



# Deviation of discrete distributions—positive and negative results

Tamás F. Móri

## ► To cite this version:

Tamás F. Móri. Deviation of discrete distributions—positive and negative results. *Statistics and Probability Letters*, 2009, 79 (8), pp.1089. 10.1016/j.spl.2008.12.015 . hal-00516883

**HAL Id: hal-00516883**

**<https://hal.science/hal-00516883>**

Submitted on 13 Sep 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

Deviation of discrete distributions – positive and negative results

Tamás F. Móri

PII: S0167-7152(08)00579-8  
DOI: [10.1016/j.spl.2008.12.015](https://doi.org/10.1016/j.spl.2008.12.015)  
Reference: STAPRO 5313

To appear in: *Statistics and Probability Letters*

Received date: 5 September 2007

Revised date: 28 October 2008

Accepted date: 12 December 2008

Please cite this article as: Móri, T.F., Deviation of discrete distributions – positive and negative results. *Statistics and Probability Letters* (2008), doi:10.1016/j.spl.2008.12.015

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



# Deviation of discrete distributions – positive and negative results<sup>1</sup>

Tamás F. Móri

*Department of Probability Theory and Statistics, Eötvös Loránd University,  
Pázmány P. s. 1/C, H-1117 Budapest, Hungary*

---

## Abstract

The paper is devoted to the problem of estimating the deviation of two discrete probability distributions in terms of the supremum distance between their generating functions over the interval  $[0, 1]$ . The deviation can be measured by the difference of the  $k$ th terms, or by total variation distance. In addition to upper bounds we illustrate the limitations of such estimations by counterexamples. Such problems arise e.g. when Poisson limit results are proved by sieve methods.

*Key words:* Poisson approximation, Rényi sieve, Bonferroni inequalities, generating function, spread polynomials

*1991 MSC:* 60E15

---

## 1 Poisson approximation by sieve methods

In discrete probability there are naturally arising situations where given are two discrete probability distributions, and one can derive estimates for the difference of the corresponding generating functions over the interval  $[0, 1]$ . Then the problem is, how those estimates can be transformed into estimations for the difference of probabilities.

For example, consider a finite collection of possibly dependent random events  $A_1, A_2, \dots, A_n$ . Let  $N$  denote the (random) number of events that occur. In many cases  $N$  is approximately Poisson distributed, particularly, when the number of events is large, their probabilities are small, and the dependence is weak. This can be quantified in several ways, like sieve methods or Chen–Stein

---

*Email address:* `moritamas@ludens.elte.hu` (Tamás F. Móri).

<sup>1</sup> Supported by the Hungarian Scientific Research Fund (OTKA) Grant K67961

approximation. A great advantage of the latter is that it provides an estimate for the total variation distance of  $N$  from the Poisson law, see Arratia et al. (1989), but there are models, lacking independence, where it is very hard to apply. Sieve methods often work with generalized Bonferroni inequalities of the form

$$P(N = k) \leq (\geq) \sum_M c(M) P\left(\bigcap_{i \in M} A_i\right), \quad (1)$$

where in the sum  $M$  runs over the subsets of  $\{1, 2, \dots, n\}$ , and the  $c(M)$  are real constants. Note that such an estimation is supposed to be valid for arbitrary events in an arbitrary probability space. Bonferroni-type inequalities with applications are collected in the monograph Galambos and Simonelli (1996).

An important example is the graph-sieve of Rényi (1976) (originally published in Hungarian in 1961), see also Móri (1990) for an application. While Rényi's formula only applies to the case  $k = 0$ , the general case is covered by Galambos (1966), but his estimates become more and more difficult to calculate as  $k$  grows. Therefore it is more convenient to show the approximate Poissonity of  $N$  by applying generating functions. If we have a lower or upper estimate for  $P(N = 0)$  of the form (1), we also have an estimate for the probability generating function of  $N$ , namely,

$$g_N(t) = E(t^N) \leq (\geq) \sum_M c(M) P\left(\bigcap_{i \in M} A_i\right) (1-t)^{|M|}, \quad (2)$$

for  $0 \leq t \leq 1$ , see Móri (1996). In this way we can estimate the deviation of  $g_N(t)$  from the generating function of the approximating Poisson distribution. Based on this information, can we estimate the deviation of  $P(N = k)$  from the corresponding Poisson probability?

More generally, let  $\mathcal{F}$  be the set of real power series  $f(t) = \sum_{k=0}^{\infty} a_k t^k$  such that  $\sum_{k=0}^{\infty} |a_k| \leq 2$  and  $\sum_{k=0}^{\infty} a_k = 0$ . Such a power series is convergent over the interval  $[0, 1]$ . Let  $\Delta = \Delta(f) = \max_{0 \leq t \leq 1} |f(t)|$ . How large can  $|a_k|$  be if  $\Delta$  is fixed?

If  $\mathbf{p} = (p_0, p_1, \dots)$  and  $\mathbf{q} = (q_0, q_1, \dots)$  are discrete probability distributions with generating functions  $g_{\mathbf{p}}(t)$  and  $g_{\mathbf{q}}(t)$ , resp., then  $f = g_{\mathbf{p}} - g_{\mathbf{q}} \in \mathcal{F}$ , and  $a_k = p_k - q_k$ .

The problem differs from those investigated in probability and mathematical analysis by the constraint that  $f$ , the generating function of the real sequence  $\mathbf{a} = (a_0, a_1, \dots)$ , is only available over the real interval  $[0, 1]$ . In a more common situation it is known over the complex unit circle (that is, we can work with the difference of characteristic functions), and estimates like the Caregradsky–Franken inequality are at hand, see Caregradsky (1958), Franken (1963/64).

In Móri (1996) it is shown that  $|a_k|$  cannot be estimated *uniformly in  $k$* . A simple counterexample is constructed by setting  $a_k = 1$ ,  $a_{k+1} = -1$ , and  $a_j = 0$  otherwise. Then  $|a_k| = 1$ , while

$$\Delta = \max_{0 \leq t \leq 1} t^k (1-t) \sim \frac{1}{ek}.$$

Let  $\varphi : [0, 1] \rightarrow \mathbb{R}$  be an arbitrary increasing function such that  $\lim_{t \downarrow 0} \varphi(t) < 1$ . Then  $\max_{j \geq 0} |a_j| = 1 > \varphi(\Delta)$ , if  $k$  is chosen sufficiently large.

Thus one can only hope to estimate  $|a_k|$  in terms of  $\Delta$  and  $k$ . Such an estimate is given by the following result of Móri (1996).

**Theorem 1.1** *Let  $\varepsilon > 0$  be arbitrarily small. Then there exists a sequence of finite constants  $c_k(\varepsilon)$ ,  $k = 0, 1, \dots$ , such that*

$$|a_k| \leq c_k(\varepsilon) \Delta^{1-\varepsilon}$$

for every  $f \in \mathcal{F}$ .

Can  $\varepsilon$  be 0? This question was posed and left unanswered in Móri (1996). As we shall see in Section 2, the answer is negative. However, the result of Theorem 1.1 can still be improved. New upper estimates are presented in Section 3.

## Acknowledgement

The author wishes to thank an anonymous referee for his/her valuable comments and suggestions.

## 2 Lower estimates

In the problem of estimating  $|a_k|$  in terms of  $\Delta$  for every fixed  $k$ , our next aim is to point out the limitations of such an estimation.

**Theorem 2.1** *Let  $k$  be an arbitrary positive integer and  $C$  a positive constant satisfying*

$$C < C_k = \frac{1}{2(2k)! \left[ \log(\sqrt{2} + 1) \right]^{2k}}.$$

*Then for every sufficiently small  $\Delta > 0$  there exists an  $f \in \mathcal{F}$  such that  $\Delta(f) = \Delta$ , and*

$$|a_k| \geq C \Delta \left( \log \frac{1}{\Delta} \right)^{2k}. \quad (3)$$

In other words, suppose that  $|a_k| \geq \varphi_k(\Delta)$  is valid for every  $f \in \mathcal{F}$  with some positive function  $\varphi_k$ . Then, by Theorem 2.1,

$$\liminf_{\Delta \downarrow 0} \frac{\varphi_k(\Delta)}{\Delta \left(\log \frac{1}{\Delta}\right)^{2k}} \geq C_k.$$

*Proof.* Let  $T_n$  denote the degree  $n$  Chebyshev polynomial of the first kind, defined by the trigonometric formula

$$\cos(n\omega) = T_n(\cos \omega).$$

It is well known that

$$T_n(t) = \frac{1}{2} \left[ \left(t + \sqrt{t^2 - 1}\right)^n + \left(t - \sqrt{t^2 - 1}\right)^n \right]. \quad (4)$$

The polynomials

$$S_n(t) = \frac{1}{2} (1 - T_n(1 - 2t))$$

are called *the spread polynomials*. They were introduced by Wildberger for use in rational trigonometry. A list of properties of spread polynomials, together with a reference, can be found in Wikipedia (available online [http://en.wikipedia.org/wiki/Spread\\_Polynomials](http://en.wikipedia.org/wiki/Spread_Polynomials)).

For our construction we will use the polynomials

$$Q_n(t) = S_n(1 - t) = \frac{1}{2} (1 - T_n(2t - 1)). \quad (5)$$

They satisfy the trigonometric identity

$$Q_n(\cos^2 \omega) = \sin^2(n\omega).$$

From (5) it is clear that  $Q_n(t) = S_n(t)$  for even  $n$ , and  $Q_n(t) = 1 - S_n(t)$  for odd  $n$ , thus

$$Q_n(t) = \frac{1 - (-1)^n}{2} + (-1)^n S_n(t).$$

Using Goh's explicit formula for the spread polynomials we obtain that

$$Q_n(t) = \frac{1 - (-1)^n}{2} + \frac{1}{2} \sum_{k=1}^n (-1)^{n-k+1} \frac{n}{k+n} \binom{k+n}{2k} (4t)^k.$$

Hence, the coefficient of  $t^k$  in  $Q_n(t)$  satisfies

$$|a_{n,k}| = \frac{n}{k+n} \binom{k+n}{2k} 2^{2k-1} \sim \frac{2^{2k-1} n^{2k}}{(2k)!}, \quad (6)$$

as  $n \rightarrow \infty$  and  $k > 0$  remains fixed. Moreover, by the inequality of arithmetic and geometric means,  $|a_{n,k}|$  is majorized by the right-hand side of (6).

Though  $\sum_{k=0}^n a_{n,k} = Q_n(1) = \frac{1}{2}(1 - T_n(1)) = 0$ ,  $Q_n$  cannot belong to  $\mathcal{F}$  when  $n$  is large enough, because  $s_n = \sum_{k=0}^n |a_{n,k}|$  tends to infinity:

$$s_n = (-1)^{n+1} Q_n(-1) = \frac{(-1)^n}{2} (T_n(-3) - 1).$$

By (4) we have  $s_n \sim 2^{-2} (3 + \sqrt{8})^n = 2^{-2} (\sqrt{2} + 1)^{2n}$ .

Let  $f = \Delta Q_n$ , where

$$n = \left\lfloor \frac{\log \frac{1}{\Delta}}{2 \log(\sqrt{2} + 1)} \right\rfloor,$$

then obviously  $\Delta(f) = \Delta$ . Since  $\Delta s_n \sim 2^{-2} \Delta (\sqrt{2} + 1)^{2n} < 1$ , it follows that  $f \in \mathcal{F}$  if  $n$  is large enough, that is, if  $\Delta$  is sufficiently small. We then have

$$\log \frac{1}{\Delta} \sim \log s_n \sim 2n \log(\sqrt{2} + 1).$$

Comparing this with (6) we conclude that

$$\lim_{\Delta \downarrow 0} \frac{|a_{n,k}|}{(\log \Delta)^{2k}} = C_k.$$

Thus  $f$  satisfies (3) if  $\Delta$  is small enough. This completes the proof.  $\square$

Let us repeat this result in terms of probability distributions.

**Corollary 2.1** *Let  $k$  be an arbitrary positive integer and  $C < C_k$  a positive constant. Then for every sufficiently small  $\Delta > 0$  there exist discrete probability distributions  $\mathbf{p}$  and  $\mathbf{q}$  with generating functions  $g_{\mathbf{p}}(t)$  and  $g_{\mathbf{q}}(t)$ , respectively, such that  $\max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)| = \Delta$ , and*

$$|p_k - q_k| \geq C \Delta \left( \log \frac{1}{\Delta} \right)^{2k}. \quad \square$$

Here both  $\mathbf{p}$  and  $\mathbf{q}$  depend on  $k$  and  $C$ , but in Poisson approximation one of the distributions is fixed. This constraint may improve the upper bound by excluding certain counterexamples. As we shall see in Section 3, in this case even the total variation distance

$$\|\mathbf{p} - \mathbf{q}\| = \sum_{k=0}^{\infty} |p_k - q_k|$$

can be estimated in terms of  $\Delta = \max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)|$ .

However, the following theorem shows that the lower bound of Corollary 2.1 remains valid, apart from the value of the threshold  $C_k$ , even if one of the distributions is fixed in such a way that its tail is not too light.

**Theorem 2.2** *Let  $\mathbf{p} = (p_0, p_1, \dots)$  be a fixed discrete probability distribution such that  $p_k > 0$  for every  $k = 0, 1, \dots$ , and*

$$\limsup_{k \rightarrow \infty} \frac{1}{2k} \log \frac{1}{p_k} = v < \infty. \quad (7)$$

*Let  $k$  be a positive integer and  $C$  a positive constant such that*

$$2(2k)!(1+v)^{2k}C < 1$$

*Then for every sufficiently small positive  $\Delta$  there exists a discrete probability distribution  $\mathbf{q} = (q_0, q_1, \dots)$ , such that  $\max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)| = \Delta$ , and*

$$|p_k - q_k| > C \Delta \left( \log \frac{1}{\Delta} \right)^{2k}.$$

*Proof.* Let  $\varepsilon > 0$  be sufficiently small to satisfy  $2(2k)!(1+v+\varepsilon)^{2k}C < 1$ . Let  $q_j = p_j + \Delta a_{n,j}$  for  $j = 0, 1, \dots, n$ , where

$$n = \left\lfloor \frac{1}{2(1+v+\varepsilon)} \log \frac{1}{\Delta} \right\rfloor, \quad (8)$$

and  $q_j = p_j$  for  $j > n$ . Note that  $\Delta \leq e^{-2n(1+v+\varepsilon)}$  by this choice of  $n$ .

First we show that this definition is correct, that is,  $\mathbf{q}$  is a probability distribution if  $\Delta$  is small enough, or equivalently,  $n$  is large enough. From (7) it follows that  $p_j \geq u e^{-(v+\varepsilon)2j}$ ,  $j = 0, 1, \dots$ , with some positive  $u$ , thus for every  $j \leq n$  we have

$$\frac{p_j}{|\Delta a_{n,j}|} \geq \frac{2u(2j)!}{\Delta(2n)^{2j}e^{(v+\varepsilon)2j}}.$$

By analyzing the ratio of consecutive terms one can see that the right-hand side is a decreasing function of  $j$ . Hence, by applying (8) and the Stirling formula we obtain that

$$\frac{p_j}{|\Delta a_{n,j}|} \geq \frac{2u(2n)!}{\Delta(2n)^{2n}e^{(v+\varepsilon)2n}} \geq \frac{2u(2n)!e^{2n}}{(2n)^{2n}} \geq 4u\sqrt{n\pi} > 1.$$

Now,  $g_{\mathbf{q}}(t) - g_{\mathbf{p}}(t) = \Delta \cdot Q_n(t)$ , hence  $\Delta(g_{\mathbf{p}} - g_{\mathbf{q}}) = \Delta$ . Therefore, by (6),

$$\frac{|p_k - q_k|}{\Delta \left( \log \frac{1}{\Delta} \right)^{2k}} = \frac{|a_{n,k}|}{\left( \log \frac{1}{\Delta} \right)^{2k}} \sim \frac{1}{2(2k)!(1+v+\varepsilon)^{2k}} > C,$$



as  $\Delta \rightarrow 0$ .  $\square$

If the tail of  $\mathbf{p}$  is lighter than exponential, the order of magnitude of the lower bound decreases. Namely, the following variant of Theorem 2.2 can be proved by similar reasoning.

**Theorem 2.3** *Let  $\mathbf{p} = (p_0, p_1, \dots)$  be a fixed discrete probability distribution such that  $p_k > 0$  for every  $k = 0, 1, \dots$ , and*

$$\limsup_{k \rightarrow \infty} \frac{1}{h(k)} \log \frac{1}{p_k} = v,$$

*where  $v$  is positive and finite,  $h$  is a positive, continuous, increasing function, regularly varying at infinity with exponent  $\alpha$ , and  $\lim_{k \rightarrow \infty} h(k)/k = \infty$  (hence  $\alpha \geq 1$ ). Let  $k$  be a positive integer and  $C$  a positive constant satisfying*

$$2^{1-2k} (2k)! v^{2k/\alpha} C < 1.$$

*Then for every sufficiently small positive  $\Delta$  there exists a discrete probability distribution  $\mathbf{q} = (q_0, q_1, \dots)$ , such that  $\max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)| = \Delta$ , and*

$$|p_k - q_k| > C \Delta \left( h^{-1} \left( \log \frac{1}{\Delta} \right) \right)^{2k}.$$

*Proof.* Choose  $\varepsilon > 0$  to satisfy

$$C < \frac{2^{2k-1}}{(2k)! (v + \varepsilon)^{2k/\alpha}}. \quad (9)$$

Let  $q_j = p_j + \Delta a_{n,j}$  for  $j = 0, 1, \dots, n$ , where

$$n = \left\lfloor h^{-1} \left( \frac{1}{v + \varepsilon} \log \frac{1}{\Delta} \right) \right\rfloor$$

and  $q_j = p_j$  for  $j > n$ . Then  $\Delta \leq e^{-h(n)(v+\varepsilon)}$ . Again,  $\mathbf{q}$  is a probability distribution if  $n$  is large enough; this can be shown in the same way as in the proof of Theorem 2.2.

This time  $\log \frac{1}{\Delta} \sim h(n)(v + \varepsilon)$ , and  $h^{-1}$  is regularly varying at infinity with exponent  $1/\alpha$ . Therefore, by (6) and (9),

$$\frac{|p_k - q_k|}{\Delta \left( h^{-1} \left( \log \frac{1}{\Delta} \right) \right)^{2k}} = \frac{|a_{n,k}|}{\left( h^{-1} \left( \log \frac{1}{\Delta} \right) \right)^{2k}} \sim \frac{2^{2k-1} n^{2k}}{(2k)! (n(v + \varepsilon)^{1/\alpha})^{2k}} > C,$$

as  $\Delta \rightarrow 0$ .  $\square$

Particularly, when  $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$ , we have

$$\lim_{k \rightarrow \infty} \frac{1}{k \log k} \log \frac{1}{p_k} = 1,$$

hence the following theorem holds.

**Corollary 2.2** *Let  $\mathbf{p}$  be the Poisson distribution with parameter  $\lambda$ . Let  $k$  be a positive integer and  $0 < C < 2^{2k-1}/(2k)!$ . Then for every sufficiently small positive  $\Delta$  there exists a discrete probability distribution  $\mathbf{q}$  such that  $\max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)| = \Delta$ , and*

$$|p_k - q_k| > C \Delta \left( \frac{\log \frac{1}{\Delta}}{\log \log \frac{1}{\Delta}} \right)^{2k}. \quad \square$$

Next, let us see what the constructions of Theorems 2.2 and 2.3 give for the total variation distance  $\|\mathbf{p} - \mathbf{q}\|$ .

**Theorem 2.4** *Let  $\mathbf{p} = (p_0, p_1, \dots)$  be a fixed discrete probability distribution satisfying the conditions of Theorem 2.2, and let*

$$C > 1 - \frac{\log(\sqrt{2} + 1)}{1 + v}$$

*Then for every sufficiently small positive  $\Delta$  there exists a discrete probability distribution  $\mathbf{q} = (q_0, q_1, \dots)$ , such that  $\max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)| = \Delta$ , and*

$$\|\mathbf{p} - \mathbf{q}\| > \Delta^C.$$

*Proof.* Let  $\mathbf{q}$  be the discrete probability distribution constructed in the proof of Theorem 2.2, with a sufficiently small  $\varepsilon$ . Then  $\|\mathbf{p} - \mathbf{q}\| = \Delta s_n$ , where  $s_n$  is the quantity introduced in the proof of Theorem 2.1. Thus, we have

$$\|\mathbf{p} - \mathbf{q}\| = \Delta \cdot 2^{-2} (\sqrt{2} + 1)^{2n} (1 + o(1)).$$

Here

$$\frac{\log \left( 2^{-2} (\sqrt{2} + 1)^{2n} \right)}{\log \Delta} = -\frac{\log(\sqrt{2} + 1)}{1 + v + \varepsilon} + \mathcal{O}(1/n),$$

which is smaller than  $C - 1$ , if  $n$  is large and  $\varepsilon$  is small enough.  $\square$

**Theorem 2.5** *Let  $\mathbf{p} = (p_0, p_1, \dots)$  be a fixed discrete probability distribution satisfying the conditions of Theorem 2.3, and let*

$$0 < C < 2 \log(\sqrt{2} + 1) v^{-1/\alpha}. \quad (10)$$

Then for every sufficiently small positive  $\Delta$  there exists a discrete probability distribution  $\mathbf{q} = (q_0, q_1, \dots)$ , such that  $\max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)| = \Delta$ , and

$$\|\mathbf{p} - \mathbf{q}\| > \Delta \cdot \exp\left(C h^{-1}\left(\log \frac{1}{\Delta}\right)\right).$$

*Proof.* Let  $\mathbf{q}$  be the discrete probability distribution constructed in the proof of Theorem 2.3, with a sufficiently small  $\varepsilon$ . Then  $\|\mathbf{p} - \mathbf{q}\| = \Delta s_n$  again. Since  $\log \frac{1}{\Delta} = h(n)(v + \varepsilon)(1 + o(1))$ , and  $h^{-1}$  is regularly varying with exponent  $1/\alpha$ , we obtain that

$$n = h^{-1}\left(\frac{1 + o(1)}{v + \varepsilon} \log \frac{1}{\Delta}\right) = \frac{1 + o(1)}{(v + \varepsilon)^{1/\alpha}} h^{-1}\left(\log \frac{1}{\Delta}\right).$$

Therefore

$$\begin{aligned} s_n &= \frac{1}{4} \exp\left(2n \log(\sqrt{2} + 1)\right) (1 + o(1)) \\ &= \exp\left(\frac{2 \log(\sqrt{2} + 1)}{(v + \varepsilon)^{1/\alpha}} (1 + o(1)) h^{-1}\left(\log \frac{1}{\Delta}\right) + \mathcal{O}(1)\right) \\ &> \exp\left(C h^{-1}\left(\log \frac{1}{\Delta}\right)\right), \end{aligned}$$

if  $n$  is large and  $\varepsilon$  is small enough.  $\square$

**Corollary 2.3** *Let  $\mathbf{p}$  be the Poisson distribution with parameter  $\lambda$ . Then for every sufficiently small positive  $\Delta$  there exist a discrete probability distribution  $\mathbf{q} = (q_0, q_1, \dots)$ , such that  $\max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)| = \Delta$ , and*

$$\|\mathbf{p} - \mathbf{q}\| > \Delta \exp\left(1.76 \frac{\log \frac{1}{\Delta}}{\log \log \frac{1}{\Delta}}\right).$$

*Proof.* In the case of Poisson distribution  $h(n) = n \log n$ ,  $\alpha = 1$ , and  $v = 1$ , thus the right-hand side of (10) equals  $1.7627 \dots$   $\square$

Let us close this section with an open problem. Are the above constructions extremal in the sense that the bounds of Corollaries 2.1 and 2.2 are sharp? That is, are there finite constants  $C'_k$  such that for arbitrary discrete distributions  $\mathbf{p}$  and  $\mathbf{q}$  we can write

$$|p_k - q_k| \leq C'_k \Delta \left(\log \frac{1}{\Delta}\right)^{2k} ?$$

Are there finite constants  $C_k''$  such that for every  $\lambda > 0$  and discrete distribution  $\mathbf{q}$

$$\left| \frac{\lambda^k}{k!} e^{-\lambda} - q_k \right| \leq C_k'' \Delta \left( \frac{\log \frac{1}{\Delta}}{\log \log \frac{1}{\Delta}} \right)^{2k} ?$$

Chebyshev polynomials are characterized by their heavy oscillation with small amplitude. I have the feeling that oscillation is somehow necessary for getting relatively large coefficients. For instance, Chebyshev polynomials are known to have the largest principal coefficient among all polynomials of the same degree with fixed  $L_\infty$ -norm over the interval  $[-1, +1]$ . In the light of these facts our construction seems extremal, at least in the order of magnitude.

### 3 Upper estimates

First we show that the  $\Delta^{1-\varepsilon}$  bound of Theorem 1.1 can be improved to  $\Delta$  multiplied by a function of  $\Delta$ , which is slowly varying at 0. All one has to do is to choose  $\varepsilon$  optimally in Theorem 2(c) of Móri (1996).

**Theorem 3.1** *For every  $f \in \mathcal{F}$ ,*

$$|a_k| \leq \Delta \exp \left( 2 \left( k \log \frac{1}{\Delta} \right)^{4/5} \right), \quad k = 0, 1, \dots$$

*Proof.* For  $k = 0$  the inequality is obvious. Let  $k \geq 1$ . If  $\log \frac{1}{\Delta} < 2^5 k^4$ , then

$$\begin{aligned} \Delta \exp \left( 2 \left( k \log \frac{1}{\Delta} \right)^{4/5} \right) &= \exp \left( 2 \left( k \log \frac{1}{\Delta} \right)^{4/5} - \log \frac{1}{\Delta} \right) \\ &= \exp \left( \left( \log \frac{1}{\Delta} \right)^{4/5} \left( \left( 2^5 k^4 \right)^{1/5} - \left( \log \frac{1}{\Delta} \right)^{1/5} \right) \right) \geq 1, \end{aligned}$$

hence we can suppose that

$$\log \frac{1}{\Delta} \geq 2^5 k^4. \quad (11)$$

By Theorem 2(c) of Móri (1996) we have

$$|a_k| \leq 2 \cdot 3^{\binom{\lceil u \rceil + 1}{2}} \Delta^{1-\varepsilon},$$

where  $u = \frac{k^2}{\pi \varepsilon^2}$ . Therefore

$$|a_k| \leq \Delta \exp \left( \log 2 + \log 3 \binom{u+2}{2} + \varepsilon \log \frac{1}{\Delta} \right).$$

For fixed  $k$  and  $\Delta$  let us choose  $\varepsilon$  to minimize the order of magnitude of the exponent above. Let

$$\varepsilon = \frac{1.6k}{\left(k \log \frac{1}{\Delta}\right)^{1/5}}. \quad (12)$$

Then

$$u = \frac{\left(k \log \frac{1}{\Delta}\right)^{2/5}}{1.6^2 k^2 \pi}, \quad (13)$$

and, by (11),  $\varepsilon \leq 0.8$ ,  $u \geq \frac{1}{\pi \varepsilon^2} > 0.4$ . Hence  $\binom{u+2}{2} u^{-2} < 11$ , thus

$$|a_k| \leq \Delta \exp\left(\log 2 + 13 u^2 + \varepsilon \log \frac{1}{\Delta}\right). \quad (14)$$

Here

$$\log 2 \leq \frac{\log 2}{2^4 k^4} \left(k \log \frac{1}{\Delta}\right)^{4/5}$$

follows by (11). Plugging this, (12), and (13) into (14) one obtains that

$$|a_k| \leq \Delta \exp\left(C \left(k \log \frac{1}{\Delta}\right)^{4/5}\right),$$

where

$$C = \frac{\log 2}{2^4 k^4} + \frac{13}{1.6^4 k^4 \pi^2} + 1.6 < 2. \quad \square$$

Though we have seen in Section 1 that  $|a_k|$  cannot be estimated uniformly in  $k$  in terms of  $\Delta$ , this is not the case when  $a_k = p_k - q_k$  with a fixed, known probability distribution  $\mathbf{p} = (p_0, p_1, \dots)$ . Then, for arbitrary discrete probability distribution  $\mathbf{q}$ , even the total variation distance

$$\|\mathbf{p} - \mathbf{q}\| = \sum_{k=0}^{\infty} |p_k - q_k|$$

can be estimated in terms of  $\Delta = \max_{0 \leq t \leq 1} |g_{\mathbf{p}}(t) - g_{\mathbf{q}}(t)|$ .

**Theorem 3.2** *Let the distribution  $\mathbf{p}$  be fixed. Then there exists an increasing function  $\varphi : [0, 1] \rightarrow \mathbb{R}$ , depending on  $\mathbf{p}$ , such that  $\lim_{t \downarrow 0} \varphi(t) = 0$ , and  $\|\mathbf{p} - \mathbf{q}\| \leq \varphi(\Delta)$  holds for arbitrary discrete probability distribution  $\mathbf{q}$ .*

*Proof.* Introduce the notation  $x^+ = \max\{x, 0\}$ , and let  $r_n = p_n + p_{n+1} + \dots$ . Then for arbitrary positive integer  $N$  we can write

$$\begin{aligned} \|\mathbf{p} - \mathbf{q}\| &= 2 \sum_{k=0}^{\infty} (p_k - q_k)^+ \leq 2 \sum_{k=0}^{N-1} |p_k - q_k| + 2r_N \\ &\leq 2 \sum_{k=0}^{N-1} \Delta \exp\left(2 \left(k \log \frac{1}{\Delta}\right)^{4/5}\right) + 2r_N, \end{aligned} \quad (15)$$

by Theorem 3.1. The right-hand side of (15) is independent of  $q$ . Let  $\varphi(\Delta)$  be its minimum as  $N$  runs over the positive integers. The minimum exists, because the first sum becomes greater than 1 if  $n$  is large enough. Let  $N$  be fixed. The first sum vanishes as  $\Delta$  tends to 0, thus  $\lim_{\Delta \downarrow 0} \varphi(\Delta) \leq r_N$ . Since  $r_N$  can be arbitrarily small, the proof is completed.  $\square$

Unfortunately, the function  $\varphi$  obtained this way is slowly varying at 0 in the most important cases, namely, whenever the tail of  $\mathbf{p}$  is not extremely light. Indeed, suppose that  $p_N > \exp(-N^{4-\varepsilon})$  holds eventually, and  $\Delta$  is small enough. The right-hand side of (15) is greater than

$$2\Delta \exp\left(2\left((N-1)\log\frac{1}{\Delta}\right)^{4/5}\right) + 2p_N.$$

If  $N > \left(\log\frac{1}{\Delta}\right)^{1/4}$ , the first term is at least 2. In the complementary case the second term is greater than

$$\exp\left(-\left(\log\frac{1}{\Delta}\right)^{1-\varepsilon/4}\right),$$

hence the minimum must be slowly varying at 0.

Finally, we return to our starting problem of Poisson approximation.

**Theorem 3.3** *Let  $\mathbf{p}$  be the Poisson distribution with parameter  $\lambda$ . Then*

$$\|\mathbf{p} - \mathbf{q}\| = o\left(\exp\left(-\left(\log\frac{1}{\Delta}\right)^{1/4}\right)\right),$$

*as  $\mathbf{q}$  varies in such a way that  $\Delta \rightarrow 0$ .*

This bound could be improved with a more careful analysis, but not essentially: the exponent 1/4 cannot be increased with our method.

*Proof.* Let  $\xi$  be a random variable with  $\text{Poisson}(\lambda)$  distribution. Then, by the Markov inequality we have

$$r_N = P(\xi \geq N) = P(N^\xi \geq N^N) \leq N^{-N} E N^\xi = e^{-\lambda} (e^{-\lambda} N)^{-N}.$$

From (15) it follows that

$$\|\mathbf{p} - \mathbf{q}\| \leq 2N\Delta \exp\left(2\left(N\log\frac{1}{\Delta}\right)^{4/5}\right) + 2e^{-\lambda} (e^{-\lambda} N)^{-N}. \quad (16)$$

Now is the time for us to choose  $N$ . Let

$$N = \left\lceil \frac{5 \left( \log \frac{1}{\Delta} \right)^{1/4}}{\log \log \frac{1}{\Delta}} \right\rceil.$$

The exponent in the first term of the right-hand side of (16) is  $o\left(\log \frac{1}{\Delta}\right)$ , thus the first term is negligible compared to what we have to prove. In the second term we can write

$$\left(e^{-\lambda} N\right)^{-N} = \exp(\lambda N - N \log N) = \exp\left(-\frac{5}{4} \left(\log \frac{1}{\Delta}\right)^{1/4} (1 + o(1))\right),$$

which proves the theorem.  $\square$

## References

- Arratia, R., Goldstein, L., Gordon, L., 1989. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.* 17, 9-25.
- Caregradsky, J. P., 1958. On uniform approximation to a binomial distribution by infinitely divisible laws (Russian). *Teor. Veroyatnost. i Primenen.* 3, 470-474.
- Franken, P., 1963/64. Approximation der Verteilungen von Summen unabhängiger nichtnegativer ganzzahliger Zufallsgrößen durch Poissonsche Verteilungen. *Math. Nachr.* 27, 303-340.
- Galambos, J., 1966. On the sieve methods in probability theory. *Studia Sci. Math. Hungar.* 1, 39-50.
- Galambos, J., Simonelli, I., 1996. *Bonferroni-type Inequalities with Applications*. Springer, New York.
- Móri, T. F., 1990. More on the waiting time till each of some given patterns occurs as a run. *Can. J. Math.* 42, 915-932.
- Móri, T. F., 1996. Bonferroni inequalities and deviations of discrete distributions. *J. Appl. Probab.* 33, 115-121.
- Rényi, A., 1976. A general method for proving theorems in probability theory and some of its applications. In: *Selected Papers of A. Rényi*. Akadémiai Kiadó, Budapest, Vol.2, pp. 581-602.