p\big(T_n(\mu) = 0\big)| + |p(S_n = 1) - p\big(T_n(\mu) = 1\big)| + \\ + |p(S_n \geq 2) - p\big(T_n(\mu) \geq 2\big)|.$$

Evaluating the probability terms that appear on the right side, we have:

$$|p(S_n = 0) - p\big(T_n(\mu) = 0\big)| = |e^{-\lambda} - e^{-\mu}| = e^{-\lambda}|1 - e^{-h}|$$

where $h = \mu - \lambda$

$$|p(S_n = 1) - p\big(T_n(\mu) = 1\big)| = e^{-\lambda}[|\sum_{i=1}^{n}(e^{\lambda_i} - 1 - \lambda_i) + \lambda - e^{-h}(h + \lambda)|]$$

$$|p(S_n \geq 2) - p\big(T_n(\mu) \geq 2\big)| = e^{-\lambda}|1 + \sum_{i=1}^{n} (e^{\lambda_i} - 1 - \lambda_i) + \\ + \lambda - e^{-h}(1 + h + \lambda)|.$$

Let $A(h)e^{-\lambda}$ be the sum of these three terms. Then:

$$2d\big(S_n, T_n(\mu)\big) \geq A(h)e^{-\lambda}.$$

The result now follows from the fact that:

a) When $h \geq 0$

$$1/2A(h) \geq (\lambda + 1) - e^{-h}(\lambda + h + 1) + \sum_{i=1}^{n} (e^{\lambda_i} - 1 - \lambda_i) \geq \\ \geq \sum_{i=1}^{n} (e^{\lambda_i} - 1 - \lambda_i) = 1/2A(0)$$

where the last inequality occurs because the function $(\lambda + 1) - e^{-h}(\lambda + h + 1)$ is always greater than zero.

b) When $h \leq 0$, using a similar argument:

$$1/2A(h) \geq \lambda - e^{-h}(\lambda + h) + \sum_{i=1}^{n} (e^{\lambda_i} - 1 - \lambda_i) \geq \sum_{i=1}^{n} (e^{\lambda_i} - 1 - \lambda_i) = 1/2A(0)$$

Finally: $d(S_n, T_n(\mu)) \geq e^{-\lambda} \sum_{i=1}^{n} (e^{\lambda_i} - 1 - \lambda_i)$.

The second inequality in (4.1) is obvious, for the function $f(x) = e^x - 1 - x - (1/2)x^2$ is always positive when $x > 0$.

In order to prove the relation (4.2), we will use the second inequality in (4.1). But as $\sum_{i=1}^{n} p_i \to 0$ then $\exp(-\sum_{i=1}^{n} \lambda_i) \to 1$.

Therefore:

$$d(S_n, T_n(\mu)) \geq 1/2 \sum_{i=1}^{n} \lambda_i^2 \exp(-\sum_{i=1}^{n} \lambda_i) \sim 1/2 \sum_{i=1}^{n} \lambda_i^2.$$

On the other hand:

$$d(S_n, T_n(\lambda)) \leq p(S_n \neq T_n(\lambda)) \leq 1 - \prod_{i=1}^{n} (1 - p(X_i \neq Y_i)) =$$

$$= 1 - \prod_{i=1}^{n} ((1 + \lambda_i)e^{-\lambda_i}) \leq \sum_{i=1}^{n} (e^{\lambda_i} - 1 - \lambda_i)e^{-\lambda_i} \leq 1/2 \sum_{i=1}^{n} \lambda_i^2.$$

Consequently:

$$\operatorname{Inf} d(S_n, T_n(\mu)) = d(S_n, T_n(\lambda)).$$

Since $\lambda_i - p_i = (1/2)p_i^2 + o(p_i^3)\ \forall i = 1, \ldots, n$ then asymptotically $\sum_{i=1}^{n} \lambda_i^2 \sim \sum_{i=1}^{n} p_i^2$ which establishes (4.2).

The expression (4.3) is obtained from (3.1), (3.2) and (3.3) for $\mu = p$.

This last result shows that whenever $p = \sum_{i=1}^{n} p_i$ tends to zero for $n \to \infty$, the choice $\mu = \lambda = \sum_{i=1}^{n} - \log(1 - p_i)$ is indeed asymptotically optimal.

However, if we assume that p tends to infinity in a certain way for $n \to \infty$, the optimal approximation Poisson random variable is that of mean $\mu = p$. In fact, we have:

Theorem 4. If $\sum_{i=1}^{n} p_i \to \infty$ and $\max\{p_1, \ldots, p_n\} \to 0$ for $n \to \infty$ then

$$d(S_n, T_n(p)) \sim (2\pi e)^{-1/2}\{\sum_{i=1}^{n} p_i^2\}/\{\sum_{i=1}^{n} p_i\}. \tag{4.4}$$

If additionally $\{\sum_{i=1}^{n} p_i\} \max(p_1, \ldots, p_n\} \to 0$ then also:

$$\operatorname*{Inf}_{\mu} d(S_n, T_n(\mu)) \sim d(S_n, T_n(p)). \tag{4.5}$$

Proof: The first relation (4.4) will be proved taking into account the theorems in section three. By the expression (3.1) for $\mu = p$, we get:

$$d(S_n, T_n(p)) = 1/4\{\sum_{i=1}^{n} p_i^2\}\|\exp((\sum_{i=1}^{n} p_i)A)A^2 E_0\| + r_n(p)$$

with $r_n(p) \xrightarrow[n\to\infty]{} 0$, which, using the expression (3.19), is asymptotically as:

$$d(S_n, T_n(p)) \sim (2\pi e)^{-1/2}\{\sum_{i=1}^{n} p_i^2\}/\{\sum_{i=1}^{n} p_i\}.$$

This proves relation (4.4).

If additionally $\{\sum_{i=1}^{n} p_i\} \max\{p_1, \ldots, p_n\} \to 0$ then

$$d(S_n, T_n(p)) \sim (2\pi e)^{-1/2}\{\sum_{i=1}^{n} p_i^2\}/\{\sum_{i=1}^{n} p_i\} \leq$$

$$\leq (2\pi e)^{-1/2} \max\{p_1, \ldots, p_n\} \xrightarrow[n\to\infty]{} 0.$$

On the other hand, for any optimal choice of μ we have:

$$d(T_n(\mu), T_n(p)) \leq d(S_n, T_n(\mu)) + d(S_n, T_n(p)) \leq 2d(S_n, T_n(p)) \to 0$$

under the assumptions considered, and since:

$$2d(T_n(\mu), T_n(p)) \geq |p(T_n(\mu) = 0) - p(T_n(p) = 0)| = |e^{-\mu} - e^{-p}|$$

we must have $\mu \sim p$. Then we can write that for any such μ, there exists some real number γ with $\mu \sim p + (\gamma/2)p^{-1}$.

So using the expressions (3.1), (3.2) and (3.3) for this μ we get:

$$d(S_n, T_n(\mu)) \sim 1/4\{\sum_{i=1}^{n} p_i^2\}\|(\gamma/\sum_{i=1}^{n} p_i)\exp((\sum_{i=1}^{n} p_i)A)A E_0 +$$

$$+ \exp((\sum_{i=1}^{n} p_i)A)A^2 E_0\|$$

because the terms $r_n(p)$ and $S_n(p, \mu)$ tend to zero as n tends to infinity. Now using the result (3.21) for $t = \sum_{i=1}^{n} p_i \to \infty$ we can conclude:

$$d(S_n, T_n(\mu)) \geq (2\pi e)^{-1/2}\{\sum_{i=1}^{n} p_i^2\}/\{\sum_{i=1}^{n} p_i\} \sim d(S_n, T_n(p)).$$

Therefore $\operatorname{Inf}_{\mu} d(S_n, T_n(\mu)) = d(S_n, T_n(p))$.

This choice corresponds to $\gamma = 0$. This proves theorem 4 completely.

5. Conclusions

In this paper we have shown two theorems allowing a precise evaluation of the total variation distance between n independent Bernoulli summands and a general Poisson distribution. The proofs of these theorems can be found in [5]. The main aim of this evaluation has been to find the mean of the Poisson random variable, which minimizes the expression obtained.

We have got it when the mean p of the random variable S_n tends to zero and to infinity.

Another open problem is to study the intermediate case in which the mean p tends to a real number $a \in (0, \infty)$.

References

1. BARBOUR, A. D. and HALL, P. (1984). "On the rate of Poisson convergence". Math. Proc. Camb. Phil. Soc. *95*, 473—480.
2. BUTZER, P. L. and BERENS, H. (1967). "Semi-Groups of Operators and Approximation". Springer, N.Y.
3. CHEN, L. H. Y. (1974). "On the convergence of Poisson binomial to Poisson distributions". Ann. Prob. *2*, 178—180.
4. LE CAM, L. (1960). "An approximation theorem for the Poisson binomial distribution". Pacific J. Math. *10*, 1181—1197.
5. GONZÁLEZ, J. M., MARTINEZ, M., CARRERA, A. R. (1987). "Application of Semi-Group Theory to the norms evaluation of probability measures". Actas de la Inter. Conf. on Linear Algebra and its Applic. (Valencia).

A. R. CARRERA
M. MARTINEZ
M. GONZALEZ
48013 Bilbao, Spain