DOI 10.1007/s10986-019-09468-3
Lithuanian Mathematical Journal, Vol. 59, No. 4, October, 2019, pp. 469–497

Nonuniform bounds in the Poisson approximation with applications to informational distances. II

Sergey G. Bobkov[a,1], **Gennadiy P. Chistyakov**[b,2], **and Friedrich Götze**[b,2]

[a]School of Mathematics, University of Minnesota,
127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455, USA
[b]Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany
(e-mail: bobkov@math.umn.edu; chistyak@math.uni-bielefeld.de; goetze@math.uni-bielefeld.de)

Dedicated to Professor Vygantas Paulauskas on the occasion of his 75th birthday

Received June 21, 2019; revised September 19, 2019

Abstract. We explore asymptotically optimal bounds for deviations of distributions of independent Bernoulli random variables from the Poisson limit in terms of the Shannon relative entropy and Rényi/relative Tsallis distances (including Pearson's χ^2). This part generalizes the results obtained in Part I and removes any constraints on the parameters of the Bernoulli distributions.

MSC: 60E, 60F

Keywords: χ^2-divergence, relative entropy, Poisson approximation

1 Introduction

Let $W = X_1 + \cdots + X_n$ be the sum of independent random variables X_j taking values 1 and 0 with respective probabilities p_j and $q_j = 1 - p_j$. Thus

$$w_k = \mathbf{P}\{W = k\} = \sum p_1^{\varepsilon_1} q_1^{1-\varepsilon_1} \cdots p_n^{\varepsilon_n} q_n^{1-\varepsilon_n}, \quad k = 0, 1, \dots, n, \tag{1.1}$$

where the summation runs over all 0–1 sequences $\varepsilon_1, \dots, \varepsilon_n$ such that $\varepsilon_1 + \cdots + \varepsilon_n = k$.

Denote by Z a Poisson random variable with parameter $\lambda = p_1 + \cdots + p_n$, that is, taking nonnegative integer values with probabilities

$$v_k = \mathbf{P}\{Z = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots . \tag{1.2}$$

[1] Research was partially supported by the Simons Foundation and the NSF grant DMS-1855575.
[2] Research was partially supported by SFB 1283.

It is well known that if all p_j are small, then the distribution of Z approximates the distribution of W in terms of the total variation distance $d(W,Z) = \sum_{k=0}^{\infty} |w_k - v_k|$. In particular, involving the functional $\lambda_2 = p_1^2 + \cdots + p_n^2$, Barbour and Hall [2] derived the two-sided bound

$$\frac{1}{32}\min\Big(1,\frac{1}{\lambda}\Big)\lambda_2 \leqslant \frac{1}{2}d(W,Z) \leqslant \frac{1-\mathrm{e}^{-\lambda}}{\lambda}\lambda_2. \tag{1.3}$$

There is considerable interest as well in the question of Poisson approximation for (stronger) informational distances, including the Rényi divergences or, equivalently, the Tsallis relative entropies in their full hierarchy. Being well-defined in the setting of abstract measure spaces (see, e.g., [4, 12]), in the discrete model specified above, these important quantities are respectively given for any parameter $\alpha > 0$ by

$$D_\alpha = D_\alpha(W\,\|\,Z) = \frac{1}{\alpha-1}\log\sum_{k=0}^{\infty}\Big(\frac{w_k}{v_k}\Big)^{\alpha} v_k$$

and

$$T_\alpha = T_\alpha(W\,\|\,Z) = \frac{1}{\alpha-1}\bigg[\sum_{k=0}^{\infty}\Big(\frac{w_k}{v_k}\Big)^{\alpha} v_k - 1\bigg].$$

The functions $\alpha \to D_\alpha$ and $\alpha \to T_\alpha = (\exp\{(\alpha-1)D_\alpha\}-1)/(\alpha-1)$ are nondecreasing, and in the particular cases $\alpha = 1$ and $\alpha = 2$, we deal with the more familiar relative entropy (Kullback–Leibler distance) and the Pearson χ^2-distance

$$D = D_1 = T_1 = \sum_{k=0}^{\infty} w_k \log\frac{w_k}{v_k}, \qquad T_2 = \chi^2 = \sum_{k=0}^{\infty}\frac{(w_k - v_k)^2}{v_k}.$$

We refer to [13] and [3] for historical references related to the lower and upper bounds as in (1.3) and to recent developments toward the problem of bounding of D and χ^2. Here let us only mention a few results in this direction.

In a rather general asymptotic regime (which is typical in applications), Borisov and Vorozheykin [5] observed that χ^2 is approximately $(\lambda_2/\lambda)^2/2$ and, more precisely,

$$\lim \frac{\chi^2}{(\lambda_2/\lambda)^2} = \frac{1}{2} \quad \text{as } \lambda^6\lambda_2 \to 0.$$

On the other hand, Harremoës, Johnson, and Kontoyiannis [7] have recently derived a universal lower bound on the relative entropy, $D \geqslant (\lambda_2/\lambda)^2/4$. Here the constant $1/4$ is the best possible and is asymptotically attained in the case of equal probabilities p_j [8]. It is therefore natural to wonder whether or not there are two-sided bounds such as

$$\frac{1}{4}\Big(\frac{\lambda_2}{\lambda}\Big)^2 \leqslant D \leqslant \chi^2 \leqslant c\Big(\frac{\lambda_2}{\lambda}\Big)^2. \tag{1.4}$$

This turns out to be true in the case where λ_2/λ is bounded away from 1. Based on orthogonal expansions in Charlier polynomials over the Poisson measure and using the Parseval identity in this context, Zacharovas and Hwang [13] obtained the superior upper bound

$$\chi^2 \leqslant 2(\sqrt{\mathrm{e}}-1)^2\Big(\frac{\lambda_2}{\lambda}\Big)^2\Big(1-\frac{\lambda_2}{\lambda}\Big)^{-3} \tag{1.5}$$

(among other similar results for different distances). Consequently, if, for example, $\lambda_2/\lambda \leqslant 1/2$, then (1.4) is fulfilled with $c = 6.74$.

The upper estimate such as (1.4) also appears as a consequence of nonuniform bounds, which have been recently studied in [3]. It was shown there that $(w_k - v_k)/v_k$ is of order at most λ_2 on a large part of the support of the Poisson measure, especially when λ is large. One of the aims of this paper is to extend (1.4) modulo absolute constants to the whole range of (λ, λ_2). To formulate results in a compact form, let us use the notation $Q_1 \sim Q_2$ whenever two positive quantities are related by $c_1 Q_1 \leqslant Q_2 \leqslant c_2 Q_1$ with some absolute constants $c_j > 0$. Introduce the quantity

$$F = \frac{\max(1, \lambda)}{\max(1,\, \lambda - \lambda_2)}.$$

Clearly, $F \geqslant 1$.

Theorem 1. *We have*

$$D \sim \left(\frac{\lambda_2}{\lambda}\right)^2 (1 + \log F), \qquad \chi^2 \sim \left(\frac{\lambda_2}{\lambda}\right)^2 \sqrt{F}. \tag{1.6}$$

If λ_2/λ is bounded away from 1, then F is bounded, and (1.6) recovers (1.4). A similar conclusion is also true when λ is not large, say $\lambda \leqslant 10$, which is typical for applications (note that for such λ, λ_2/λ may be close to 1, and then (1.5) fails to be optimal). On the other hand, if these two assumptions on λ and λ_2 are violated (which we henceforth call the "degenerate case"), then both distances are bounded away from zero and can be large, since then

$$D \sim \log \frac{\lambda}{\max\{1, \lambda - \lambda_2\}}, \qquad \chi^2 \sim \left(\frac{\lambda}{\max\{1, \lambda - \lambda_2\}}\right)^{1/2}.$$

This shows that the lower bound for D in (1.4) may not be reversed in general. Indeed, in the extreme case with all $p_j = 1$, we have $\lambda_2 = \lambda = n$. Here $\mathbf{P}\{W = n\} = 1$, and hence, as $n \to \infty$,

$$D = \log \frac{1}{\mathbf{P}\{Z = n\}} = \log\left(\frac{n!}{n^n} e^n\right) \sim \log n,$$

$$\chi^2 = \frac{1}{\mathbf{P}\{Z = n\}} - 1 = \frac{n!}{n^n} e^n - 1 \sim \sqrt{2\pi n}.$$

As a next step, we employ the nonuniform bounds of [3] to extend (1.4) and (1.6) to all Tsallis entropies.

Theorem 2. *Given* $\alpha > 1$,

$$T_\alpha \sim \left(\frac{\lambda_2}{\lambda}\right)^2 F^{(\alpha-1)/2}$$

with involved constants depending on α. *In particular,* $T_\alpha \leqslant c_\alpha \chi^2$ *as long as* $\lambda_2/\lambda \leqslant 1/2$.

Let us finally mention one application of Theorem 1 to the problem of the estimation of the difference of entropies

$$H(W \,\|\, Z) = H(Z) - H(W), \tag{1.7}$$

where H stands for the Shannon entropy, that is,

$$H(Z) = -\sum_k v_k \log v_k, \qquad H(W) = -\sum_k w_k \log w_k.$$

The property that $H(W \,\|\, Z)$ is positive is a consequence of the assertion, recently proved by Hillion and Johnson [9], that $H(p) \equiv H(W)$ is a concave function of the vector $p = (p_1, \dots, p_n)$. Indeed, since $H(p)$ is invariant under permutations of p_j, this entropy attains its maximum on the simplex $p_j \geqslant 0$, $p_1 + \cdots + p_n = \lambda$ at the point where all coordinates coincide, that is, for $p_j = \lambda/n$. But in that case the distribution of W represents the binomial law with parameters n and λ/n, whose entropy is dominated by $H(Z)$, as was shown by Harremoës [6].

Thus the difference of entropies in this particular model may be viewed as kind of informational distance. Sason [11] proposed to bound $H(W \,\|\, Z)$ for equal p_j by means of the so-called maximal coupling. Here we show that this distance may be controlled in terms of $\chi^2(W, Z)$, which, together with the upper bound on the Pearson distance as in (1.4)–(1.5), leads to the following estimate.

Corollary 1. *With some constants C_λ depending only on λ, we have*

$$H(W \,\|\, Z) \leqslant C_\lambda \frac{\lambda_2}{\lambda}. \tag{1.8}$$

If $\lambda_2 \leqslant \lambda/2$, then one may take $C_\lambda = C \log(2 + \lambda)$ with an absolute constant C.

Further, we start with some general bounds involving the relative entropy and the Pearson distance (Section 2). In Section 3, we describe several results obtained in [3] in the nondegenerate case and employ there some bounds for the probability function of the Poisson law. The remaining parts are devoted to the proof of Theorems 1 and 2 in the degenerate case (Sections 4–10) and of Corollary 1 (Section 11).

2 General bounds on relative entropy and χ^2

Before turning to the problem of lower and upper bounds for the relative entropy and χ^2-distance, we first collect several useful general inequalities. If two discrete random elements W and Z in a measurable space Ω take at most countably many values $\omega_k \in \Omega$ with probabilities $w_k = \mathbf{P}\{W = \omega_k\}$ and $v_k = \mathbf{P}\{Z = \omega_k\}$, then the above distances are defined canonically by

$$D(W \,\|\, Z) = \sum_k w_k \log \frac{w_k}{v_k}, \qquad \chi^2(W, Z) = \sum_k \frac{(w_k - v_k)^2}{v_k}.$$

Proposition 1. *We have*

$$-\sum_{w_k < v_k} w_k \log \frac{w_k}{v_k} \leqslant 1. \tag{2.1}$$

Moreover,

$$D(W \,\|\, Z) \geqslant \frac{1}{2} \sum_k \frac{(w_k - v_k)^2}{\max\{w_k, v_k\}}. \tag{2.2}$$

Proof. Using the Taylor formula for the logarithmic function, write

$$\begin{aligned} \sum_{w_k < v_k} w_k \log \frac{w_k}{v_k} &= \sum_{w_k < v_k} \big(v_k - (v_k - w_k)\big) \log\Big(1 - \frac{v_k - w_k}{v_k}\Big) \\ &= \sum_{w_k < v_k} (w_k - v_k) + \sum_{w_k < v_k} \sum_{m=2}^{\infty} \frac{1}{m(m-1)} \frac{(v_k - w_k)^m}{v_k^{m-1}}, \end{aligned}$$

where

$$\sum_{w_k<v_k}(w_k-v_k) = -\frac{1}{2}\sum_{k=0}^{\infty}|w_k-v_k| \geqslant -1,$$

thus proving the first statement. Similarly, we have the second identity

$$\begin{aligned}\sum_{w_k>v_k} w_k \log\frac{w_k}{v_k} &= -\sum_{w_k>v_k} w_k\log\frac{v_k}{w_k} = -\sum_{w_k>v_k} w_k\log\left(1-\frac{w_k-v_k}{w_k}\right)\\ &= \sum_{w_k>v_k}(w_k-v_k) + \sum_{w_k>v_k}\sum_{m=2}^{\infty}\frac{1}{m}\frac{(w_k-v_k)^m}{w_k^{m-1}}.\end{aligned}$$

Adding the two identities, we get

$$\sum_k w_k\log\frac{w_k}{v_k} \geqslant \frac{1}{2}\sum_{w_k>v_k}\frac{(w_k-v_k)^2}{w_k} + \frac{1}{2}\sum_{w_k<v_k}\frac{(w_k-v_k)^2}{v_k},$$

which is the desired inequality (2.2). □

Proposition 2. *Let W_1 and W_2 be independent nonnegative integer-valued random variables with finite means, and let Z_1 and Z_2 be independent Poisson random variables with $\mathbf{E}Z_1 = \mathbf{E}W_1$ and $\mathbf{E}Z_2 = \mathbf{E}W_2$. Then*

$$D(W_1+W_2 \,\|\, Z_1+Z_2) \leqslant D(W_1\,\|\,Z_1) + D(W_2\,\|\,Z_2). \tag{2.3}$$

In addition,

$$\chi^2(W_1+W_2,\, Z_1+Z_2)+1 \leqslant \bigl(\chi^2(W_1,Z_1)+1\bigr)\bigl(\chi^2(W_2,Z_2)+1\bigr). \tag{2.4}$$

For the proof, we refer to Johnson [10, pp. 133–134]. Let us only mention that (2.4) is obtained in [10] in the more general form

$$\sum_{k=0}^{\infty}\frac{\mathbf{P}\{W_1+W_2=k\}^{\alpha}}{\mathbf{P}(Z_1+Z_2=k\}^{\alpha-1}} \leqslant \sum_{k=0}^{\infty}\frac{\mathbf{P}\{W_1=k\}^{\alpha}}{\mathbf{P}\{Z_1=k\}^{\alpha-1}}\sum_{k=0}^{\infty}\frac{\mathbf{P}\{W_2=k\}^{\alpha}}{\mathbf{P}\{Z_2=k\}^{\alpha-1}}$$

with arbitrary $\alpha \geqslant 1$, which represents a Poisson analog of weighted convolution inequalities due to Andersen [1]. Here, for $\alpha = 1$, there is an equality, and comparing the derivatives of both sides at this point, we arrive at relation (2.3).

3 Poisson approximation in the nondegenerate case

Now we restrict ourselves to the random variables $W = X_1 + \cdots + X_n$ and Z with distributions described in (1.1)–(1.2). In particular,

$$\begin{aligned}\mathbf{P}\{Z=0\} &= \mathrm{e}^{-(p_1+\cdots+p_n)} = \mathrm{e}^{-\lambda},\\ \mathbf{P}\{W=0\} &= (1-p_1)\cdots(1-p_n) \leqslant \mathbf{P}\{Z=0\}.\end{aligned} \tag{3.1}$$

The bounds (1.4) follow from the following two statements proved in [3]. To compare the lower and upper bounds, we recall the lower bound (1.4) of Harremoës, Johnson, and Kontoyiannis [7].

Proposition 3. *If* $\max_j p_j \leqslant 1/2$*, then*

$$\frac{1}{4}\left(\frac{\lambda_2}{\lambda}\right)^2 \leqslant D(W \,\|\, Z) \leqslant \chi^2(W,Z) \leqslant C_\lambda \left(\frac{\lambda_2}{\lambda}\right)^2,$$

where C_λ *depends on* $\lambda \geqslant 0$ *and is an increasing continuous function with* $C_0 = 2$*. In particular, if* $\lambda \leqslant 1/2$*, then*

$$\chi^2(W,Z) \leqslant 15\left(\frac{\lambda_2}{\lambda}\right)^2.$$

Proposition 4. *If* $\lambda \geqslant 1/2$ *and* $\lambda_2 \leqslant \kappa\lambda$ *with* $\kappa \in (0,1)$*, then*

$$\frac{1}{4}\left(\frac{\lambda_2}{\lambda}\right)^2 \leqslant D(W \,\|\, Z) \leqslant \chi^2(W,Z) \leqslant c_\kappa \left(\frac{\lambda_2}{\lambda}\right)^2, \tag{3.2}$$

where we may take $c_\kappa = c(1-\kappa)^{-3}$ *with some absolute constant, for example,* $c = 7 \cdot 10^6$*.*

A natural approach to the Poisson approximation is based on the comparison of characteristic functions. Since the random variables W and Z take nonnegative integer values only, we may equivalently consider the associated generating functions, similar as in [3]. The generating function for the Poisson law with parameter $\lambda > 0$ is given by

$$\varphi(w) = \mathbf{E}w^Z = \sum_{k=0}^{\infty} \mathbf{P}\{Z = k\} w^k = \mathrm{e}^{\lambda(w-1)} = \prod_{j=1}^{n} \mathrm{e}^{p_j(w-1)},$$

which is an entire function of the complex variable w. Correspondingly, the generating function for the distribution of the random variable W is

$$g(w) = \mathbf{E}w^W = \sum_{k=0}^{\infty} \mathbf{P}\{W = k\} w^k = \prod_{j=1}^{n} (q_j + p_j w),$$

which is a polynomial of degree n. Hence the difference between the involved probabilities may be expressed with the help of the contour integrals by the Cauchy formula

$$\mathbf{P}\{W = k\} - \mathbf{P}\{Z = k\} = \int_{|w|=r} w^{-k}\big(g(w) - \varphi(w)\big)\,\mathrm{d}\mu_r(w),$$

where μ_r is the uniform probability measure on the circle $|w| = r$ of arbitrary radius $r > 0$. This identity for the difference of probabilities was used in [3] in the derivation of the upper bound in (3.2), whereas here the representation

$$\mathbf{P}\{W = k\} = \int_{|w|=r} w^{-k} g(w)\,\mathrm{d}\mu_r(w) \tag{3.3}$$

will be particularly helpful in the study of the degenerate case.

When estimating the Poisson probabilities

$$f(k) = \mathbf{P}\{Z = k\} = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}$$

for a fixed parameter $\lambda > 0$, it is convenient to use the well-known Stirling-type two-sided bound

$$\sqrt{2\pi}\,k^{k+1/2}\mathrm{e}^{-k} \leqslant k! \leqslant \mathrm{e}k^{k+1/2}\mathrm{e}^{-k} \quad (k \geqslant 1). \tag{3.4}$$

In particular, it implies the following Gaussian type estimates (see [3]).

Lemma 1. *For all $k \geqslant 1$,*

$$f(k) \leqslant \frac{1}{\sqrt{2\pi k}}. \tag{3.5}$$

Moreover, if $1 \leqslant k \leqslant 2\lambda$, then

$$\frac{1}{\mathrm{e}\sqrt{k}}\mathrm{e}^{-(k-\lambda)^2/\lambda} \leqslant f(k) \leqslant \frac{1}{\sqrt{2\pi k}}\mathrm{e}^{-(k-\lambda)^2/(3\lambda)}. \tag{3.6}$$

Here, the lower bound may be improved in the region $k \geqslant \lambda$ as

$$f(k) \geqslant \frac{1}{\mathrm{e}\sqrt{k}}\mathrm{e}^{-(k-\lambda)^2/(2\lambda)}. \tag{3.7}$$

4 Upper bounds on D and χ^2

We now turn to Theorem 2 in the degenerate case, where the optimal bounds on the relative entropy and χ^2 have a different behavior. As an intermediate step, let us derive the following upper bounds for the χ^2-distance and the relative entropy by using the quantity

$$Q = \frac{\lambda}{\max\{1,\, \lambda - \lambda_2\}}.$$

Proposition 5. *For $\lambda \geqslant 1/2$, we have*

$$\chi^2(W, Z) \leqslant 19\sqrt{Q}, \tag{4.1}$$

$$D(W \,\|\, Z) \leqslant 23\log(\mathrm{e}Q). \tag{4.2}$$

These bounds are sharp when $\lambda_2 \geqslant \kappa\lambda$; see Propositions 6 and 7.

Proof. Setting $g(w) = \prod_{l=1}^n (q_l + p_l w)$, $w \in \mathbb{C}$, we exploit the contour integral representation (3.3), that is,

$$\mathbf{P}\{W = k\} = \frac{1}{2\pi} r^{-k} \int_{-\pi}^{\pi} g\bigl(r\mathrm{e}^{\mathrm{i}\theta}\bigr)\mathrm{e}^{-\mathrm{i}k\theta}\,\mathrm{d}\theta, \quad r > 0.$$

It yields the upper bound

$$\mathbf{P}\{W = k\} \leqslant R_k(r) I(r), \tag{4.3}$$

where

$$R_k(r) = r^{-k}\prod_{l=1}^n (q_l + p_l r) \quad \text{and} \quad I(r) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\prod_{l=1}^n \frac{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r}\,\mathrm{d}\theta.$$

Let us choose $r = k/\lambda$. Since $q_j + p_j r \leqslant e^{p_j(r-1)}$, we have

$$R_k(r) \leqslant r^{-k} \prod_{j=1}^n (q_j + p_j r) \leqslant e^{\lambda(r-1) - k \log r} = \left(\frac{e\lambda}{k}\right)^k e^{-\lambda}.$$

Moreover, applying $(e/k)^k \leqslant e\sqrt{k}/k!$ (cf. (3.4), this is simplified to

$$R_k(r) \leqslant e\sqrt{k}\frac{\lambda^k}{k!}e^{-\lambda} = e\sqrt{k}f(k), \tag{4.4}$$

where $f(k)$ is the density of the Poisson law with parameter λ.

Now, to bound $I(r)$, for all $|\theta| \leqslant \pi$, using $\sin(\theta/2) \geqslant \theta/pi$, we have

$$\prod_{l=1}^n \frac{|q_l + p_l r e^{i\theta}|}{q_l + p_l r} = \prod_{l=1}^n \left(1 - \frac{4q_l p_l r}{(q_l + p_l r)^2}\sin^2\frac{\theta}{2}\right)^{1/2} \leqslant \exp\left\{-2\sin^2\frac{\theta}{2}\sum_{l=1}^n \frac{q_l p_l r}{(q_l + p_l r)^2}\right\}$$
$$\leqslant \exp\left\{-\frac{2\theta^2}{\pi^2}\sum_{l=1}^n \frac{q_l p_l r}{(q_l + p_l r)^2}\right\}.$$

Here

$$\sum_{l=1}^n \frac{q_l p_l r}{(q_l + p_l r)^2} \geqslant \begin{cases} \frac{1}{r}\sum_{l=1}^n q_l p_l = \frac{1}{r}(\lambda - \lambda_2) & \text{in case } r \geqslant 1, \\ r\sum_{l=1}^n q_l p_l = r(\lambda - \lambda_2) & \text{in case } r \leqslant 1. \end{cases}$$

These right-hand sides have the form

$$\psi(r) = \min\left\{r, \frac{1}{r}\right\}(\lambda - \lambda_2),$$

and we get

$$I(r) \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left\{-\frac{2}{\pi^2}\psi(r)\theta^2\right\} d\theta = \frac{1}{4\psi(r)^{1/2}}\int_{-2\sqrt{\psi(r)}}^{2\sqrt{\psi(r)}} e^{-x^2/2}\, dx$$
$$\leqslant \frac{1}{4\psi(r)^{1/2}}\min\left\{\sqrt{2\pi}, 4\psi(r)^{1/2}\right\} \leqslant \min\left\{1, \psi(r)^{-1/2}\right\}.$$

First, we consider the region $\lambda/4 \leqslant k \leqslant 4\lambda$, in which case $1/4 \leqslant r \leqslant 4$ and $\psi(r) \geqslant (\lambda - \lambda_2)/4$, and thus

$$I(r) \leqslant \min\left\{1, \frac{2}{\sqrt{\lambda - \lambda_2}}\right\} \leqslant 2\sqrt{Q_0}, \quad Q_0 = \frac{1}{\max\{1,\ \lambda - \lambda_2\}}.$$

Applying this bound together with (4.4) in (4.3), we get

$$\mathbf{P}\{W = k\} \leqslant 2e\sqrt{Q_0}\sqrt{k}f(k). \tag{4.5}$$

As for the regions $1 \leqslant k < \lambda/4$ and $k > 4\lambda$, we use the property $|I(r)| \leqslant 1$, which yields the simpler upper bounds

$$\mathbf{P}\{W = k\} \leqslant \left(\frac{e\lambda}{k}\right)^k e^{-\lambda} \leqslant e\sqrt{k}f(k). \tag{4.6}$$

Now recall that $\mathbf{P}\{W=0\} \leqslant f(0)$ (as mentioned in (3.1)) and write

$$\chi^2(W,Z) = \sum_{k=0}^{\infty} \frac{\mathbf{P}\{W=k\}^2}{f(k)} - 1 \leqslant S_1 + S_2 + S_3$$
$$= \biggl(\sum_{1\leqslant k<\lambda/4} + \sum_{\lambda/4\leqslant k\leqslant 4\lambda} + \sum_{k>4\lambda} \biggr) \frac{\mathbf{P}\{W=k\}^2}{f(k)}.$$

By (4.5) we have

$$S_2 \leqslant 2\mathrm{e}\sqrt{Q_0} \sum_{\lambda/4\leqslant k\leqslant 4\lambda} \sqrt{k}\, \mathbf{P}\{W=k\} \leqslant 4\mathrm{e}\sqrt{Q} \sum_{\lambda/4\leqslant k\leqslant 4\lambda} \mathbf{P}\{W=k\} \leqslant 4\mathrm{e}\sqrt{Q}.$$

To estimate S_1, first note that $S_1 = 0$ for $\lambda < 4$. For $\lambda \geqslant 4$, since the function $k \to (\mathrm{e}\lambda/k)^k$ is increasing for $k < \lambda$, we obtain from (4.6) that

$$S_1 \leqslant \mathrm{e}^{-\lambda+1} \sum_{k<\lambda/4} \sqrt{k} \biggl(\frac{\mathrm{e}\lambda}{k}\biggr)^k \leqslant \frac{1}{2}\sqrt{\lambda}\,\mathrm{e}^{-\lambda+1} \sum_{1\leqslant k<\lambda/4} \biggl(\frac{\mathrm{e}\lambda}{k}\biggr)^k$$
$$\leqslant \frac{1}{2}\sqrt{\lambda}\,\mathrm{e}^{-\lambda+1} \sum_{1\leqslant k<\lambda/4} (4\mathrm{e})^{\lambda/4} \leqslant \mathrm{e}\biggl(\frac{\lambda}{4}\biggr)^{3/2} \biggl(\frac{4}{\mathrm{e}^3}\biggr)^{\lambda/4}$$
$$\leqslant \mathrm{e}\biggl(\frac{3}{2\mathrm{e}\log(\mathrm{e}^3/4)}\biggr)^{3/2} < 0.544.$$

Here we applied the inequality

$$x^p c^x \leqslant \biggl(\frac{p}{\mathrm{e}\log(1/c)}\biggr)^p, \quad p, x > 0,\ 0 < c < 1, \tag{4.7}$$

with $p = 3/2$ and $c = 4/\mathrm{e}^3$.

To estimate S_3, we may bound the sequence $\sqrt{k}(\mathrm{e}\lambda/k)^k$ for $k > 4\lambda \geqslant 2$ by the geometric progression Ab^k with suitable parameters $A > 0$ and $0 < b < 1$. To this aim, consider the function

$$u(x) = \log\biggl(\sqrt{x}\biggl(\frac{\mathrm{e}\lambda}{x}\biggr)^x\biggr) - \log b^x$$
$$= \frac{1}{2}\log x + x + x\log\lambda - x\log x - x\log b, \quad x \geqslant 4\lambda.$$

We have

$$u'(x) = \frac{1}{2x} + \log\lambda - \log x - \log b \leqslant \frac{1}{4} + \log\frac{1}{4b} \leqslant 0$$

if $b \geqslant \mathrm{e}^{1/4}/4$, which we assume. In this case, u is decreasing, so that $u(x) \leqslant u(4\lambda) = \log(2\sqrt{\lambda}(\mathrm{e}/(4b))^{4\lambda}) \leqslant \log A$, where

$$A = 2\sup_{\lambda\geqslant 1/2} \sqrt{\lambda}\biggl(\frac{\mathrm{e}}{4b}\biggr)^{4\lambda} = \sup_{y\geqslant 2} \sqrt{y}\biggl(\frac{\mathrm{e}}{4b}\biggr)^y = \biggl(\frac{1}{2\mathrm{e}\log(3/\mathrm{e})}\biggr)^{1/2} < 1.366,$$

where on the last step, we chose $b = 3/4$ and applied (4.7) with $p = 1/2$ and $c = \mathrm{e}/3$. Thus, putting $k_0 = [4\lambda] + 1$ and noting that $k_0 \geqslant 2$, we get

$$S_3 \leqslant \mathrm{e}^{-\lambda+1} \sum_{k>4\lambda} \sqrt{k}\Big(\frac{\mathrm{e}\lambda}{k}\Big)^k \leqslant \sqrt{\mathrm{e}} \sum_{k\geqslant k_0} A\Big(\frac{3}{4}\Big)^k = 4A\sqrt{\mathrm{e}}\Big(\frac{3}{4}\Big)^{k_0} \leqslant \frac{9}{4}A\sqrt{\mathrm{e}} < 5.067.$$

Finally, using $Q = \lambda Q_0 \geqslant 1/2$ (due to $\lambda \geqslant 1/2$), we get $S_1 + S_3 < 5.611 \leqslant 5.611\sqrt{2Q}$. This gives $S_1 + S_2 + S_3 < (5.611\sqrt{2} + 4e)\sqrt{Q} < 18.81\sqrt{Q}$, so (4.1) follows.

Turning to the second statement and using $\mathbf{P}\{W = 0\} \leqslant f(0)$, write similarly

$$D(W \,\|\, Z) = \sum_{k=0}^{\infty} \mathbf{P}\{W = k\} \log \frac{\mathbf{P}\{W = k\}}{\mathbf{P}\{Z = k\}} = T_1 + T_2 + T_3$$
$$\leqslant \bigg(\sum_{1\leqslant k<\lambda/4} + \sum_{\lambda/4\leqslant k\leqslant 4\lambda} + \sum_{k>4\lambda}\bigg) \mathbf{P}\{W = k\} \log \frac{\mathbf{P}\{W = k\}}{f(k)}.$$

For the region $\lambda/4 \leqslant k \leqslant 4\lambda$, we can apply the bound (4.5) again, which gives

$$\mathbf{P}\{W = k\} \leqslant 2\sqrt{Q_0}\mathrm{e}\sqrt{k} f(k) \leqslant 4\mathrm{e}\sqrt{Q} f(k),$$

and therefore, since $Q \geqslant 1/2$,

$$T_2 \leqslant \log(4\mathrm{e}) + \frac{1}{2}\log Q \leqslant \frac{\log(4\mathrm{e}) - \log(2)/2}{\log(\mathrm{e}/2)} \log(\mathrm{e}Q) < 6.65 \log(\mathrm{e}Q).$$

Using (4.6) together with the inequality $\log(\mathrm{e}t) \leqslant t$ $(t > 0)$, we obtain, similarly to the derivation of the bound on T_1 in the χ^2-case, that

$$T_1 \leqslant \mathrm{e}^{-\lambda} \sum_{1\leqslant k<\lambda/4} \Big(\frac{\mathrm{e}\lambda}{k}\Big)^k \log(\mathrm{e}\sqrt{k}) \leqslant \mathrm{e}^{-\lambda} \log\bigg(\mathrm{e}\sqrt{\frac{\lambda}{4}}\bigg) \sum_{1\leqslant k<\lambda/4} \Big(\frac{\mathrm{e}\lambda}{k}\Big)^k$$
$$\leqslant \Big(\frac{\lambda}{4}\Big)^{3/2} \Big(\frac{4}{\mathrm{e}^3}\Big)^{\lambda/4} \leqslant \bigg(\frac{3}{2\mathrm{e}\log(\mathrm{e}^3/4)}\bigg)^{3/2} < 0.2.$$

Choosing again $k_0 = [4\lambda] + 1$, similarly to the derivation of the bound on S_3 in the χ^2-case, we also get

$$T_3 \leqslant \mathrm{e}^{-\lambda} \sum_{k>4\lambda} \Big(\frac{\mathrm{e}\lambda}{k}\Big)^k \log(\mathrm{e}\sqrt{k}) \leqslant \mathrm{e}^{-\lambda+1} \sum_{k\geqslant k_0} \sqrt{k}\Big(\frac{\mathrm{e}\lambda}{k}\Big)^k < 5.067.$$

Hence $T_1 + T_3 < 5.087 < 16.578 \log(\mathrm{e}Q)$, and (4.2) follows as well. □

5 Lower bound on χ^2

Here we complement Proposition 5 by a similar lower bound for the χ^2-distance in terms of the same quantity $Q = \lambda/\max\{1,\ \lambda - \lambda_2\}$. Let $c_0 = 2.5 \cdot 10^{-6}$.

Proposition 6. *If $\lambda \geqslant 1/2$, then, with some absolute constant $c \in [c_0, 1)$,*

$$1 + \chi^2(W, Z) \geqslant c\sqrt{Q}. \tag{5.1}$$

Moreover, as long as $\lambda_2 \geqslant (1 - c^2/4)\lambda$,

$$\chi^2(W, Z) \geqslant \frac{c}{9}\sqrt{Q}. \tag{5.2}$$

Suppose that $\lambda_2 \geqslant (1 - c^2/4)\lambda$. To derive (5.2) from (5.1), it is sufficient to require that $c\sqrt{Q} \geqslant 2$, since then $c\sqrt{Q} - 1 \geqslant c\sqrt{Q}/2$. This condition is fulfilled, as long as $\lambda \geqslant \lambda_0 = 4/c^2$, and then we obtain (5.2). In the remaining case $1/2 \leqslant \lambda \leqslant \lambda_0$, inequality (5.2) follows from the lower bound

$$\chi^2(W, Z) \geqslant \frac{1}{4}\left(\frac{\lambda_2}{\lambda}\right)^2;$$

cf. (1.4). Indeed, in this case, $\lambda - \lambda_2 \leqslant c^2\lambda/4 \leqslant 1$, so that $Q = \lambda \leqslant 4/c^2$, and thus $c\sqrt{Q}/9 \leqslant 2/9$, whereas $(\lambda_2/\lambda)^2/4 \geqslant (1 - c^2/4)^2/4$.

Thus it remains to derive the first inequality in (5.1). First, we will prove it assuming that $\lambda - \lambda_2$ is sufficiently large. As in Section 4, for any fixed $r > 0$, we apply the Cauchy theorem and write

$$\mathbf{P}\{W = k\} = \int\limits_{|w|=r} w^{-k}\prod_{l=1}^{n}(q_l + p_l w)\,\mathrm{d}\mu_r(w) = R_k(r)I_k(r)$$

with integration over the uniform distribution μ_r on the circle $|w| = r$ of the complex plane. Here and further,

$$R_k(r) = r^{-k}\prod_{l=1}^{n}(q_l + p_l r)$$

and

$$I_k(r) = \frac{1}{2\pi}\int\limits_{-\pi}^{\pi}\prod_{l=1}^{n}\frac{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r}\exp\left\{-\mathrm{i}k\theta + \mathrm{i}\sum_{l=1}^{n}\mathrm{Im}\big(\log\big(q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}\big)\big)\right\}\mathrm{d}\theta.$$

We split the integration over the two regions so that to work with the representation

$$\mathbf{P}\{W = k\} = R_k(r)I_k(r) = R_k(r)\big(I_{k1}(r) + I_{k2}(r)\big),$$

where

$$I_{k1}(r) = \frac{1}{2\pi}\int\limits_{-\pi/2}^{\pi/2}\prod_{l=1}^{n}\frac{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r}\exp\left\{-\mathrm{i}k\theta + \mathrm{i}\sum_{l=1}^{n}\mathrm{Im}\big(\log\big(q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}\big)\big)\right\}\mathrm{d}\theta,$$

$$I_{k2}(r) = \frac{1}{2\pi}\int\limits_{\pi/2<|\theta|<\pi}\prod_{l=1}^{n}\frac{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r}\exp\left\{-\mathrm{i}k\theta + \mathrm{i}\sum_{l=1}^{n}\mathrm{Im}\big(\log\big(q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}\big)\big)\right\}\mathrm{d}\theta.$$

To properly estimate $I_k(r)$ from below, $I_{k2}(r)$ needs to be estimated from above (in absolute value), whereas $I_{k1}(r)$, which is a real number, should be estimated from below.

Furthermore, the quantity $R_k(r)$ needs to be estimated from below as well. To this aim, we choose the radius $r = r(k) > 0$ by the condition $R_k'(r) = 0$ or, equivalently,

$$F(r) \equiv \sum_{l=1}^{n} \frac{p_l r}{q_l + p_l r} = k. \tag{5.3}$$

Since the function F is monotone and $F(0) = 0$, $F(\infty) = n$, there is a unique solution, say r, to this equation as long as $n > k$ (which may be assumed). We also assume that not all p_k are equal to 0 or 1, so that $\lambda_2 < \lambda$.

Let us also emphasize that F is concave on the positive half-axis. Since $F(1) = \lambda$, we necessarily have $r(k) < 1$ in case $k < \lambda$ and $r(k) > 1$ in case $k > \lambda$.

Lemma 2. *For any $k = 0, \dots, n-1$, the solution $r = r(k)$ to equation* (5.3) *satisfies*

$$r \geqslant 1 + \frac{k - \lambda}{\lambda - \lambda_2}.$$

Moreover, in case $|k - \lambda| \leqslant (\lambda - \lambda_2)/6$, we have $5/6 \leqslant r \leqslant 6/5$, and in fact with some $0 \leqslant b_i \leqslant 1$ we have

$$r = 1 + \left(\frac{6}{5}\right)^2 b_1 \frac{k-\lambda}{\lambda - \lambda_2} = 1 + \frac{k-\lambda}{\lambda-\lambda_2} + \left(\frac{6}{5}\right)^9 b_2 \frac{\lambda_2 - \lambda_3}{\lambda - \lambda_2} \left(\frac{k-\lambda}{\lambda-\lambda_2}\right)^2.$$

Proof. We have

$$F'(r) = \sum_{l=1}^{n} \frac{p_l q_l}{(q_l + p_l r)^2}, \qquad F'(1) = \lambda - \lambda_2.$$

The inverse function $F^{-1} : [0, n) \to [0, \infty)$ is increasing and convex. Hence, for any $s \in [0, n)$,

$$\begin{aligned} F^{-1}(s) &\geqslant F^{-1}(\lambda) + \big(F^{-1}\big)'(\lambda)(s - \lambda) = F^{-1}(\lambda) + \frac{1}{F'(F^{-1}(\lambda))}(s - \lambda) \\ &= 1 + \frac{1}{\lambda - \lambda_2}(s - \lambda). \end{aligned}$$

Plugging $s = k$, we obtain the first inequality.

Now, since $q_l + p_l r \leqslant 1$ for $r \leqslant 1$, we conclude that $F'(r) \geqslant \sum_{l=1}^n p_l q_l = \lambda - \lambda_2$ and $F(1) - F(r) \geqslant (1 - r)(\lambda - \lambda_2)$. Thus, if $k \leqslant \lambda$, then we obtain that

$$\frac{1}{6}(\lambda - \lambda_2) \geqslant |k - \lambda| = F(1) - F\big(r(k)\big) \geqslant \big(1 - r(k)\big)(\lambda - \lambda_2),$$

implying $r(k) \geqslant 5/6$. For $r \geqslant 1$, we may use $q_l + p_l r \leqslant r$, which gives $F'(r) \geqslant (\lambda - \lambda_2)/r^2$ and $F(r) - F(1) \geqslant (1 - 1/r)(\lambda - \lambda_2)$. Hence, again by the assumption,

$$\frac{1}{6}(\lambda - \lambda_2) \geqslant k - \lambda = F\big(r(k)\big) - F(1) \geqslant \left(1 - \frac{1}{r(k)}\right)(\lambda - \lambda_2),$$

implying $r(k) \leqslant 6/5$. In both cases, $5/6 \leqslant r(k) \leqslant 6/5$, proving the second statement of the lemma.

Now, in the interval $5/6 \leqslant r \leqslant 65$, we necessarily have $5/6 \leqslant q_l + p_l r \leqslant 6/5$, so that

$$\left(\frac{5}{6}\right)^2 (\lambda - \lambda_2) \leqslant F'(r) \leqslant \left(\frac{6}{5}\right)^2 (\lambda - \lambda_2).$$

In addition,

$$-F''(r) = 2\sum_{l=1}^{n} \frac{p_l^2 q_l}{(q_l + p_l r)^3} \leqslant 2 \cdot \left(\frac{6}{5}\right)^3 \sum_{l=1}^{n} p_l^2 q_l = 2 \cdot \left(\frac{6}{5}\right)^3 (\lambda_2 - \lambda_3).$$

Let us now write the Taylor expansion up to the linear and quadratic terms for the inverse function $F^{-1}(s)$ around the point λ. Then we get

$$\begin{aligned} F^{-1}(s) &= 1 + \frac{1}{F'(F^{-1}(s_1))}(s - \lambda) \\ &= 1 + \frac{1}{F'(1)}(s - \lambda) - \frac{1}{2F'(F^{-1}(s_2))^3} F''\big(F^{-1}(s_2)\big)(s - \lambda)^2, \end{aligned}$$

where the points s_1 and s_2 lie between λ and s. Putting $r = F^{-1}(s)$ and $r_i = F^{-1}(s_i)$, this is simplified as

$$r = 1 + \frac{1}{F'(r_1)}(s - \lambda) = 1 + \frac{1}{\lambda - \lambda_2}(s - \lambda) - \frac{1}{2F'(r_2)^3} F''(r_2)(s - \lambda)^2,$$

where r_1 and r_2 lie between 1 and r. It remains to apply these equalities with $s = k$, that is, $r = r(k)$, and note that $1/F'(r_1) \leqslant (6/5)^2(1/(\lambda - \lambda_2)$, whereas

$$\frac{1}{2F'(r_2)^3}\big|F''(r_2)\big| \leqslant \frac{1}{2(\frac{5}{6})^6(\lambda - \lambda_2)^3} \cdot 2 \cdot \left(\frac{6}{5}\right)^3 (\lambda_2 - \lambda_3) = \left(\frac{6}{5}\right)^9 \frac{\lambda_2 - \lambda_3}{(\lambda - \lambda_2)^3}.$$

Note that $(6/5)^2 = 1.44$ and $(6/5)^9 < 5.16$. $\square$

Lemma 3. *Let $r = r(k)$ be the solution of* (5.3) *for $0 \leqslant \lambda - k \leqslant (\lambda - \lambda_2)/6$. Then*

$$R_k(r) = r^{-k} \prod_{l=1}^{n} (q_l + p_l r) \geqslant \exp\left\{-4\frac{(\lambda - k)^2}{\lambda - \lambda_2}\right\}.$$

Proof. The function

$$\psi_k(r) = \log R_k(r) = \sum_{l=1}^{n} \log(q_l + p_l r) - k \log r, \quad r > 0,$$

is vanishing at $r = 1$ and has the derivative

$$\psi_k'(r) = \sum_{l=1}^{n} \frac{p_l}{q_l + p_l r} - \frac{k}{r} = \frac{F(r) - k}{r} = \frac{F(r) - F(r(k))}{r}.$$

Since F is increasing and concave, $F(a) - F(b) \leqslant F'(b)(a - b)$ whenever $a \geqslant b > 0$. In particular, in the interval $r(k) \leqslant r \leqslant 1$, we have

$$\psi_k'(r) \leqslant \frac{F'(r(k))}{r}\big(r - r(k)\big) \leqslant \frac{F'(r(k))}{r(k)}\big(1 - r(k)\big),$$

which implies

$$\psi_k\big(r(k)\big) = \psi_k\big(r(k)\big) - \psi_k(1) \geqslant -\frac{F'(r(k))}{r(k)}\big(1 - r(k)\big)^2.$$

By Lemma 2, $5/6 \leqslant r(k) \leqslant 1$ and $1 - r(k) \leqslant (6/5)^2(\lambda - k)/(\lambda - \lambda_2)$. Moreover, as was shown in the proof, $F'(r(k)) \leqslant (6/5)^2(\lambda - \lambda_2)$. Hence

$$\frac{F'(r(k))}{r(k)}\big(1 - r(k)\big)^2 \leqslant \frac{(\frac{6}{5})^2(\lambda - \lambda_2)}{5/6}\bigg(\Big(\frac{6}{5}\Big)^2\frac{k-\lambda}{\lambda - \lambda_2}\bigg)^2 = \Big(\frac{6}{5}\Big)^7\frac{(k-\lambda)^2}{\lambda - \lambda_2}.$$

Here $(6/5)^7 < 3.6$. □

Lemma 4. *Let* $\lambda - \lambda_2 \geqslant 100$. *Then, for* $0 \leqslant \lambda - k \leqslant (\lambda - \lambda_2)/6$,

$$I_k\big(r(k)\big) \geqslant \frac{1}{10\sqrt{\lambda - \lambda_2}}.$$

Proof. By Lemma 2, $1 \geqslant r(k) \geqslant 5/6$. As in the proof of Proposition 5, recall that for $r > 0$ and $-\pi \leqslant \theta \leqslant \pi$,

$$\prod_{l=1}^n \frac{|q_l + p_l r \mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r} = \prod_{l=1}^n \bigg(1 - \frac{4q_l p_l r}{(q_l + p_l r)^2}\sin^2\frac{\theta}{2}\bigg)^{1/2} \leqslant \exp\bigg\{-2\sum_{l=1}^n \frac{q_l p_l r}{(q_l + p_l r)^2}\sin^2\frac{\theta}{2}\bigg\}.$$

For $5/6 \leqslant r \leqslant 1$, necessarily $q_l + p_l r \leqslant 1$ and

$$\sum_{l=1}^n \frac{q_l p_l r}{(q_l + p_l r)^2} \geqslant \sum_{l=1}^n q_l p_l r = (\lambda - \lambda_2) r \geqslant \frac{5}{6}(\lambda - \lambda_2).$$

Hence

$$\begin{aligned}\big|I_{k2}(r)\big| &\leqslant \frac{1}{2\pi}\int_{\pi/2 \leqslant |\theta| \leqslant \pi} \prod_{l=1}^n \frac{|q_l + p_l r \mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r}\,\mathrm{d}\theta \leqslant \frac{1}{2\pi}\int_{\pi/2 \leqslant |\theta| \leqslant \pi} \exp\bigg\{-\frac{5}{3}(\lambda - \lambda_2)\sin^2\frac{\theta}{2}\bigg\}\,\mathrm{d}\theta \\ &\leqslant \frac{1}{2}\mathrm{e}^{-(5/6)(\lambda - \lambda_2)}.\end{aligned}$$

Let us now estimate $I_{k1}(r)$ from below. Using $4q_l p_l r \leqslant (q_l + p_l r)^2$, which is equivalent to $(q_l - p_l r)^2 \geqslant 0$, we have, for $|\theta| \leqslant \pi/2$,

$$\frac{4q_l p_l r}{(q_l + p_l r)^2}\sin^2\frac{\theta}{2} \leqslant \frac{1}{2}, \quad l = 1, \dots, n.$$

In the region $0 \leqslant \varepsilon \leqslant \varepsilon_0 < 1$, there is a lower bound $1 - \varepsilon \geqslant \mathrm{e}^{-c\varepsilon}$ with best attainable constant when $\varepsilon = \varepsilon_0$. In the case $\varepsilon_0 = 1/2$, this constant is given by $c = 2\log 2$. Therefore, for $|\theta| \leqslant \pi/2$,

$$\prod_{l=1}^n \frac{|q_l + p_l r \mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r} \geqslant \exp\bigg\{-\log 2\sum_{l=1}^n \frac{4q_l p_l r}{(q_l + p_l r)^2}\sin^2\frac{\theta}{2}\bigg\}.$$

Here the involved function

$$w_l(r) = \frac{r}{(q_l + p_l r)^2}, \quad r \geqslant 0,$$

is increasing in $0 \leqslant r \leqslant r_l \equiv q_l/p_l$ and decreasing in $r \geqslant r_l$. Hence, if $r_l \geqslant 1$, then $\max_{5/6 \leqslant r \leqslant 1} w_l(r) = w_l(1) = 1$. If $r_l \leqslant 5/6$, that is, when $p_l \geqslant 6/11$, we have

$$\max_{5/6 \leqslant r \leqslant 1} w_l(r) = w_l\Big(\frac{5}{6}\Big) = \frac{\frac{5}{6}}{(q_l + \frac{5}{6}p_l)^2} \leqslant \frac{6}{5}.$$

Finally, if $5/6 \leqslant r_l \leqslant 1$, which is equivalent to $1/2 \leqslant p_l \leqslant 6/11$, then we have

$$\max_{5/6 \leqslant r \leqslant 1} w_l(r) = w_l(r_l) = \frac{1}{4 p_l q_l} \leqslant \frac{1}{4 \cdot \frac{6}{11} \cdot \frac{5}{11}} = \frac{121}{120}.$$

Thus, in all cases, $w_l(r) \leqslant 6/5$ on the interval $5/6 \leqslant r \leqslant 1$, so that

$$\prod_{l=1}^{n} \frac{|q_l + p_l r \mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r} \geqslant \exp\left\{ -\frac{6}{5} \log 2 \sum_{l=1}^{n} 4 q_l p_l \sin^2 \frac{\theta}{2} \right\} \geqslant \exp\left\{ -\frac{6}{5} \log(2)(\lambda - \lambda_2)\theta^2 \right\},$$

and thus

$$\begin{aligned}
\frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} \prod_{l=1}^{n} \frac{|q_l + p_l r \mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r} \,\mathrm{d}\theta &\geqslant \frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} \exp\left\{ -\frac{6}{5} \log(2)(\lambda - \lambda_2)\theta^2 \right\} \mathrm{d}\theta \\
&= \frac{1}{2\pi \sqrt{\frac{6}{5} \log(4)(\lambda - \lambda_2)}} \int_{-(\pi/2)\sqrt{(6/5)\log(4)(\lambda-\lambda_2)}}^{(\pi/2)\sqrt{(6/5)\log(4)(\lambda-\lambda_2)}} \exp\left\{ -\frac{1}{2} x^2 \right\} \mathrm{d}x \\
&\geqslant 0.3093 \frac{1}{\sqrt{\lambda - \lambda_2}}.
\end{aligned}$$

Here we used $\lambda - \lambda_2 \geqslant 100$, which ensures that

$$\begin{aligned}
\frac{1}{2\pi \sqrt{\frac{6}{5} \log 4}} \int_{-(\pi/2)\sqrt{(6/5)\log(4)(\lambda-\lambda_2)}}^{(\pi/2)\sqrt{(6/5)\log(4)(\lambda-\lambda_2)}} \mathrm{e}^{-x^2/2} \,\mathrm{d}x &\geqslant \frac{1}{2\pi \sqrt{\frac{6}{5} \log 4}} \int_{-5\pi\sqrt{(6/5)\log 4}}^{5\pi\sqrt{(6/5)\log 4}} \mathrm{e}^{-x^2/2} \,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi \frac{6}{5} \log 4}} \mathbf{P}\left\{ |\xi| \leqslant 5\pi \sqrt{\frac{6}{5} \log 4} \right\} > 0.3093,
\end{aligned}$$

where $\xi \sim N(0,1)$. In addition (recalling one of the upper bounds when bounding the integral I_{k2} from above), using the inequality $\sin(\theta/2) \geqslant (\sqrt{2}/\pi)\theta$ for $0 \leqslant \theta \leqslant \pi/2$, we get that

$$\begin{aligned}
\frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \prod_{l=1}^{n} \frac{|q_l + p_l r \mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r} \theta^6 \,\mathrm{d}\theta &\leqslant \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \exp\left\{ -\frac{5}{3}(\lambda - \lambda_2) \sin^2 \frac{\theta}{2} \right\} \theta^6 \,\mathrm{d}\theta \\
&\leqslant \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \exp\left\{ -\frac{10}{3\pi^2}(\lambda - \lambda_2)\theta^2 \right\} \theta^6 \,\mathrm{d}\theta
\end{aligned}$$

$$\leqslant \frac{1}{\pi}\Big(\frac{20}{3\pi^2}(\lambda-\lambda_2)\Big)^{-7/2}\int_{-\infty}^{\infty} e^{-x^2/2}x^6\,dx$$

$$= \pi^{13/2}\Big(\frac{3}{20}\Big)^{7/2} 15\sqrt{2}\frac{1}{(\lambda-\lambda_2)^{7/2}} < \frac{48}{(\lambda-\lambda_2)^{7/2}}.$$

Now, assumption (5.3) may be rewritten as

$$\mathrm{Im}\Big(\sum_{l=1}^n \log\big(q_l+p_l r e^{i\theta}\big)\Big)'\Big|_{\theta=0} = \Big(\sum_{l=1}^n \mathrm{Im}\big(\log\big(q_l+p_l r e^{i\theta}\big)\big)\Big)'\Big|_{\theta=0} = k.$$

Here the functions $\mathrm{Im}(\log(q_l+p_l r e^{i\theta}))$ are odd, so their second derivatives are vanishing at zero. We now apply the Taylor formula up to the cubic term to the function

$$A_k(r,\theta) = -k\theta + \mathrm{Im}\sum_{l=1}^n \log\big(q_l+p_l r e^{i\theta}\big)$$

on the interval $\theta\in[-\pi/2,\pi/2]$ to get that

$$A_k(r,\theta) = \frac{1}{6}\Big(\mathrm{Im}\sum_{l=1}^n \log\big(q_l+p_l r e^{iv}\big)\Big)'''\Big|_{v=\theta_0}\theta^3$$

with some $\theta_0\in[-\pi/2,\pi/2]$. To perform differentiation, consider a function of the form

$$h(v) = \log\big(q+pre^{iv}\big), \quad p,q,r>0.$$

We have

$$h'(v) = \frac{prie^{iv}}{q+pre^{iv}} = i\Big(1-\frac{q}{q+pre^{iv}}\Big) = i - iq\big(q+pre^{iv}\big)^{-1},$$

$$h''(v) = -pqre^{iv}\big(q+pre^{iv}\big)^{-2},$$

$$h'''(v) = -pqr\big(ie^{iv}\big(q+pre^{iv}\big)^{-2} - 2ipre^{2iv}\big(q+pre^{iv}\big)^{-3}\big).$$

Therefore

$$-\Big(\mathrm{Im}\sum_{l=1}^n \log\big(q_l+p_l r e^{i\theta}\big)\Big)''' = \mathrm{Im}\Big(i\sum_{l=1}^n \frac{p_l q_l r e^{i\theta}}{(q_l+p_l r e^{i\theta})^2}\Big) - 2\mathrm{Im}\Big(i\sum_{l=1}^n \frac{q_l p_l^2 r^2 e^{2i\theta}}{(q_l+p_l r e^{i\theta})^3}\Big),$$

implying that

$$\Big|\Big(\mathrm{Im}\sum_{l=1}^n \log\big(q_l+p_l r e^{i\theta}\big)\Big)'''\Big| \leqslant \sum_{l=1}^n \frac{p_l q_l r}{|q_l+p_l r e^{i\theta}|^2} + 2\sum_{l=1}^n \frac{q_l p_l^2 r^2}{|q_l+p_l r e^{i\theta}|^3}.$$

But, for $5/6\leqslant r\leqslant 1$ and $|\theta|\leqslant\pi/2$,

$$\big|q_l+p_l r e^{i\theta}\big|^2 = (q_l+p_l r)^2\Big(1-\frac{4q_l p_l r}{(q_l+p_l r)^2}\sin^2\frac{\theta}{2}\Big) \geqslant (q_l+p_l r)^2 - 2q_l p_l r = q_l^2+p_l^2 r^2.$$

Hence

$$\frac{r}{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|^2} \leqslant \frac{r}{q_l^2 + p_l^2 r^2} = u_l(r) \leqslant \frac{121}{60}.$$

Here we used the property that $u_l(r)$ is increasing in $r \leqslant r_l = q_l/p_l$ and decreasing in $r \geqslant r_l$. If $r_l \geqslant 1$, then this gives $u_l(r) \leqslant u_l(1) = 1/(q_l^2 + p_l^2) \leqslant 2$. If $r_l \leqslant 5/6$, that is, when $p_l \geqslant 6/11$, we get $u_l(r) \leqslant u_l(5/6) = (5/6)/(q_l^2 + (5/6)p_l^2)$. The latter expression is minimized at $p_l = 6/11$, where it has the value $121/66$. Finally, if $5/6 \leqslant r_l \leqslant 1$, which is equivalent to $1/2 \leqslant p_l \leqslant 6/11$, then we have

$$u_l(r) \leqslant u_l(r_l) = \frac{1}{2p_l q_l} \leqslant \frac{1}{2 \cdot \frac{6}{11} \cdot \frac{5}{11}} = \frac{121}{60}.$$

From this it follows that

$$\frac{r^2}{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|^3} \leqslant \left(\frac{r^{4/3}}{q_l^2 + p_l^2 r^2}\right)^{3/2} \leqslant \left(\frac{r}{q_l^2 + p_l^2 r^2}\right)^{3/2} = u_l(r)^{3/2} \leqslant \left(\frac{121}{60}\right)^{3/2},$$

so that

$$\left|\left(\operatorname{Im}\sum_{l=1}^{n} \log\bigl(q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}\bigr)\right)'''\right| \leqslant \frac{121}{60}\sum_{l=1}^{n} p_l q_l + 2\left(\frac{121}{60}\right)^{3/2}\sum_{l=1}^{n} q_l p_l^2 \leqslant c_0(\lambda - \lambda_2)$$

with $c_0 = 121/60 + 2(121/60)^{3/2} < 7.744438$. Thus

$$\bigl|A_k(r,\theta)\bigr| \leqslant \frac{c_0}{6}(\lambda - \lambda_2)|\theta|^3, \quad \frac{5}{6} \leqslant r \leqslant 1,\ |\theta| \leqslant \frac{\pi}{2}.$$

Now, as we mentioned before, the function A_k is odd in θ, so that $I_{k1}(r)$ is a real number given by

$$\begin{aligned} I_{k1}(r) &= \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} \prod_{l=1}^{n} \frac{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r} \cos\bigl(A_k(r,\theta)\bigr)\,\mathrm{d}\theta \\ &= \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} \prod_{l=1}^{n} \frac{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r}\,\mathrm{d}\theta - \frac{1}{\pi}\int_{-\pi/2}^{\pi/2} \prod_{l=1}^{n} \frac{|q_l + p_l r\mathrm{e}^{\mathrm{i}\theta}|}{q_l + p_l r} \sin^2\frac{A(r,\theta)}{2}\,\mathrm{d}\theta. \end{aligned}$$

Hence, using

$$\sin^2\frac{A(r,\theta)}{2} \leqslant \frac{1}{4}A_k(r,\theta)^2 \leqslant \frac{c_0^2}{144}(\lambda - \lambda_2)^2\theta^6,$$

from the previous estimates we may deduce the lower bound

$$\begin{aligned} I_{k1}(r) &\geqslant 0.3093\frac{1}{\sqrt{\lambda - \lambda_2}} - \frac{c_0^2}{144}(\lambda - \lambda_2)^2\frac{48}{(\lambda - \lambda_2)^{7/2}} = 0.3093\frac{1}{\sqrt{\lambda - \lambda_2}} - \frac{c_0^2}{3}\frac{1}{(\lambda - \lambda_2)^{3/2}} \\ &\geqslant \frac{1}{\sqrt{\lambda - \lambda_2}}\left(0.3093 - \frac{20}{\lambda - \lambda_2}\right) \geqslant 0.1093\frac{1}{\sqrt{\lambda - \lambda_2}}, \end{aligned}$$

where on the last step, we assume that $\lambda - \lambda_2 \geqslant 100$. Together with the upper bound on I_{k2}, we arrive at the lower bound

$$I_k(r) \geqslant 0.1093\frac{1}{\sqrt{\lambda-\lambda_2}} - \frac{1}{2}\mathrm{e}^{-(5/6)(\lambda-\lambda_2)} \geqslant \big(0.1093 - 5\mathrm{e}^{-500/6}\big)\frac{1}{\sqrt{\lambda-\lambda_2}} > \frac{0.1}{\sqrt{\lambda-\lambda_2}}.$$

Thus Lemma 4 is proved. □

Proof of Proposition 6. We conclude from Lemmas 3 and 4 that

$$\mathbf{P}\{W = k\} \geqslant \frac{1}{10\sqrt{\lambda-\lambda_2}}\mathrm{e}^{-4(\lambda-k)^2/(\lambda-\lambda_2)} \tag{5.4}$$

for $0 \leqslant \lambda - k \leqslant (\lambda-\lambda_2)/6$ under the assumption $\lambda - \lambda_2 \geqslant 100$.

On the other hand, $f(k) = \mathbf{P}\{Z = k\} \leqslant 1/\sqrt{2\pi k}$, cf. (3.5). Since $k \geqslant \lambda - (\lambda-\lambda_2)/6 \geqslant (5/6)\lambda$, we have

$$f(k) \leqslant \frac{\sqrt{6/5}}{\sqrt{2\pi\lambda}} < \frac{1}{2\sqrt{\lambda}}.$$

As a consequence,

$$\begin{aligned} 1 + \chi^2(W,Z) &\geqslant \sum_{0\leqslant\lambda-k\leqslant\sqrt{\lambda-\lambda_2}/6} \frac{\mathbf{P}\{W=k\}^2}{f(k)} \\ &\geqslant \frac{\sqrt{\lambda}}{50(\lambda-\lambda_2)} \sum_{0\leqslant\lambda-k\leqslant\sqrt{\lambda-\lambda_2}/6} \mathrm{e}^{-8(\lambda-k)^2/(\lambda-\lambda_2)} \geqslant 0.001\sqrt{\frac{\lambda}{\lambda-\lambda_2}}. \end{aligned}$$

To clarify the last inequality, note that the condition $\lambda - \lambda_2 \geqslant 100$ implies that $\lambda > 100$. The last summation is performed over all integers k from the interval $\lambda - \sqrt{\lambda-\lambda_2}/6 \leqslant x \leqslant \lambda$ of length at least $10/6$. It contains at least one integer point, and actually the number of integer points in it is at least $h = \sqrt{\lambda-\lambda_2}/6$. Moreover,

$$\begin{aligned} \sum_{0\leqslant\lambda-k\leqslant h} \mathrm{e}^{-8(\lambda-k)^2/(\lambda-\lambda_2)} &\geqslant \sum_{[\lambda-h]+1\leqslant k\leqslant[\lambda]} \int_{\lambda-k}^{\lambda-k+1} \mathrm{e}^{-8x^2/(\lambda-\lambda_2)}\,\mathrm{d}x = \int_{\lambda-[\lambda]}^{\lambda-[\lambda-h]} \mathrm{e}^{-8x^2/(\lambda-\lambda_2)}\,\mathrm{d}x \\ &\geqslant \frac{1}{4}\sqrt{\lambda-\lambda_2}\int_{2/5}^{2/3} \mathrm{e}^{-y^2/2}\,\mathrm{d}y = \frac{\sqrt{2\pi}}{4}\sqrt{\lambda-\lambda_2}\left(\varPhi\left(\frac{2}{3}\right) - \varPhi\left(\frac{2}{5}\right)\right) \\ &\geqslant 0.056\sqrt{\lambda-\lambda_2}. \end{aligned}$$

Here we used the bounds $4(\lambda-[\lambda])/\sqrt{\lambda-\lambda_2} \leqslant 2/5$ and $4(\lambda-[\lambda-h])/\sqrt{\lambda-\lambda_2} \geqslant 4(\lambda-[\lambda-10/6])/10 \geqslant 2/3$ together with $\varPhi(2/3) - \varPhi(2/5) > 0.09$.

To treat the region $\lambda - \lambda_2 \leqslant 100$, we apply Proposition 2. Let $W_1 = W$ and $W_2 = Y_1 + \cdots + Y_m$, where $Y_1, \dots Y_m$ are independent Bernoulli random variables taking values 1 and 0 with probabilities $1/2$, and $m = 400$. Assume as well that W and W_2 are independent. Then $\tilde{\lambda} = \lambda + m/2$ and $\tilde{\lambda}_2 = \lambda_2 + m/4$ satisfy the condition $\tilde{\lambda} - \tilde{\lambda}_2 \geqslant 100$.

Denote by Z_2 a Poisson random variable with $\mathbf{E}Z_2 = m/2$ independent of $Z_1 = Z$. By the previous step and inequality (2.4) of Proposition 2,

$$0.001\sqrt{\frac{\tilde{\lambda}}{\tilde{\lambda}-\tilde{\lambda}_2}} \leqslant \chi^2(W_1+W_2, Z_1+Z_2)+1 \leqslant \bigl(\chi^2(W_1,Z_1)+1\bigr)\bigl(\chi^2(W_2,Z_2)+1\bigr).$$

Here, by (4.1), $\chi^2(W_2, Z_2) \leqslant 19\sqrt{2}$. Moreover, since $\lambda - \lambda_2 \leqslant 100$, we have

$$\sqrt{\frac{\tilde{\lambda}}{\tilde{\lambda}-\tilde{\lambda}_2}} = \sqrt{\frac{\lambda+m/2}{\lambda-\lambda_2+m/4}} \geqslant \sqrt{\frac{\lambda+200}{200}} \geqslant \frac{1}{10\sqrt{2}}\sqrt{\frac{\lambda}{\max\{1,\lambda-\lambda_2\}}}.$$

It follows that

$$1+\chi^2(W,Z) \geqslant \frac{0.001}{10\sqrt{2}(19\sqrt{2}+1)}\sqrt{\frac{\lambda}{\max\{1,\lambda-\lambda_2\}}} > 2.5\cdot 10^{-6}\sqrt{\frac{\lambda}{\max\{1,\lambda-\lambda_2\}}}.$$

Hence Proposition 6 holds in the case $\lambda - \lambda_2 \leqslant 100$ as well.

6 Lower bound on D

An analogue of Proposition 6 is the following statement for the relative entropy. Recall that $Q = \lambda/\max\{1, \lambda - \lambda_2\}$.

Proposition 7. *If $\lambda_2 \geqslant \kappa_0\lambda$ and $\lambda \geqslant \lambda_0$, then*

$$D(W\,\|\,Z) \geqslant c_0 \log(\mathrm{e}Q), \tag{6.1}$$

where $\kappa_0 = 1-\exp\{-2\cdot 10^7\}$, $\lambda_0 = \exp\{2\cdot 10^7\}$, and $c_0 = \mathrm{e}^{-14}$.

Proof. Let us recall two estimates from the previous section:

$$w_k = \mathbf{P}\{W=k\} \geqslant \frac{1}{10\sqrt{\lambda-\lambda_2}}\mathrm{e}^{-4(\lambda-k)^2/(\lambda-\lambda_2)}, \qquad v_k = \mathbf{P}\{Z=k\} \leqslant \frac{1}{\sqrt{2\pi k}}.$$

The first one is valid under the conditions $0 \leqslant \lambda - k \leqslant (\lambda-\lambda_2)/6$ and $\lambda - \lambda_2 \geqslant 100$; cf. (5.4). Clearly, they are fulfilled if $0 \leqslant \lambda - k \leqslant (5/3)\sqrt{\lambda-\lambda_2}$ and $\lambda-\lambda_2 \geqslant 100$. If additionally $\lambda_2 \geqslant \kappa\lambda$, $0<\kappa<1$, then

$$w_k \geqslant \frac{1}{10\sqrt{\lambda-\lambda_2}}\mathrm{e}^{-100/9} \geqslant \frac{1}{10\sqrt{(1-\kappa)\lambda}}\mathrm{e}^{-100/9}.$$

Since $k \geqslant \frac{5}{6}\lambda$, we also have the upper bound

$$v_k \leqslant \frac{1}{\sqrt{5\pi\lambda/3}}.$$

In order that $w_k \geqslant v_k$, it is therefore sufficient to require that $\mathrm{e}^{-100/9}/(10\sqrt{1-\kappa}) \geqslant 1/\sqrt{5\pi/3}$, that is, $1-\kappa \leqslant (\pi/60)\mathrm{e}^{-200/9}$. Moreover, we have

$$\log\frac{w_k}{v_k} \geqslant \frac{1}{2}\log\frac{\mathrm{e}\lambda}{\lambda-\lambda_2} + \log\biggl(\frac{\sqrt{5\pi/3e}}{10}\mathrm{e}^{-100/9}\biggr) \geqslant \frac{1}{2}\log\frac{\mathrm{e}\lambda}{\lambda-\lambda_2} - 14.$$

Now, applying inequality (2.1) of Proposition 1, we get

$$\begin{aligned} D(W \,\|\, Z) &\geqslant \sum_{w_k \geqslant v_k} w_k \log \frac{w_k}{v_k} - 1 \geqslant \sum_{0 \leqslant \lambda - k \leqslant (5/3)\sqrt{\lambda - \lambda_2}} w_k \log \frac{w_k}{v_k} - 1 \\ &\geqslant \sum_{0 \leqslant \lambda - k \leqslant (5/3)\sqrt{\lambda - \lambda_2}} w_k \bigg(\frac{1}{2} \log \frac{\mathrm{e}\lambda}{\lambda - \lambda_2} - 14\bigg) - 1 \\ &\geqslant \frac{1}{2} \log \frac{\mathrm{e}\lambda}{\lambda - \lambda_2} \sum_{0 \leqslant \lambda - k \leqslant (5/3)\sqrt{\lambda - \lambda_2}} \frac{1}{10\sqrt{\lambda - \lambda_2}} \mathrm{e}^{-4(\lambda - k)^2/(\lambda - \lambda_2)} - 15. \end{aligned}$$

Note that if $\lambda - \lambda_2 \geqslant 100$, then the x-interval $0 \leqslant \lambda - x \leqslant (5/3)\sqrt{\lambda - \lambda_2}$ has length at least $50/3$, so the total number of integer points in this interval is at least $50/3$ as well. Moreover, it is easy to see that the last sum can be bounded from below by

$$\frac{1}{10\sqrt{\lambda - \lambda_2}} \mathrm{e}^{-9} \sum_{0 \leqslant \lambda - k \leqslant (3/2)\sqrt{\lambda - \lambda_2}} > \mathrm{e}^{-11}.$$

Thus

$$D(W \,\|\, Z) \geqslant \frac{1}{2} \mathrm{e}^{-11} \log \frac{\mathrm{e}\lambda}{\lambda - \lambda_2} - 15. \tag{6.2}$$

Moreover, if $\lambda_2 \geqslant \kappa\lambda$ with $\kappa \geqslant \kappa_1 = 1 - \exp\{-60\mathrm{e}^{11}\}$, then

$$\frac{1}{4} \mathrm{e}^{-11} \log \frac{\mathrm{e}\lambda}{\lambda - \lambda_2} \geqslant \frac{1}{4} \mathrm{e}^{-11} \log \frac{1}{1 - \kappa} \geqslant 15,$$

and (6.2) yields

$$D(W \,\|\, Z) \geqslant \frac{1}{4} \mathrm{e}^{-11} \log \frac{\mathrm{e}\lambda}{\lambda - \lambda_2}. \tag{6.3}$$

The proposition is thus proved under the conditions $\lambda - \lambda_2 \geqslant 100$ and $\lambda_2 \geqslant \kappa\lambda$ with $\kappa_1 \leqslant \kappa < 1$. It remains to eliminate the first condition by assuming that $\lambda - \lambda_2 < 100$ and again that $\lambda_2 \geqslant \kappa\lambda$ with κ being sufficiently close to 1. To this aim, we appeal to Proposition 2 again like in the last step of the proof of Proposition 6. Namely, using the same notations and assumptions, from inequality (2.3) and using (6.3), we obtain that

$$\begin{aligned} \frac{1}{4} \mathrm{e}^{-11} \log \frac{\mathrm{e}\tilde{\lambda}}{\max\{1, \tilde{\lambda} - \tilde{\lambda}_2\}} &\leqslant D(W_1 + W_2 \,\|\, Z_1 + Z_2) \\ &\leqslant D(W_1 \,\|\, Z_1) + D(W_2 \,\|\, Z_2), \end{aligned} \tag{6.4}$$

where $W_1 = W$ and $Z_1 = Z$, which holds as long as $\tilde{\lambda}_2 \geqslant \kappa\tilde{\lambda}$, that is, $\lambda_2 + m/4 \geqslant \kappa(\lambda + m/2)$. Since $\lambda - \lambda_2 < 100$, the latter follows from

$$\lambda - 100 + \frac{m}{4} \geqslant \kappa\Big(\lambda + \frac{m}{2}\Big),$$

which is solved as

$$\lambda \geqslant 200 \frac{\kappa}{1 - \kappa}.$$

Moreover, by (4.2) we have $D(W_2 \,\|\, Z_2) \leqslant 23\log(2\mathrm{e})$. This bound may be used in (6.4), which gives

$$D(W \,\|\, Z) \geqslant \frac{1}{4}\mathrm{e}^{-11}\log\frac{\mathrm{e}\tilde{\lambda}}{\max\{1,\, \tilde{\lambda}-\tilde{\lambda}_2\}} - 23\log(2\mathrm{e}) \geqslant \frac{1}{8}\mathrm{e}^{-11}\log\frac{\mathrm{e}\tilde{\lambda}}{\max\{1,\, \tilde{\lambda}-\tilde{\lambda}_2\}},$$

where the second inequality holds when $1-\kappa$ is sufficiently small. Namely,

$$\frac{1}{8}\mathrm{e}^{-11}\log\frac{\mathrm{e}\tilde{\lambda}}{\tilde{\lambda}-\tilde{\lambda}_2} \geqslant \frac{1}{8}\mathrm{e}^{-11}\log\frac{1}{1-\kappa} \geqslant 23\log(2\mathrm{e})$$

if $\tilde{\lambda}_2 \geqslant \kappa\tilde{\lambda}$ and $1-\kappa \leqslant \exp\{-8\cdot 23\cdot\log(2\mathrm{e})\cdot\mathrm{e}^{11}\}$. Since the product in the exponent is smaller than $1.87\cdot 10^7$, we may choose $\kappa = 1-\exp\{-1.87\cdot 10^7\} > \kappa_1$. In this case,

$$D(W \,\|\, Z) \geqslant c_1\log\frac{\mathrm{e}\tilde{\lambda}}{\tilde{\lambda}-\tilde{\lambda}_2}, \quad c_1 = \frac{1}{8}\mathrm{e}^{-11},$$

provided that $\lambda \geqslant 200\kappa/(1-\kappa)$. However,

$$\log\frac{\mathrm{e}\tilde{\lambda}}{\tilde{\lambda}-\tilde{\lambda}_2} = \log\frac{\mathrm{e}(\lambda+200)}{\lambda-\lambda_2+100} \geqslant \frac{1}{2}\log\frac{\mathrm{e}\lambda}{\max\{1,\, \lambda-\lambda_2\}}$$

for all $\lambda \geqslant 4\cdot 10^4$. It remains to note that $200\kappa/(1-\kappa) < \lambda_0$, $\kappa < \kappa_0$, $c_1/2 > c_0$. $\square$

7 Proof of Theorem 1

Let us summarize. Using the quantity

$$F = F(\lambda, \lambda_2) = \frac{\max(1, \lambda)}{\max(1,\, \lambda-\lambda_2)},$$

the results on Poisson approximation obtained for different regions of λ and λ_2 can be combined in the form of the following two-sided bounds:

$$c_1\left(\frac{\lambda_2}{\lambda}\right)^2(1+\log F) \leqslant D(W \,\|\, Z) \leqslant c_2\left(\frac{\lambda_2}{\lambda}\right)^2(1+\log F), \tag{7.1}$$

$$c_1\left(\frac{\lambda_2}{\lambda}\right)^2\sqrt{F} \leqslant \chi^2(W, Z) \leqslant c_2\left(\frac{\lambda_2}{\lambda}\right)^2\sqrt{F}, \tag{7.2}$$

which are valid up to some absolute positive constants c_1 and c_2. Let us describe the proof of Theorem 1 and provide explicit values for these constants. As we will see, (7.1)–(7.2) hold with $c_1 = 10^{-8}$ and $c_2 = 5.6\cdot 10^7$.

The upper bound in (7.1). If $\lambda \leqslant 1/2$, then these bounds simplify and are made precise via

$$\frac{1}{4}\left(\frac{\lambda_2}{\lambda}\right)^2 \leqslant D(W \,\|\, Z) \leqslant \chi^2(W, Z) \leqslant 15\left(\frac{\lambda_2}{\lambda}\right)^2. \tag{7.3}$$

Here, the left inequality holds for all λ and λ_2 (see [7]), whereas the right inequality is part of Proposition 3. Note that $\lambda \leqslant 1/2$ implies $\lambda_2 \leqslant \lambda/2$.

If $\lambda \geqslant 1/2$ and $\lambda_2 \leqslant \lambda/2$, then by Proposition 4 we have

$$D(W \,\|\, Z) \leqslant \chi^2(W, Z) \leqslant 56 \cdot 10^6 \Big(\frac{\lambda_2}{\lambda}\Big)^2,$$

so that

$$D(W \,\|\, Z) \leqslant 56 \cdot 10^6 \Big(\frac{\lambda_2}{\lambda}\Big)^2 (1 + \log F). \tag{7.4}$$

In the case where $\lambda \geqslant 1/2$ and $\lambda_2 > \lambda/2$, we may apply (4.2), which gives

$$D(W \,\|\, Z) \leqslant 23(1 + \log F) \leqslant 4 \cdot 23 \Big(\frac{\lambda_2}{\lambda}\Big)^2 (1 + \log F).$$

Here the right-hand side contains a better numerical constant in comparison with (7.4), and we finally get (7.1) with a constant $c_2 = 56 \cdot 10^6$.

The lower bound in (7.1). If $\lambda \leqslant 1$, then $F = 1$, so that the lower bound in (7.3) yields (7.1) with $c_1 = 1/4$. If $\lambda \geqslant 1$, then inequality (7.4) may be reversed by (6.1), which gives

$$D(W \,\|\, Z) \geqslant c_0(1 + \log F) \geqslant c_0 \Big(\frac{\lambda_2}{\lambda}\Big)^2 (1 + \log F) \tag{7.5}$$

with $c_0 = \mathrm{e}^{-14}$, provided that $\lambda_2 \geqslant \kappa_0 \lambda$ and $\lambda \geqslant \lambda_0$, where $\kappa_0 = 1 - \exp\{-2 \cdot 10^7\}$ and $\lambda_0 = \exp\{2 \cdot 10^7\}$. But the remaining regions belong to the nondegenerate case, where F is bounded by a quantity that depends on κ_0 or λ_0. Indeed, if $\lambda_2 \leqslant \kappa_0 \lambda$, then $\log F \leqslant -\log(1 - \kappa_0) = 2 \cdot 10^7$, so

$$D(W \,\|\, Z) \geqslant \frac{1}{4(1 + 2 \cdot 10^7)} \Big(\frac{\lambda_2}{\lambda}\Big)^2 (1 + \log F).$$

This means that the left inequality in (7.1) holds with a constant $c_1 = (1 + 2 \cdot 10^7)/4$, which is smaller than c_0 in the analogous inequality (7.5). Similarly, if $1 \leqslant \lambda < \lambda_0$, then $F \leqslant \lambda < \lambda_0$, and by the lower bound in (7.3) we get

$$D(W \,\|\, Z) \geqslant \frac{1}{4(1 + \log \lambda_0)} \Big(\frac{\lambda_2}{\lambda}\Big)^2 (1 + \log F).$$

This means that the left inequality in (7.1) holds with the same constant c_1 as before. Thus the lower bound in (7.1) holds with constant c_1 ($> 10^{-8}$).

The upper bound in (7.2). If $\lambda \leqslant 1/2$, then we have (7.3), which implies (7.2) with $c_2 = 15$.

If $\lambda \geqslant 1/2$ and $\lambda_2 \leqslant \lambda/2$, then a stronger version of (7.4) is provided by Proposition 4, which gives

$$\chi^2(W, Z) \leqslant 56 \cdot 10^6 \Big(\frac{\lambda_2}{\lambda}\Big)^2,$$

so that (7.2) holds with $c_2 = 56 \cdot 10^6$. In the case where $\lambda \geqslant 1/2$ and $\lambda_2 > \lambda/2$, we may apply (4.1), which gives

$$\chi^2(W, Z) \leqslant 76 \Big(\frac{\lambda_2}{\lambda}\Big)^2 \sqrt{F}.$$

Here the right-hand side contains a better numerical constant, and we finally get (7.2) with the same constant c_2 as in (7.1).

The lower bound in (7.2). If $\lambda \leqslant 1$, then $F = 1$, so that the lower bound in (7.3) yields (7.1) with $c_1 = 1/4$. Assume that $\lambda \geqslant 1$, in which case $F = Q = \lambda/\max(1,\ \lambda - \lambda_2)$. By (5.2) we have

$$\chi^2(W, Z) \geqslant \frac{c_0}{9}\sqrt{F}$$

with $c_0 = 2.5 \cdot 10^{-6}$, provided that $\lambda_2 \geqslant \kappa_0 \lambda$, $\kappa_0 = 1 - c_0^2/4$. This gives

$$\chi^2(W, Z) \geqslant \frac{c_0}{9}\left(\frac{\lambda_2}{\lambda}\right)^2 \sqrt{F}, \tag{7.6}$$

and we obtain the left inequality in (7.2) with $c_1 = c_0/9 > 10^{-7}$.

The remaining region belongs to the nondegenerate case, where F is bounded. Indeed, if $\lambda_2 \leqslant \kappa_0 \lambda$, then $1/\sqrt{F} \geqslant \sqrt{1 - \kappa_0} = c_0/2 = 0.8 \cdot 10^{-6}$, so that, by the left inequality in (7.3),

$$\chi^2(W, Z) \geqslant \frac{1}{4}\left(\frac{\lambda_2}{\lambda}\right)^2 \geqslant 0.2 \cdot 10^{-6}\left(\frac{\lambda_2}{\lambda}\right)^2 \sqrt{F}.$$

This means that the left inequality in (7.1) holds with constant $c_1 = 2 \cdot 10^{-7}$, which is slightly better than the constant in the analogous inequality (7.6). Thus the lower bound in (7.2) holds with constant $c_1 = 10^{-7}$. □

8 Tsallis versus Vajda–Pearson

We now turn to the Tsallis relative entropies of other indexes. To make an application of nonuniform bounds more convenient, let us first relate T_α to the Vajda–Pearson distance

$$\chi_\alpha(X, Z) = \int \left|\frac{p - q}{q}\right|^\alpha q \,\mathrm{d}\pi.$$

It is defined for arbitrary random elements X and Z in a measure space (Ω, π) whose distributions are absolutely continuous and have densities p and q, respectively, with respect to the measure π on Ω (the definition does not depend on the choice of π).

Recall that

$$T_\alpha(X \,\|\, Z) = \frac{1}{\alpha - 1}\left[\int \left(\frac{p}{q}\right)^\alpha q \,\mathrm{d}\pi - 1\right],$$

so that $T_2 = \chi_2$ is the classical Pearson distance, and note that $T_\alpha = \chi_\alpha = \infty$ as long as the distribution of X is not absolutely continuous with respect to the distribution of Z. We need the following auxiliary result.

Proposition 8. *For any* $\alpha \geqslant 2$,

$$T_\alpha(W \,\|\, Z) \leqslant \frac{2^\alpha}{\alpha - 1}\bigl(T_2(W \,\|\, Z) + \chi_\alpha(W, Z)\bigr).$$

Proof. We may assume that the distribution of X is absolutely continuous with respect to the distribution of Z with $\chi_\alpha(W, Z) < \infty$. In this case the (nonnegative) function $\xi = p/q$ is well defined a.e. with respect to the probability measure $Q = q \,\mathrm{d}\pi$. We consider it as a random variable on the probability space (Ω, Q) with finite moment of order α. Note that

$$(\alpha - 1)T_\alpha(W \,\|\, Z) = \mathbf{E}\bigl(\xi^\alpha - 1\bigr) \quad \text{and} \quad \chi_\alpha(W, Z) = \mathbf{E}|\xi - 1|^\alpha.$$

Putting $\eta = \xi - 1 \geqslant -1$, define the function $\psi(t) = \mathbf{E}(1+t\eta)^\alpha - 1$, $t \geqslant 0$, so that $\psi(1) = (\alpha-1)T_\alpha(W \,\|\, Z)$. By the integral Taylor formula,

$$\psi(1) = \alpha(\alpha-1)\mathbf{E}\eta^2 \int_0^1 (1-t)(1+t\eta)^{\alpha-2}\,\mathrm{d}t.$$

Introducing the sets $A = \{\xi \leqslant 2\} = \{\eta \leqslant 1\}$ and $B = \{\xi > 2\} = \{\eta > 1\}$, we have

$$\mathbf{E}\mathbf{1}_A\eta^2 \int_0^1 (1-t)(1+t\eta)^{\alpha-2}\,\mathrm{d}t \leqslant \mathbf{E}\eta^2 \int_0^1 (1-t)(1+t)^{\alpha-2}\,\mathrm{d}t \leqslant \frac{2^\alpha}{\alpha(\alpha-1)} T_2(W \,\|\, Z)$$

and

$$\mathbf{E}\mathbf{1}_B\eta^2 \int_0^1 (1-t)(1+t\eta)^{\alpha-2}\,\mathrm{d}t \leqslant \mathbf{E}\mathbf{1}_B\eta^\alpha \int_0^1 (1-t)(1+t)^{\alpha-2}\,\mathrm{d}t \leqslant \frac{2^\alpha}{\alpha(\alpha-1)} \chi_\alpha(W, Z).$$

We obtain the statement of the proposition from the last two bounds. □

9 Estimates of Vajda–Pearson distances

For the proof of Theorem 2, we need the following propositions. We thus return to the setting of Bernoulli trials. Let us denote by $c(\alpha)$ a positive constant depending on α only, which may vary from place to place.

Proposition 9. *For $\alpha > 1$ and $\lambda \leqslant 1/2$, we have*

$$\chi_\alpha(W, Z) \leqslant c(\alpha) \frac{\lambda_2^\alpha}{\lambda^{2(\alpha-1)}}.$$

Proof. Applying Lemmas III.1–2 and repeating the argument used in the proof of Proposition III.4 from [3], we obtain that

$$\begin{aligned} \frac{\mathrm{e}^\lambda}{\lambda_2^\alpha}\chi_\alpha(W, Z) &\leqslant 1 + \lambda\Big(\frac{\lambda + \mathrm{e} - 1}{\lambda}\Big)^\alpha \\ &\quad + 3^{\alpha-1}\sum_{k=2}^\infty \frac{\lambda^k}{k!}\Big(1 + \Big(\frac{\mathrm{e}^\lambda - 1}{\lambda}\Big)^\alpha\Big(\Big(\frac{k}{\lambda}\Big)^\alpha + \Big(\frac{k(k-1)}{\lambda^2}\Big)^\alpha\Big)\Big) \\ &\leqslant c(\alpha)\Big(1 + \frac{1}{\lambda^{\alpha-1}} + \frac{1}{\lambda^{2(\alpha-1)}}\Big). \qquad \square \end{aligned}$$

Proposition 10. *Let $\alpha > 1$. If $\lambda \geqslant 1/2$ and $\lambda_2 \leqslant \kappa\lambda$ with $\kappa \in (0,1)$, then*

$$\chi_\alpha(W, Z) \leqslant \frac{c(\alpha)}{(1-\kappa)^{3\alpha/2}}\Big(\frac{\lambda_2}{\lambda}\Big)^\alpha.$$

Proof. Write

$$\chi_\alpha(W, Z) = \sum_{k=0}^\infty \frac{|\Delta_k|^\alpha}{f(k)^{\alpha-1}} = S_1 + S_2 = \Big(\sum_{k=0}^{[2\lambda]} + \sum_{k=[2\lambda]+1}^\infty\Big)\frac{|\Delta_k|^\alpha}{f(k)^{\alpha-1}}.$$

In the range $0 \leqslant k \leqslant [2\lambda]$, we apply inequality (VI.2) from [3], which gives

$$|\Delta_k|^\alpha \leqslant \frac{c(\alpha)}{(1-\kappa)^{3\alpha/2}} \left(\frac{|k-\lambda|^{2\alpha}}{\lambda^\alpha} + 1\right) \left(\frac{\lambda_2}{\lambda}\right)^\alpha f^\alpha(k).$$

Therefore

$$S_1 \leqslant \frac{c(\alpha)}{(1-\kappa)^{3/2}} \left(\frac{\mathbf{E}|Z-\lambda|^{2\alpha}}{\lambda^\alpha} + 1\right) \left(\frac{\lambda_2}{\lambda}\right)^\alpha \leqslant \frac{c(\alpha)}{(1-\kappa)^{3\alpha/2}} \left(\frac{\lambda_2}{\lambda}\right)^\alpha,$$

where we used the upper bound $\mathbf{E}|Z-\lambda|^{2\alpha} \leqslant c(\alpha)\lambda^\alpha$.

To estimate S_2, we use inequalities (VI.3) and (II.1) from [3] to get

$$\begin{aligned} S_2 &\leqslant \frac{c(\alpha)}{(1-\kappa)^{3\alpha/2}} \sum_{k=[2\lambda]+1}^{\infty} \left(\frac{k}{\lambda}\right)^{3\alpha} \lambda_2^\alpha f(k) \\ &\leqslant \frac{c(\alpha)}{(1-\kappa)^{3\alpha/2}} \lambda_2^\alpha f\big([2\lambda]+1\big) \sum_{k=0}^{\infty} \left(1 + \left(\frac{k}{\lambda}\right)^{3\alpha}\right) \frac{1}{2^k} \\ &\leqslant \frac{c(\alpha)}{(1-\kappa)^{2\alpha/2}} \lambda_2^\alpha f\big([2\lambda]+1\big) \leqslant \frac{c(\alpha)}{(1-\kappa)^{3\alpha/2}} \lambda_2^\alpha \mathrm{e}^{\lambda \log(\mathrm{e}/4)/2}. \end{aligned}$$

The statement of the proposition immediately follows from the last two estimates. □

10 Proof of Theorem 2

To complete the proof of Theorem 2, we need the following two lemmas. Recall that $Q = \lambda / \max\{1,\ \lambda - \lambda_2\}$.

Lemma 5. *For $\alpha > 1$ and $\lambda \geqslant 1/2$,*

$$T_\alpha(W \,\|\, Z) \leqslant c(\alpha) Q^{(\alpha-1)/2}.$$

Proof. By the definition of the Tsallis distance we have

$$\begin{aligned} T_\alpha(W \,\|\, Z) &\leqslant \frac{1}{\alpha-1} \sum_{k=0}^{\infty} \left(\frac{w_k}{v_k}\right)^\alpha v_k \leqslant S_1 + S_2 + S_3 \\ &= \frac{1}{\alpha-1} \left(\sum_{1 \leqslant k < \lambda/4} + \sum_{\lambda/4 \leqslant k \leqslant 4\lambda} + \sum_{k > 4\lambda}\right) \left(\frac{w_k}{v_k}\right)^{\alpha-1} w_k. \end{aligned}$$

By (4.5) we get

$$(\alpha-1) S_2 \leqslant \sum_{\lambda/4 \leqslant k \leqslant 4\lambda} \big(2\mathrm{e}\, Q_0^{1/2} k^{1/2}\big)^{\alpha-1} w_k \leqslant (4\mathrm{e})^{\alpha-1} Q^{(\alpha-1)/2}.$$

Using (4.6) and repeating the argument of Section 4, we obtain the upper bounds $S_1 + S_3 \leqslant c(\alpha)$. The three last estimates give the statement of the proposition. □

Lemma 6. *For $\alpha > 1$ and $\lambda \geqslant 1/2$, with some constant $c_1(\alpha) \in (0,1)$,*

$$1 + T_\alpha(W \,\|\, Z) \geqslant c(\alpha) Q^{(\alpha-1)/2}. \tag{10.1}$$

Moreover,

$$T_\alpha(W \,\|\, Z) \geqslant \frac{c_1(\alpha)}{9} Q^{(\alpha-1)/2} \tag{10.2}$$

as long as $\lambda_2 \geqslant (1 - c_1(\alpha)^2/4)\lambda$.

Proof. Bound (10.2) follows from bound (10.1) in the same way as (5.2) follows from (5.1). Therefore we omit the proof.

To prove (10.1), we use the lower bound (5.4). Repeating the argument of the proof of Proposition 6, we easily obtain the lower bound under the assumption $\lambda - \lambda_2 \geqslant 100$:

$$\begin{aligned} 1 + T_\alpha(W \,\|\, Z) &\geqslant \sum_{0 \leqslant \lambda - k \leqslant \sqrt{\lambda - \lambda_2}/6} \left(\frac{w_k}{v_k}\right)^{\alpha-1} w_k \\ &\geqslant \frac{1}{10} \cdot \left(\frac{1}{5}\right)^{\alpha-1} \left(\frac{\lambda}{\lambda - \lambda_2}\right)^{(\alpha-1)/2} \frac{1}{\sqrt{\lambda - \lambda_2}} \sum_{0 \leqslant \lambda - k \leqslant \sqrt{\lambda - \lambda_2}/6} \mathrm{e}^{-4\alpha(\lambda-k)^2/(\lambda-\lambda_2)} \\ &\geqslant c(\alpha)\left(\frac{\lambda}{\lambda - \lambda_2}\right)^{(\alpha-1)/2}. \end{aligned}$$

To treat the region $\lambda - \lambda_2 \leqslant 100$, we refer to Johnson [10, pp. 133–134] and repeat the argument at the end of Section 5. □

Proof of Theorem 2. Assuming that $\lambda_2/\lambda \leqslant 1 - c_1(\alpha)^2/4$, we have $F \sim 1$ with involved constants depending on α, and then we need to show that $T_\alpha(W \,\|\, Z) \sim (\lambda_2/\lambda)^2$.

In the case $1 < \alpha \leqslant 2$, we have

$$\frac{1}{4}\left(\frac{\lambda_2}{\lambda}\right)^2 \leqslant T_\alpha(W \,\|\, Z) \leqslant T_2(W \,\|\, Z) = \chi^2(W, Z).$$

Turning to the case $\alpha \geqslant 2$, first let $\lambda \leqslant 1/2$. Since $\lambda_2 \leqslant \lambda^2$, by Propositions 8 and 9 we have

$$\begin{aligned} T_\alpha(W \,\|\, Z) &\leqslant c(\alpha)\big(T_2(W \,\|\, Z) + \chi_\alpha(W, Z)\big) \\ &\leqslant c(\alpha)\left(T_2(W \,\|\, Z) + \frac{\lambda_2^\alpha}{\lambda^{2(\alpha-1)}}\right) \leqslant c(\alpha)\left(\frac{\lambda_2}{\lambda}\right)^2. \end{aligned}$$

Now let $\lambda \geqslant 1/2$. Then by Propositions 8 and 10 we conclude that

$$\begin{aligned} T_\alpha(W \,\|\, Z) &\leqslant c(\alpha)\big(T_2(W \,\|\, Z) + \chi_\alpha(W, Z)\big) \\ &\leqslant c(\alpha)\left(T_2(W \,\|\, Z) + \left(\frac{\lambda_2}{\lambda}\right)^\alpha\right) \leqslant c(\alpha)\left(\frac{\lambda_2}{\lambda}\right)^2. \end{aligned}$$

It remains to consider the region $\lambda_2/\lambda \geqslant 1 - c_1(\alpha)^2/4$. But in this case the statement of the theorem immediately follows from Lemmas 5 and 6. □

11 Difference of entropies

For the proof of Corollary 1, we will use the other functional

$$H_2(Z) = \big(\mathbf{E}\log^2 v(Z)\big)^{1/2} = \bigg(\sum_k v_k \log^2 v_k\bigg)^{1/2}, \quad v_k = \mathbf{P}\{Z = k\},$$

where Z is an integer-valued random variable. Thus, whereas the Shannon entropy $H(Z) = -\mathbf{E}\log v(Z)$ describes the average of the informational content $-\log v(Z)$, the informational quantity $H_2(Z)$ represents the second moment of this random variable.

An application of Theorem 1 is based upon the following elementary relation.

Proposition 11. *For all integer-valued random variables W and Z with finite entropies, we have*

$$H(W \,\|\, Z) \leqslant \chi^2(W,Z) + H_2(Z)\sqrt{\chi^2(W,Z)}. \tag{11.1}$$

Proof. We may assume that the distribution of W is absolutely continuous with respect to the distribution of Z (since otherwise $\chi^2(W,Z) = \infty$). Equivalently, for all $k \in \mathbb{Z}$, $v_k = 0 \Rightarrow w_k = 0$, where $w_k = \mathbf{P}\{W = k\}$. Define $t_k = w_k/v_k$ in case $v_k > 0$. Recalling definition (1.7), we then have

$$H(W \,\|\, Z) = \sum_{v_k>0} (t_k \log t_k) v_k + \sum_{v_k>0} (t_k - 1) v_k \log v_k.$$

We now apply the inequality $t\log t \leqslant (t-1) + (t-1)^2$ ($t \geqslant 0$), obtaining

$$\begin{aligned} H(W \,\|\, Z) &\leqslant \sum_{v_k>0} (t_k - 1)v_k + \sum_{v_k>0} (t_k-1)^2 v_k + \sum_{v_k>0} (t_k-1) v_k \log v_k \\ &= \sum_k \frac{(w_k - v_k)^2}{v_k} + \sum_{v_k>0} (w_k - v_k)\log v_k. \end{aligned}$$

Here the first sum in the last bound is exactly $\chi^2(W,Z)$, whereas, by Cauchy's inequality, the square of the last sum is bounded from above by

$$\sum_k \frac{(w_k - v_k)^2}{v_k} \sum_k v_k \log^2 v_k = \chi^2(W,Z) H_2^2(Z). \qquad \square$$

In view of (11.1), we also need the following:

Proposition 12. *If Z has a Poisson distribution with parameter λ, then*

$$H_2(Z) \leqslant \begin{cases} \sqrt{50}\log(1+\lambda) & \text{if } \lambda \geqslant 1, \\ 5\sqrt{\lambda}\log(\mathrm{e}/\lambda) & \text{if } \lambda \leqslant 1. \end{cases}$$

Proof. Put $v_k = \mathbf{P}\{Z = k\}$. In particular, $v_0 \log^2 v_0 = \lambda^2 \mathrm{e}^{-\lambda}$ and $v_1 \log^2 v_1 = \lambda \mathrm{e}^{-\lambda}(\lambda + \log(1/\lambda))^2$. This shows that the stated upper bound for small λ can be reversed up to a constant. For $\lambda \leqslant 1$, given $k \geqslant 1$, from

$$\log\frac{1}{v_k} = \lambda + \log k! + k\log\frac{1}{\lambda} \leqslant k^2 \log\frac{\mathrm{e}}{\lambda}$$

we get

$$\sum_{k\geqslant 1} v_k \log^2 v_k \leqslant \mathbf{E}Z^4 \log^2 \frac{\mathrm{e}}{\lambda} \leqslant 24\lambda \log^2 \frac{\mathrm{e}}{\lambda}.$$

Hence $H_2^2(Z) \leqslant 25\lambda \log^2(\mathrm{e}/\lambda)$, thus proving the second upper bound of the lemma.

Now, assuming that $\lambda \geqslant 1$, let us apply the lower bounds (3.6)–(3.7) from Lemma 1, which for all $k \geqslant 1$ give

$$\log \frac{1}{v_k} \leqslant 1 + \frac{1}{2}\log k + \frac{1}{\lambda}(k-\lambda)^2 \leqslant \log(\mathrm{e}k) + \frac{1}{\lambda}(k-\lambda)^2$$

and

$$\log^2 \frac{1}{v_k} \leqslant 2\log^2\big(\mathrm{e}(k+1)\big) + \frac{2}{\lambda^2}(k-\lambda)^4.$$

Note that this bound is also true for $k=0$. Using the concavity of the function $\log^2 x$ in $x \geqslant e$ and applying Jensen's inequality, we therefore obtain that

$$\begin{aligned}\sum_{k=0}^{\infty} v_k \log^2 v_k &\leqslant 2\mathbf{E}\log^2\big(\mathrm{e}(Z+1)\big) + \frac{2}{\lambda^2}\mathbf{E}(Z-\lambda)^4 \\ &\leqslant 2\log^2\big(\mathrm{e}(\lambda+1)\big) + \frac{6(\lambda+2)}{\lambda} \\ &\leqslant 2\big(1+\log(1+\lambda)\big)^2 + 18.\end{aligned}$$

Hence $H_2(Z) \leqslant Cx$, $x = \log(1+\lambda) \geqslant \log 2$, with $C^2 = 2(1+1/x)^2 + 18/x^2 < 50$.

Applying the upper bound (3.6) from Lemma 1, we also see that this upper bound on H_2 can be reversed up to a constant as well. $\square$

Remark 1. With similar arguments, it follows that

$$H(Z) \leqslant \begin{cases} c\log(1+\lambda) & \text{if } \lambda \geqslant 1, \\ c\lambda\log(\mathrm{e}/\lambda) & \text{if } \lambda \leqslant 1, \end{cases}$$

which can be reversed modulo an absolute factor $c>0$. Hence $H_2(Z) \sim H(Z)$ as long as λ stays bounded away from zero.

Proof of Corollary 1. By Theorem 1 with W as in (1.1) and with a Poisson random variable Z with parameter λ, we have

$$\chi^2(W,Z) \leqslant C\left(\frac{\lambda_2}{\lambda}\right)^2 \sqrt{2+\lambda}$$

up to some absolute constant C. Using this estimate in (11.1) and applying Proposition 12, we immediately get the desired inequality (1.8) (in view of $\lambda_2 \leqslant \lambda$).

To derive a more precise inequality illustrating the asymptotic behavior in λ in the typical case $\lambda_2 \leqslant \lambda/2$, let us apply once more Theorem 1 with its sharper bound

$$\chi^2(W,Z) \leqslant C\left(\frac{\lambda_2}{\lambda}\right)^2,$$

as in Proposition 3. By Proposition 11 this gives

$$H(W \,\|\, Z) \leqslant C\big(1 + H_2(Z)\big)\frac{\lambda_2}{\lambda}.$$

It remains to note that $1 + H_2(Z) \leqslant C \log(2 + \lambda)$ by Proposition 12. $\square$

Acknowledgments. We would like to thank the referee for drawing our attention to the work by V. Zacharovas and H.-K. Hwang. Thanks also to A. Zaitsev for drawing our attention to the work by I.S. Borisov and I.S. Vorozheykin.

References

1. K.F. Andersen, Weighted inequalities for iterated convolutions, *Proc. Am. Math. Soc.*, **127**(9):2643–2651, 1999.
2. A.D. Barbour and P. Hall, On the rate of Poisson convergence, *Math. Proc. Camb. Philos. Soc.*, **95**(3):473–480, 1984.
3. S.G. Bobkov, G.P. Chistyakov, and F. Götze, Non-uniform bounds in the Poisson approximation with applications to informational distances. I, *IEEE Trans Inf. Theory*, **65**(9):5283–5293, 2019.
4. S.G. Bobkov, G.P. Chistyakov, and F. Götze, Rényi divergence and the central limit theorem, *Ann. Probab.*, **47**(1): 270–323, 2019.
5. I.S. Borisov and I.S. Vorozheykin, Accuracy of approximation in the Poisson theorem in terms of χ^2 distance, *Sib. Mat. Zh.*, **49**(1):8–22, 2008 (in Russian). English transl.: *Sib. Math. J.*, **49**(1):5–17, 2008.
6. P. Harremoës, Binomial and Poisson distributions as maximum entropy distributions, *IEEE Trans. Inf. Theory*, **47**(5):2039–2041, 2001.
7. P. Harremoës, O. Johnson, and I. Kontoyiannis, Thinning and information projections, 2016, arXiv:1601.04255.
8. P. Harremoës and P.S. Ruzankin, Rate of convergence to Poisson law in terms of information divergence, *IEEE Trans. Inf. Theory*, **50**(9):2145–2149, 2004.
9. E. Hillion and O. Johnson, A proof of the Shepp–Olkin entropy concavity conjecture, *Bernoulli*, **23**(4B):3638–3649, 2017.
10. O. Johnson, *Information Theory and Central Limit Theorem*, Imperial College Press, London, 2004.
11. I. Sason, Entropy bounds for discrete random variables via maximal coupling, *IEEE Trans. Inf. Theory*, **59**(11):7118–7131, 2013.
12. T. van Erven and P. Harremoës, Rényi divergence and Kullback–Leibler divergence, *IEEE Trans. Inf. Theory*, **60**(7):3797–3820, 2014.
13. V. Zacharovas and H.-K. Hwang, A Charlier–Parseval approach to Poisson approximation and its applications, *Lith. Math. J.*, **50**(1):88–119, 2010.