A NEW SEMIGROUP TECHNIQUE IN POISSON APPROXIMATION

P. Deheuvels, D. Pfeifer and Madan L. Puri

ABSTRACT. We present a unified and self-contained approach to Poisson approximation problems for independent Bernoulli summands with respect to several metrics by a general semigroup technique, expanding and completing earlier work on this subject by the first two authors [4], [5], [6].

1. Introduction

We consider the approximation of the distribution of $S_n = \sum_{k=1}^{n} X_k$ where $X_1, \ldots, X_n$ are independent Bernoulli random variables with success probabilities $p_k = P(X_k = 1) = 1 - P(X_k = 0)$ by suitable Poisson distributions. As a measure of deviation, it is convenient to consider probability metrics over a set $\mathcal{M}$ of probability measures defined on a measurable space $(\Omega, \mathcal{A})$, for example

$$(1.1) \quad d(P,Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)| \qquad \text{(total variation)}$$

$$(1.2) \quad d_0(P,Q) = \sup_{x \in \mathbf{R}} |F_P(x) - F_Q(x)| \qquad \text{(uniform or Kolmogorov metric)}$$

AMS 1980 *subject classifications.* Primary 60F05, 62E20; Secondary 47D05, 20M30.

Key words and phrases. Poisson approximation, operator semigroups, Kolmogorov metric, Fortet-Mourier-Wasserstein metric, minimal metrics.

$$(1.3)\quad d_1(P,Q) = \inf\{E_L|X-Y| \mid \mathcal{L}(X,Y) = L,\ \mathcal{L}(X) = P,\ \mathcal{L}(Y) = Q\}$$
(Fortet-Mourier or Wasserstein metric).

Here F_P denotes the cumulative distribution function of the measure P, E_L means expectation with respect to the measure L, and $\mathcal{L}(X,Y)$ denotes the distribution of (X,Y). Note that the definitions of d_0 and d_1 make sense only in the case $(\Omega,\mathcal{A}) = (\mathbf{R},\mathcal{B})$ where $\mathcal{B}$ is the σ-field of Borel sets. Moreover, d and d_1 are special cases of minimal metrics d_ρ obtained via

$$(1.4)\quad d_\rho = \inf\{\rho(\mathcal{L}(X,Y)) \mid \mathcal{L}(X) = P,\ \mathcal{L}(Y) = Q\}$$

where ρ is a probability metric in the sense of Zolotarev ([21], [22]; see also Deheuvels, Karr, Pfeifer and Serfling [7]).

Until very recently, only estimates and expansions for $d(\mathcal{L}(S_n),\mathcal{P}(\mu))$ and $d_0(\mathcal{L}(S_n),\mathcal{P}(\mu))$ (where $\mathcal{P}(\mu)$ denotes a Poisson distribution with expectation $\mu > 0$) have been studied in the literature, of which the following ones are of specific importance:

$$(1.5)\quad d(\mathcal{L}(S_n),\mathcal{P}(\sum_{k=1}^n p_k)) \le \sum_{k=1}^n p_k^2 \qquad \text{(Le Cam [12])}$$

$$(1.6)\quad d(\mathcal{L}(S_n),\mathcal{P}(\sum_{k=1}^n \nu_k)) \le \tfrac{1}{2}\sum_{k=1}^n \nu_k^2 \qquad \text{(Serfling [18], [19])}$$
where $\nu_k = -\log(1-p_k),\ 1 \le k \le n$,

$$(1.7)\quad d(\mathcal{L}(S_n),\mathcal{P}(\sum_{k=1}^n p_k)) \le (1 - exp(-\sum_{k=1}^n p_k))\theta_n \qquad \text{(Barbour and Hall [2])}$$
where $\theta_n = \sum_{k=1}^n p_k^2 / \sum_{k=1}^n p_k$,

$$(1.8)\quad d_0(\mathcal{L}(S_n),\mathcal{P}(\sum_{k=1}^n p_k)) \le \tfrac{1}{2}\sum_{k=1}^n p_k^2 \qquad \text{(Daley; cf. Serfling [19])}$$

$$(1.9)\quad d_0(\mathcal{L}(S_n),\mathcal{P}(\sum_{k=1}^n \nu_k)) \le \tfrac{1}{2}\sum_{k=1}^n \nu_k^2 \qquad \text{(Serfling [19])}$$

$$(1.10)\quad d_0(\mathcal{L}(S_n),\mathcal{P}(\sum_{k=1}^n p_k)) \le \tfrac{\pi}{4}(\sum_{k=1}^n p_k^2/(1-p_k))/(\sum_{k=1}^n p_k(1-p_k))$$
(Hipp [9])

$$(1.11)\quad d_0(\mathcal{L}(S_n),\mathcal{P}(\sum_{k=1}^n p_k)) \le (\tfrac{1}{2} + \sqrt{\tfrac{\pi}{8}})\theta_n/(1-\sqrt{\theta_n}) \qquad \text{(Shorgin [20]).}$$

Although most of these authors consider $\mu = \sum_{k=1}^n p_k = E(S_n)$ as the favorable choice for an approximation, it is noteworthy that for $n = 1$, Serfling's [18] choice for d is actually the best possible, i.e.

(1.12) $d(\mathcal{L}(S_1), \mathcal{P}(\nu_1)) = \min_\mu d(\mathcal{L}(S_1), \mathcal{P}(\mu)) = 1 - (1+\nu_1)e^{-\nu_1} \le \frac{1}{2}\nu_1^2$

whenever $p_1 \le 1 - 1/e$.

Although it was not explicitly stated in Serfling's paper [19], it is easy to see that a corresponding optimization procedure for d_0 would yield

(1.13) $\min_\mu d_0(\mathcal{L}(S_1), \mathcal{P}(\mu)) = d_0(\mathcal{L}(S_1), \mathcal{P}(\overline{p})) = 1 - (1+\overline{p})e^{-\overline{p}}$

where $\overline{p}$ is the unique positive solution of the equation

(1.14) $p_1 = 2 - 2e^{-\overline{p}} - \overline{p}e^{-\overline{p}}$.

Some further analysis shows in fact that

(1.15) $p_1 < p_1 + \frac{1}{6}p_1^3 - \frac{1}{12}p_1^4 \le \overline{p} \le p_1 + \frac{1}{6}p_1^3 < \nu_1$

which in turn gives rise to the estimate

(1.16) $d_0(\mathcal{L}(S_n), \mathcal{P}(\sum_{k=1}^n \overline{p}_k)) \le \frac{1}{2}\sum_{k=1}^n p_k^2(1 - \frac{1}{3}p_k + \frac{1}{2}p_k^2 + \frac{1}{36}p_k^4)$

where $\overline{p}_k$ is the corresponding solution of (1.14) when p_1 is replaced by p_k. Note that when for example $max(p_1, \ldots, p_n) \le \frac{1}{2}$, then the right hand side of (1.6) is actually smaller than the right hand side of (1.8).

Similarly, it can easily be shown that the optimal $\mu = \widetilde{p}$ for minimizing $d_1(\mathcal{L}(S_1), \mathcal{P}(\mu))$ with respect to μ is given by

(1.17) $\widetilde{p} = \min(\nu_1, \log 2)$

with a corresponding distance given by

(1.18) $d_1(\mathcal{L}(S_1), \mathcal{P}(\overline{p})) = \begin{cases} \nu_1 - p_1, & p_1 \le \frac{1}{2} \\ \log 2 - 1 + p_1, & p_1 > \frac{1}{2} \end{cases} = \Delta(p_1)$, say.

Let (X_k, Y_k) be a coupling such that $E|X_k - Y_k| = d_1(\mathcal{L}(X_k), \mathcal{L}(Y_k)) = d_1(\mathcal{L}(X_k), \mathcal{P}(\widetilde{p}_k))$, $1 \le k \le n$ (which always exists). Then

$$\begin{aligned} d_1(\mathcal{L}(S_n), \mathcal{L}(\sum_{k=1}^n Y_k)) &\le E|\sum_{k=1}^n X_k - \sum_{k=1}^n Y_k| \le \sum_{k=1}^n E|X_k - Y_k| \\ &\le \sum_{k=1}^n d_1(\mathcal{L}(X_k), \mathcal{L}(Y_k)), \end{aligned}$$

or

$$(1.19)\quad d_1(\mathcal{L}(S_n), \mathcal{P}(\sum_{k=1}^{n} \tilde{p}_k)) \le \sum_{k=1}^{n} \Delta(p_k).$$

Note that by (1.18), $\Delta(p_k)$ is always larger than $\frac{1}{2}p_k^2$, but bounded by $\frac{1}{2}p_k^2/(1-p_k)$, for $p_k < 1$.

The last examples show that we cannot hope for an explicit closed solution for the problem of the best choice for the Poisson parameter, not even for commonly used distance measures. However, it was possible to approach the problem at least asymptotically (cf. Deheuvels and Pfeifer [4], [5], [6]) by a successively refined application of convolution semigroups on different Banach spaces, in the spirit of Le Cam's [12] paper which at the same time also allowed for asymptotic expansions of the metrics under consideration with essentially arbitrary choices of μ. Unfortunately, the 'telescoping technique' used there gave only results in the range where $\sum_{k=1}^{n} p_k^2 = o(1)$ or $\sum_{k=1}^{n} p_k^2 = 0(1)$ and $\sum_{k=1}^{n} p_k$ unbounded, whereas some of the above estimates show that Poisson convergence with respect to d or d_0 takes place as long as $\theta_n = \sum_{k=1}^{n} p_k^2 / \sum_{k=1}^{n} p_k = o(1)$ which is a weaker condition. In the present paper we present a new approach in the semigroup setting which covers also this case, and at the same time provides in a unified and self-contained fashion, upper and lower bounds for the metrics under consideration, which are sometimes considerably sharper than all the competitors outlined in (1.5) to (1.11). A particular advantage of this technique is also the fact that we do not need any longer different methods for different metrics, such as characteristic functions (Shorgin [20], Hipp [9]), couplings (Serfling [18], [19]) or other specific approaches such as Stein's method (Barbour and Hall [2], Barbour [1]), and that we are also able to present at least nearly optimal solutions to the problem of the best parameter choice for the approximating Poisson distribution.

2. The Semigroup Setting

Let $\mathcal{X}$ denote either the Banach space ℓ^1 of summable sequences $f = (f(0), f(1), \ldots)$ or the Banach space ℓ^∞ of all absolutely bounded sequences. Define the shift operator $B: \mathcal{X} \to \mathcal{X}$ by

$$(2.1)\quad Bf = (0, f), \text{ i.e. } Bf(n) = \begin{cases} 0 & \text{if } n = 0 \\ f(n-1) & \text{if } n \ge 1. \end{cases}$$

Actually B corresponds to a convolution of the unit mass ε_1 concentrated at 1 with f (cf. Deheuvels and Pfeifer [4], [5]). Then, the operator $A = B - I$ (I denotes the identity) generates the Poisson convolution semigroup

$$(2.2)\quad e^{tA}f = e^{-t}e^{tB}f = e^{-t}\sum_{k=0}^{\infty} \tfrac{t^k}{k!}B^k f = \mathcal{P}(t) * f,\ f \in \mathcal{X}$$

where $*$ means convolution. Similarly,

(2.3) $(I + p_k A)f = ((1 - p_k)I + p_k B)f = \mathcal{L}(X_k)^* f,\ f \in \mathcal{X}.$

Hence, by independence,

(2.4) $\sum_{k=1}^{n} (I + p_k A)f = \mathcal{L}(S_n)^* f,\ f \in \mathcal{X}.$

Then, we have the following relationship between such semigroups and the metrics outlined in the introduction.

Lemma 2.1. *Let* $\|\cdot\|_{(\mathcal{X},\mathcal{X})}$ *denote the corresponding operator norm. Then*

(2.5) $d(\mathcal{L}(S_n), \mathcal{P}(\mu)) = \frac{1}{2}\|e^{\mu A} - \prod_{k=1}^{N} (I + p_k A)\|_{(\ell^1,\ell^1)}$

(2.6) $d_1(\mathcal{L}(S_n), \mathcal{P}(\mu)) = \|\{e^{\mu A} - \prod_{k=1}^{N} (I + p_k A)\}A^{-1}\|_{(\ell^1,\ell^1)}$

Proof. Relation (2.5) is essentially given in Deheuvels and Pfeifer [4]; see also Deheuvels and Pfeifer [6]. To prove (2.6), observe that with $f_0 = (1, 0, 0, \cdots)$ we have $\|f_0\|_{\ell^1} = 1$, $A^{-1}f_0 = (1, 1, 1 \ldots)$, and $d_1(\mathcal{L}(S_n), \mathcal{P}(\mu)) = \|\{e^{\mu A} - \prod_{k=1}^{N} (I + p_k A)\}(A^{-1}f_0)\|_{\ell^1}$ (Deheuvels and Pfeifer [5]) from which the assertion follows. (Note that $-A^{-1}$ is also called the potential operator associated with the semigroup.)

Unfortunately, the corresponding d_0-metric cannot be expressed directly by an operator norm; however, it is valid that

(2.7) $d_0(\mathcal{L}(S_n), \mathcal{P}(\mu)) = \|\{e^{\mu A} - \prod_{k=1}^{N} (I + p_k A)\}(A^{-1}f_0)\|_{\ell^\infty}$

(Deheuvels and Pfeifer [5]), which will also be sufficient for our purposes.

The following theorem is a key result for a unified treatment of estimates and expansions for the expressions above.

Theorem 2.1. *Let* $\lambda = \sum_{k=1}^{n} p_k$, $\lambda_m = \sum_{k=1}^{n} p_k^m$, $m \geq 2$. *Then, for an arbitrary operator semigroup* $\{e^{tA},\ t \geq 0\}$ *with bounded generator* A,

(2.8) $e^{\lambda A} - \prod_{k=1}^{n} (I + p_k A) = \frac{1}{2}\lambda_2 A^2 e^{\lambda A} - \sum_{k=3}^{\infty} a_k (-A)^k e^{\lambda A}$

where a_k *is recursively defined by*

(2.9) $a_k = -\frac{1}{k}(\lambda_k + \sum_{i=2}^{k-2} a_i\lambda_{k-i}),\ k \geq 3,$ *and* $a_2 = -\frac{1}{2}\lambda_2.$

The convergence in (2.8) is absolute and to be understood in the uniform operator topology. Likewise

(2.10) $\{e^{\lambda A} - \prod_{k=1}^{n}(I + p_k A)\}A^{-1} = \frac{1}{2}\lambda_2 A e^A + \sum_{k=3}^{\infty} a_k(-A)^{k-1}e^{\lambda A}.$

Proof. By Lemma 3 of Shorgin [20], it follows that

(2.11) $\prod_{k=1}^{n}(1 + p_k z)e^{-p_k z} = 1 - \frac{1}{2}\lambda_2 z^2 + \sum_{k=3}^{\infty} a_k(-z)^k,$

where the right hand side converges absolutely for every real z since by Lemma 5 of Shorgin [20], we have

(2.12) $|a_k| \leq (e\lambda_2/k)^{k/2} \leq |2z|^{-k}$ for sufficiently large k $(z \neq 0)$.

Thus, we have

(2.13) $\prod_{k=1}^{n}(1 + p_k A)e^{-p_k A} = 1 - \frac{1}{2}\lambda_2 A^2 + \sum_{k=3}^{\infty} a_k(-A)^k$

where the right hand side coverges absolutely, from which (2.8) follows by application of $e^{\lambda A}$. (2.10) is proved similarly.

It should be pointed out that Theorem 2.1 also allows for a full expansion related to the Charlier-Poisson expansion in Barbour [1] since e.g. for the metric d, we have

(2.14) $\| \sum_{j=2}^{\infty} a_j(-A)^j e^{\lambda A}\|_{(\ell^1,\ell^1)} = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}|\sum_{j=2}^{\infty}(-1)^j a_j \sum_{i=0}^{j}(-1)^i\binom{j}{i}\frac{k_{(i)}}{\lambda^i}|$

where $k_{(i)} = \prod_{m=0}^{i-1}(k - m)$; similar ideas lead to the expansions for the metrics d_0 and d_1 even in the more general setting of this paper.

We should like to mention that Theorem 1 also applies to the more general Banach space setting used in connection with minimal metrics described in Deheuvels, Karr, Pfeifer and Serfling [7]. It was shown there that such metrics can be estimated by

(2.15) $d_\alpha(\mathcal{L}(S_n), \mathcal{P}(\mu)) = \sum_{k=0}^{\infty} \alpha_k |P(S_n = k) - \mathcal{P}(\mu; k)|$

where $\{\alpha_k\}$ is a non-negative sequence which does not grow faster than at a geometric rate, i.e. $\sup_{k\geq 1}(\alpha_{k+1}/\alpha_k) = M < \infty$. Note that d_α is itself a metric for all such α, and that d_α can be represented as

$$(2.16)\quad d_\alpha(\mathcal{L}(S_n), \mathcal{P}(\mu)) = \|\{e^{\mu A} - \prod_{k=1}^{n}(I + p_k A)\} f_0\|_{\ell^1_\alpha}$$

where ℓ^1_α is the Banach space of all α-summable sequences with norm

$$(2.17)\quad \|f\|_{\ell^1_\alpha} = \sum_{k=0}^{\infty} \alpha_k |f(k)|.$$

Then $\|A\|_{(\ell^1_\alpha),(\ell^1_\alpha)} \leq 1 + M$, and (2.8) also holds for this case. Note that d_α-metrics have also been considered earlier by Johnson and Simons [10], however, for the case of equal p_k's only. The next result will provide a tool for estimating some distances in the case of $\mu = \lambda$.

Lemma 2.2. *Suppose there exists a constant* $K > 0$ *such that*

$$(2.18)\qquad \|Ae^{\lambda A}\|_{(\mathcal{X},\mathcal{X})} \leq \sqrt{K/e\lambda}.$$

Then

$$(2.19)\qquad \|e^{\lambda A} - \prod_{k=1}^{n}(I + p_k A)\|_{(\mathcal{X},\mathcal{X})} \leq \frac{K\theta}{1 - \sqrt{K\theta}}$$

provided $\theta = \theta_n < \frac{1}{k}$.

Proof. By (2.18),

$$(2.20)\qquad \|A^k e^{\lambda A}\|_{(\mathcal{X},\mathcal{X})} \leq \|Ae^{(\lambda/k)A}\|^k_{(\mathcal{X},\mathcal{X})} \leq \sqrt{Kk/e\lambda}^{\,k},$$

hence by (2.8) and (2.12),

$$\|e^{\lambda A} - \prod_{k=1}^{n}(I + p_k A)\|_{(\mathcal{X},\mathcal{X})} \leq \sum_{k=2}^{\infty} |a_k|\ \|A^k e^{\lambda A}\|_{(\mathcal{X},\mathcal{X})} \leq \sum_{k=2}^{\infty} \sqrt{K\theta}^{\,k}$$

which proves the result.

Note that in a similar way we obtain

$$(2.21)\qquad |\{e^{\lambda A} = \prod_{k=1}^{n}(I + p_k A)\} A^{-1}\|_{(\ell^1,\ell^1)} \leq \sqrt{e\lambda}\frac{K\theta}{1 - \sqrt{K\theta}},\ \theta < \frac{1}{K}.$$

As has been shown in Deheuvels and Pfeifer [4], [5] we have

$$(2.22)\qquad \|Ae^{\lambda A}\|_{(\ell^1,\ell^1)} = 2e^{-\lambda}\frac{\lambda^{[\lambda]}}{[\lambda]!} \leq \sqrt{2/e\lambda}$$

with equality reached for $\lambda = \frac{1}{2}$, hence we may choose $K = 2$ in Lemma 2.2. Here $[\lambda]$ denotes the integral part of λ. Thus we obtain:

Corollary 2.1.

$$d(\mathcal{L}(S_n), \mathcal{P}(\lambda)) \leq \frac{\theta}{1-\sqrt{2\theta}}, \ \theta < \frac{1}{2}, \tag{2.23}$$

$$d_1(\mathcal{L}(S_n), \mathcal{P}(\lambda)) \leq \frac{4\sqrt{\lambda}\theta}{1-\sqrt{2\theta}}, \ \theta < \frac{1}{2}. \tag{2.24}$$

Better estimations are obtained if the leading term $\frac{\lambda_2}{2}\|A^2 e^{\lambda A}\|_{(\mathcal{X},\mathcal{X})}$ is considered separately; since it is completely known in closed form (see Deheuvels and Pfeifer [4], [5]) one can easily derive upper and lower bounds by estimating the remainder terms in the series expansion (2.8). By application of some of Shorgin's [20] auxiliary results in the case of d_0, one obtains, for instance,

Corollary 2.2.

$$|d(\mathcal{L}(S_n), \mathcal{P}(\lambda)) - \frac{\lambda_2}{2}\, e^{-\lambda}\{\frac{\lambda^{a-1}(a-\lambda)}{a!} + \frac{\lambda^{b-1}(\lambda-b)}{b!}\}| \leq \frac{(2\theta)^{3/2}}{2(1-\sqrt{2\theta})}, \tag{2.25}$$

$$\theta < 1/2,$$

where $a = [\lambda + 1/2 + \sqrt{\lambda + 1/4}]$, $b = [\lambda + 1/2 - \sqrt{\lambda + 1/4}]$;

$$|d_o(\mathcal{L}(S_n), \mathcal{P}(\lambda)) - \frac{\lambda_2}{2}\, e^{-\lambda} \max\{\frac{\lambda^{a-1}(a-\lambda)}{a!}, \frac{\lambda^{b-1}(\lambda-b)}{b!}\}| \leq \frac{C\theta^{3/2}}{1-\sqrt{\theta}}, \tag{2.26}$$

$$\theta < 1;$$

$$|d_1(\mathcal{L}(S_n), \mathcal{P}(\lambda)) - \lambda_2\, e^{-\lambda}\frac{\lambda^{[\lambda]}}{[\lambda]!}| \leq \frac{2\sqrt{\lambda}(2\theta)^{3/2}}{1-\sqrt{2\theta}}, \ \theta < 1/2. \tag{2.27}$$

Here we have used $C = \frac{1}{2} + \sqrt{\frac{\pi}{8}} < 1.13$.

This shows clearly that the condition $\theta \to 0$ is necessary and sufficient for Poisson convergence with $\mu = \lambda$ for the metrics d and d_o, and $\sqrt{\lambda}\theta \to 0$ for the metric d_1. The bounds obtained in Corollary 2.2 are usually much sharper than all the competitors from (1.5) to (1.11). This will be demonstrated in the next section.

In order to attack the problem of the (nearly) best Poisson parameter not only in an asymptotic setting as in Deheuvels and Pfeifer ([4], [5]) we proceed as follows. Instead of using the parameter $\mu = \lambda$ we make an adjustment of the form

$$\mu = \lambda + \delta\lambda_2 \ \text{ for some } \ 0 \leq \delta \leq 1/2. \tag{2.28}$$

Then

$$e^{\mu A}-\prod_{k=1}^{n}(I+p_kA)=\frac{\lambda_2}{2}(A^2+2\delta A)e^{\lambda A}+\frac{\delta^2\lambda_2^2}{2}A^2e^{\lambda A}+\sum_{k=3}^{\infty}\{\frac{(\delta\lambda_2)^k}{k!}-a_k(-1)^k\}A^ke^{\lambda A} \tag{2.29}$$

which is obvious from $e^{\mu A}-e^{\lambda A}=e^{\lambda A}(e^{\delta\lambda_2 A}-I)$ and the Taylor expansion for the second factor. Since the leading term in the above expansion will be the dominating one in general, a suboptimal solution is obtained by choosing δ such that $\|(A^2+2\delta A)e^{\lambda A}\|_{(\mathcal{X},\mathcal{X})}$ is minimal, paralleling the selection procedure in the asymptotic case. Since also the last norm terms can explicitly be given, the determination of δ is mainly a numerical problem. For these we shall restate Example 3.1 in Deheuvels and Pfeifer [5] giving the minimizing δ for each case. In what follows, $L(\mu)$ denotes the resulting leading term for the metric obtained from (2.29) and Lemma 2.1.

Lemma 2.3. *I. For the metric* d, *we have*

$$\delta=1/2,\ L(\mu)=\frac{1}{2}\lambda_2e^{-\lambda}\ \text{ if }\ 0<\lambda<1, \tag{2.30}$$

$$\delta=1/2,\ L(\mu)=\frac{1}{2}\lambda_2\lambda e^{-\lambda}\ \text{ if }\ 1<\lambda<\sqrt{2}, \tag{2.31}$$

$$\delta=\frac{1}{2}-\frac{3}{2\lambda}\cdot\frac{2-\lambda}{3-\lambda},\ L(\mu)=\frac{1}{2}\lambda_2e^{-\lambda}\{\lambda+(1-\frac{\lambda^2}{2})\frac{3}{\lambda}\cdot\frac{2-\lambda}{3-\lambda}\}, \tag{2.32}$$

$$\delta=0,\ L(\mu)=\frac{1}{2}\lambda_2e^{-\lambda}\{\frac{\lambda^2}{2}+(1-\frac{\lambda^3}{6})\}\ \text{ if }\ \sqrt[3]{6}<\lambda\le\sqrt{2}. \tag{2.33}$$

II. For the metric d_0, *we have*

$$\delta=\frac{1}{2}-\frac{1}{2(1+\lambda)},\ L(\mu)=\frac{\lambda_2e^{-\lambda}}{2(1+\lambda)}\ \text{ if }\ 0<\lambda\le\sqrt{3}-1, \tag{2.34}$$

$$\delta=\frac{1}{2}-\frac{\lambda}{2+\lambda^2},\ L(\mu)=\frac{\lambda_2\lambda e^{-\lambda}}{(2+\lambda^2)}\ \text{ if }\ \sqrt{3}-1<\lambda\le 1. \tag{2.35}$$

III. For the metric d_1, *we have*

$$\delta=\frac{1}{2},\ L(\mu)=\lambda_2/2\ \text{ if }\ 0<\lambda<\log 2, \tag{2.36}$$

$$\delta=0,\ L(\mu)=\lambda_2e^{-\lambda}\ \text{ if }\ \log 2<\lambda\le 1, \tag{2.37}$$

$$\delta=\frac{1}{2}-\frac{1}{2\lambda},\ L(\mu)=\frac{1}{2}\lambda_2\{1-\frac{1}{\lambda}(1-2e^{-\lambda})\}\ \text{ if }\ 1<\lambda\le\alpha \tag{2.38}$$

where α *is the positive root of the equation* $2e^{-\alpha}(1+\alpha)=1$,

$$\delta=0,\ L(\mu)=\lambda_2\lambda e^{-\lambda}\ \text{ if }\ \alpha<\lambda\le 2. \tag{2.39}$$

If $\delta>0$ *is an arbitrary choice, and if again* $L(\mu)$ *denotes the resulting leading term, then (2.29) leads to the following:*

Corollary 2.3.

$$(2.40)\qquad |d(\mathcal{L}(S_n),\mathcal{P}(\mu))-L(\mu)|\le\frac{\delta^2\lambda_2^2}{2e\lambda}+\frac{1}{2}\left|\frac{\delta^3\lambda_2^3}{6}-\frac{\lambda_3}{3}\right|\left(\frac{6}{e\lambda}\right)^{3/2}+\frac{2\theta^2}{\sqrt{1-2\theta}}+8\frac{\delta^4\lambda_2^4}{e^2\lambda^2}e^{2\delta\lambda_2},\ \theta<1/2;$$

$$(2.41)\qquad |d_o(\mathcal{L}(S_n),\mathcal{P}(\mu))-L(\mu)|\le\frac{\delta^2\lambda_2^2}{2e\lambda}+\left|\frac{\delta^3\lambda_2^3}{6}-\frac{\lambda_3}{3}\right|\left(\frac{3}{2\lambda}\right)^{3/2}+\frac{C\theta^2}{1-\sqrt{\theta}}+\frac{4\delta^4\lambda_2^4}{e^2\lambda^2}e^{2\delta\lambda_2},\ \theta<1;$$

$$(2.42)\qquad |d_1(\mathcal{L}(S_n),\mathcal{P}(\mu))-L(\mu)|\le\frac{\delta^2\lambda_2^2}{\sqrt{2e}\lambda}+\left|\frac{\delta^3\lambda_2^3}{6}-\frac{\lambda_3}{3}\right|\frac{4}{e\lambda}+\sqrt{e\lambda_2}\frac{(2\theta)^{3/2}}{\sqrt{1-2\theta}}+\delta^4\lambda_2^4\sqrt{6/e}\ e^{2\delta\lambda_2},\quad \theta<1/2.$$

In a similar manner, estimates and expansions can be obtained for the d_α-metrics which for lack of space we omit here.

Of course, evaluations of the above distances can be made up to an arbitrary precision using more terms from the full expansion given in Theorem 2.1, and a corresponding remainder treatment as in Lemma 2.2 or Corollary 2.3, together with (2.29). The advantage of the one-term procedure, however, relies on the fact that the expressions involved are much more easy to compute.

3. Applications and Numerical Comparisons

In order to show the power of the approach introduced above, we shall briefly investigate a problem stemming from the theory of extremes. Let $\{Z_n, n\ge 1\}$ be a sequence of independent and identically distributed random variables with continuous distribution function F. Set $Z_{(n)}=\max(Z_1,\dots,Z_n)$, and

$$(3.1)\qquad X_n=\begin{cases}1 & \text{if } Z_{(n)}>Z_{(n-1)}\\ 0 & \text{otherwise}\end{cases}$$

for $n\ge 2$ (by convention $X_1\equiv 1$).

We say that $Z_{(n)}$ is a *record value* of $\{Z_n, n\ge 1\}$ if for the corresponding n we have $X_n=1$. By a well known result of Rényi [16] we know that $\{X_n\}$ is an independent sequence with $p_k=\frac{1}{k}$, $k\in\mathbf{N}$. Hence $S_n(m)=\sum\limits_{k=n+1}^{nm}X_k$, the number of record values between observation $n+1$ and nm, is approximately Poisson-distributed with $E(S_n(m)) = \sum_{k=n+1}^{nm} \frac{1}{k} \sim \log \frac{n \cdot m}{n} = \log m$, for large n. The following table provides a comparison between the different expansions and estimates outlined above, for the Poisson approximation of $\mathcal{L}(S_n(m))$.

Figure 1

Metric d, $n = 10$, $m = 5$ $\log m = 1.6094$

μ	δ	$L(\mu)$	lower bound	upper bound	authors
		.0124	.0020	.0228	Deheuvels, Pfeifer, Puri
1.5702	0	--	.0015	.0380	Barbour and Hall
		--	--	.0754	LeCam
		.0124	.0017	.0229	Deheuvels, Pfeifer, Puri
1.6094	.5199	--	--	.0398	Serfling
1.5863	.2129	.0092	--	.0196	Deheuvels, Pfeifer, Puri

Figure 2

Metric d_o, $n = 10$, $m = 2$ $\log m = .6931$

μ	δ	$L(\mu)$	lower bound	upper bound	authors
		.0119	.0038	.0200	Deheuvels, Pfeifer, Puri
.6688	0	--	--	.0232	Daley
		--	--	.0631	Hipp
		--	--	.1060	Shorgin
		.0119	.0037	.0201	Deheuvels, Pfeifer, Puri
.6931	.5247	--	--	.0246	Serfling
.6780	.2004	.0071	.0011	.0142	Deheuvels, Pfeifer, Puri

Note that in last line of each figure, δ is chosen to minimize the corresponding leading term from (2.29).

We should like to point out that with the former semigroup approach developed in Deheuvels and Pfeifer ([4], [5], [6]), no reasonable upper and lower bounds are to be obtained for this case, for the metrics d_0 and d_1.

Further situations in which Poisson approximation problems of such kind (i.e. with varying success probabilities p_k) occur are to be found in Ross [17] (average-case analysis for the simplex method in Linear Programming), Kemp [11] (average-case analysis of search algorithms), Gastwirth and Bhattacharya [8] (chain letter systems, multilevel marketing systems), Bruss [3] (secretary problems), Nevzorov [13] (generalized record problems), Pfeifer [14] (extremal processes), and Pfeifer [15] (best choice problems, non-homogeneous extremal processes). The results developed in this paper are especially suited to provide a unified treatment of the approximation problems considered in these papers with a strong emphasis on numerical tractability.

Acknowledgements

This paper was written while the second author was visiting the Department of Mathematics at Indiana University, Bloomington. He is grateful to the Deutsche Forschungsgemeinschaft for financial support by the Heisenberg Research Foundation.

We are also indebted to Dr. H. J. Witte for pointing out and correcting an error in an earlier version of the paper.

References

[1] Barbour, A. D. (1987), *Asymptotic expansions in the Poisson limit theorem*, Ann. Prob. **15**, 748-766.

[2] Barbour, A. D. and Hall, P. (1984), *On the rate of Poisson convergence*, Math. Proc. Cambridge Philos. Soc. **95**, 473-480.

[3] Bruss, F. T. (1984), *Patterns of relative maxima in random permutations*, Ann. Soc. Sci. Bruxelles **98 (I)**, 19-28.

[4] Deheuvels, P. and Pfeifer, D. (1986 a), *A semigroup approach to Poisson approximation*, Ann. Prob. **14**, 663-676.

[5] Deheuvels, P. and Pfeifer, D. (1986 b), *Operator semigroups and Poisson convergence in selected metrics*, Semigroup Forum **34**, 203-224.

[6] Deheuvels, P. and Pfeifer, D. (1987), *Semigroups and Poisson approximation. In: New Perspectives in Theoretical and Applied Statistics*, M. L. Puri, J. P. Vilaplana, W. Wertz (eds.), John Wiley, N.Y., 439-448.

[7] Deheuvels, P., Karr, A. F., Pfeifer, D. and Serfling, R. J. (1988), *Poisson approximations in selected metrics by coupling and semigroup methods with applications.* To appear in: J. Stat. Planning Inf.

[8] Gastwirth, J. L. and Bhattacharya, P. K. (1984), *Probability models of pyramid or chain letter systems demonstrating that their promotional claims are unreliable*, Operat. Res. **32**, 527-536.

[9] Hipp, C. (1985), *Approximation of aggregate claims distributions by compound Poisson distributions.* Insurance: Math. and Econ. **4**, 227-232.

[10] Johnson, N. L. and Simons, G. (1971), *On the convergence of binomial to Poisson distributions*, Ann. Math. Stat. **42**, 1735-1736.

[11] Kemp, R. (1984), *Fundamentals of the Average Case Analysis of Particular Algorithms*, Wiley-Teubner, N.Y.

[12] Le Cam, L. (1960), *An approximation theorem for the Poisson binomial distribution*, Pacific J. Math. **10**, 1181-1197.

[13] Nevzorov, V. B. (1986), *Two characterizations using records. In: Stability Problems for Stochastic Models*, Lecture Notes in Math. **1233**, Springer, Berlin, 79-85.

[14] Pfeifer, D. (1986), *Extremal processes, record times and strong approximation*, Pub. Inst. Stat. Univ. Paris **31**, 47-65.

[15] Pfeifer, D. (1988), *Extremal processes, secretary problems and the 1/e-law.* To appear in: J. Appl. Prob.

[16] Rényi, A. (1962), *Théorie des éléments saillants d'une suite d'observations*, Coll. Comb. Meth. Prob. Th. **104-115**, Mathematisk Institut, Aarhus Universitet, Denmark.

[17] Ross, S. M. (1982), *A simple heuristic approach to simplex efficiency*, Eur. J. Oper. Res. **9**, 344-346.

[18] Serfling, R. J. (1975), *A general Poisson approximation theorem*, Ann. Prob. **3**, 726-731.

[19] Serfling, R. J. (1978), *Some elementary results on Poisson approximation in a sequence of Bernoulli trials*, SIAM Rev. **20**, 567-579.

[20] Shorgin, S. Y. (1977), *Approximation of a generalized binomial distribution*, Th. Prob. Appl. **22**, 846-850.

[21] Zolotarev, V. M. (1976), *Metric distances in spaces of random variables and their distributions*, Math. U.S.S.R. Sbornik **30**, 3, 373-401.

[22] Zolotarev, V. M. (1984), *Probability metrics*, Th. Prob. Appl. **28**, 278-302.