NON-UNIFORM BOUNDS IN THE POISSON APPROXIMATION WITH APPLICATIONS TO INFORMATIONAL DISTANCES

S. G. BOBKOV[1], G. P. CHISTYAKOV[2], AND F. GÖTZE[2]

Abstract. We explore asymptotically optimal bounds for deviations of Bernoulli convolutions from the Poisson limit in terms of the Shannon relative entropy and the Pearson χ^2-distance. The results are based on proper non-uniform estimates for densities.

1. Introduction

Let $X_1, \dots, X_n$ be independent Bernoulli random variables taking the two values, 1 (interpreted as a success) and 0 (as a failure) with respective probabilities p_j and $q_j = 1 - p_j$. The total number of successes $W = X_1 + \dots + X_n$ takes values $k = 0, 1, \dots, n$ with probabilities

$$\mathbb{P}\{W = k\} = \sum p_1^{\varepsilon_1} q_1^{1-\varepsilon_1} \dots p_n^{\varepsilon_n} q_n^{1-\varepsilon_n}, \tag{1.1}$$

where the summation runs over all 0-1 sequences $\varepsilon_1, \dots, \varepsilon_n$ such that $\varepsilon_1 + \dots + \varepsilon_n = k$. Although this expression is difficult to determine in case of arbitrary p_j and large n, it can be well approximated by the Poisson probabilities under quite general assumptions. Putting

$$\lambda = p_1 + \dots + p_n,$$

let Z be a Poisson random variable with parameter $\lambda > 0$ (for short, $Z \sim P_\lambda$), i.e.,

$$\mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad k = 0, 1, \dots$$

It is well-known for a long time that, if $\max_{j \le n} p_j$ is small, the distribution P_λ approximates the distribution P_W of W, which may be quantified by means of the total variation distance

$$\begin{aligned} d(W,Z) &= \|P_W - P_\lambda\|_{\rm TV} \\ &= 2 \sup_{A \subset \mathbb{Z}} |\,\mathbb{P}\{W \in A\} - \mathbb{P}\{Z \in A\}| \,=\, \sum_{k=0}^\infty |w_k - v_k|, \end{aligned}$$

where $w_k = \mathbb{P}\{W = k\}$ and $v_k = \mathbb{P}\{Z = k\}$. In particular, based on the Stein-Chen method, there is the following remarkable two-sided bound due to Barbour and Hall involving the functional

$$\lambda_2 = p_1^2 + \dots + p_n^2.$$

1991 *Mathematics Subject Classification.* Primary 60E, 60F.

Key words and phrases. χ^2-divergence, Relative entropy, Poisson approximation.

1) School of Mathematics, University of Minnesota, USA; research was partially supported by SFB 1283, Humboldt Foundation, and NSF grant DMS-1855575.

2) Faculty of Mathematics, University of Bielefeld, Germany; research was partially supported by SFB 1283.

Theorem 1.1 [2]. *One has*

$$\frac{1}{32}\min(1,1/\lambda)\,\lambda_2 \,\leq\, \frac{1}{2}\,d(W,Z) \,\leq\, \frac{1-e^{-\lambda}}{\lambda}\,\lambda_2. \tag{1.2}$$

Here, the parameter λ_2, or more precisely – the ratio λ_2/λ (for λ bounded away from zero), plays a similar role as the Lyapunov ratio L_3 in the central limit theorem.

In the i.i.d. case with $p_j = \lambda/n$ and fixed $\lambda > 0$, both sides of (1.2) are of the same order $1/n$. In the case $\lambda \leq 1$, the upper bound in (1.2) is sharp also in the sense that the second inequality becomes an equality for $p_1 = 1$, $p_j = 0$ $(2 \leq j \leq n)$.

Theorem 1.1 refined many previous results in this direction, starting from bounds for the i.i.d. case by Prokhorov [19] and bounds for the general case by Le Cam [16]. In particular, Le Cam obtained the upper bound

$$d(W,Z) \leq 2\lambda_2. \tag{1.3}$$

For large λ Kerstan [14] and respectively Chen [5] improved these bounds to

$$d(W,Z) \,\leq\, \frac{2.1}{\lambda}\,\lambda_2 \quad \text{if} \quad \max_{j\leq n} p_j \leq \frac{1}{4}, \qquad \text{respectively} \quad d(W,Z) \,\leq\, \frac{10}{\lambda}\,\lambda_2.$$

See also [12], [26], [24], [20], [21], [3] and the references therein. A certain refinement of the lower bound in (1.2) was obtained in Sason [22].

While (1.2) provides a sharp estimate for the total variation distance, one may wonder whether or not similar approximation bounds still hold for the stronger informational distances. As a first interesting example, one may consider the relative entropy

$$D(W||Z) = \sum_{k=0}^{\infty} w_k \log\frac{w_k}{v_k},$$

often called the Kullback-Leibler distance, or an informational divergence of P_W from P_λ. It dominates the total variation distance in view of the Pinsker inequality

$$D(W||Z) \,\geq\, \frac{1}{2}\,d(W,Z)^2.$$

In this context, a number of lower and upper bounds on the relative entropy were studied by Harremoës [7], [8], and Harremoës and Ruzankin [10]. In particular, in the i.i.d. case $p_j = p$, it was shown in [10] that

$$\begin{aligned} \frac{-\log(1-p)-p}{2} - \frac{14p^2}{n\,(1-p)^3} \;&\leq\; D(W||Z) \\ &\leq\; \frac{-\log(1-p)-p}{2} - \frac{(1+p)\,p^2}{4n\,(1-p)^3}. \end{aligned}$$

If $p = \lambda/n$ with a fixed (or just bounded) value of λ, these estimates provide the rate of Poisson approximation

$$D(W||Z) \,=\, \frac{\lambda^2}{4n^2} + O(1/n^3) \quad \text{as} \quad n\to\infty. \tag{1.4}$$

The general non-i.i.d. scenario (with not necessarily equal probabilities p_j) has been partially studied as well. A simple upper estimate $D(W||Z) \leq \lambda_2$, analogous to the Le Cam bound (1.3), may be found in [7], cf. also Johnson [13]. It is however not so sharp as (1.4). A much tighter upper bound

$$D(W||Z) \ \le \ \frac{1}{\lambda} \sum_{j=1}^{n} \frac{p_j^3}{1 - p_j} \tag{1.5}$$

was later derived by Kontoyiannis, Harremoës and Johnson [15]. If all $p_j = \lambda/n$ with $\lambda \le n/2$, it yields $D(W||Z) \le 2\lambda^2/n^2$ reflecting a correct decay with respect to n up to a constant, according to (1.4). Nevertheless, in the general case, Pinsker's inequality and the bounds (1.2) and (1.3) suggest that a further sharpening such as

$$D(W||Z) \le A_\lambda \lambda_2^2 \tag{1.6}$$

might be possible by involving λ_2 rather than the functional $\lambda_3 = p_1^3 + \cdots + p_n^3$. To compare the two quantities, note that, by Cauchy's inequality, $\lambda_2^2 \le \lambda\lambda_3$. Hence, the inequality (1.6) would be sharper compared to (1.5), modulo a λ-dependent factor. An upper bound such as (1.6) may also be inspired by the lower bound

$$D(W||Z) \ge \frac{1}{4} \Big(\frac{\lambda_2}{\lambda}\Big)^2 \tag{1.7}$$

recently derived by Harremoës, Johnson and Kontoyiannis [9]. It is consistent with (1.4) and also shows that the constant 1/4 is best possible.

As it turns out, (1.6) does hold in the so-called non-degenerate situation, and in essence, the inequality (1.7) may be reversed. Moreover, one can further sharpen (1.6) by replacing the relative entropy with the Pearson χ^2-distance, as well as with other Rényi/Tsallis distances. To avoid technical complications, let us restrict ourselves to the χ^2-divergence which is given by

$$\chi^2(W,Z) = \sum_{k=0}^{\infty} \frac{(w_k - v_k)^2}{v_k}.$$

It is a divergence type quantity which dominates the relative entropy via the inequality

$$\chi^2(W,Z) \ \ge \ D(W||Z). \tag{1.8}$$

For a general theory of informational distances, we refer interested readers to the recent review by van Erven and Harremoës [6]; additional material may be found in the books [17], [18], [25], [13].

To formulate the main result of this paper in a compact form, let us use the notation $Q_1 \sim Q_2$, whenever two positive quantities are related by $c_1 Q_1 \le Q_2 \le c_2 Q_1$ with some absolute constants $c_j > 0$. Introduce the quantity

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

Theorem 1.2. *We have*

$$D(W||Z) \sim \Big(\frac{\lambda_2}{\lambda}\Big)^2 (1 + \log F), \qquad \chi^2(W,Z) \sim \Big(\frac{\lambda_2}{\lambda}\Big)^2 \sqrt{F}.$$

Clearly, $F \geq 1$, and if $\lambda \leq 1$, then $F = 1$. If λ is not large (say $\lambda \leq 10$, which is typical for applications), or if λ_2/λ is bounded away from 1 (for instance, when $\max_j p_j \leq 1/2$), the quantity F is bounded, and both equivalences are simplified to

$$D(W||Z) \,\sim\, \chi^2(W,Z) \,\sim\, \Big(\frac{\lambda_2}{\lambda}\Big)^2. \tag{1.9}$$

Hence in the above regime, (1.6) holds with a factor $A_\lambda \sim 1/\lambda^2$, which tends to infinity as λ is approaching zero, in contrast with the lower estimate in (1.2).

If the above assumptions on λ and λ_2 are violated (which we call the "degenerate case"), both distances are bounded away from zero and can be large, since then

$$D(W||Z) \,\sim\, \log\frac{\lambda}{\max\{1,\lambda-\lambda_2\}}, \qquad \chi^2(W,Z) \,\sim\, \Big(\frac{\lambda}{\max\{1,\lambda-\lambda_2\}}\Big)^{1/2}.$$

For example, in the case where all $p_j = 1$, we have $\lambda_2 = \lambda = n$. Here $\mathbb{P}\{W = n\} = 1$, hence as $n \to \infty$

$$D(W||Z) = \log\frac{1}{\mathbb{P}\{Z = n\}} = \log\Big(\frac{n!}{n^n}\,e^n\Big) \sim \log n,$$

$$\chi^2(W,Z) = \frac{1}{\mathbb{P}\{Z = n\}} - 1 = \frac{n!}{n^n}\,e^n - 1 \sim \sqrt{2\pi n}.$$

These examples show that the lower bound (1.7) may not be reversed in general.

For the study of the asymptotic behavior of D and χ^2 in terms of λ and λ_2, we derive new bounds for the difference between densities of W and Z, that is, for

$$\Delta_k = w_k - v_k = \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}.$$

To this aim, one has to consider different zones of λ's, distinguishing between "small" and "large" values. The case $\lambda \leq \frac{1}{2}$ can be handled directly leading to the non-uniform density bound

$$|\Delta_k| \,\leq\, 2\,\lambda_2\,\mathbb{P}\{k-2 \leq Z \leq k\}.$$

It easily yields sharp upper bounds for all above distances as in Theorems 1.1-1.2 in the case of small λ, at least up to numerical factors. To treat larger values of λ, a more sophisticated analysis in the complex plane is involved – using the closeness of the generating functions associated with the sequences w_k and v_k. In particular, the following statement may be of independent interest.

Theorem 1.3. *For all integer $k \geq 0$, we have*

$$|\Delta_k| \,\leq\, 3\lambda_2\,e^{-\lambda}. \tag{1.10}$$

Moreover, putting $\rho = (\lambda - \lambda_2)\,\min\{\frac{k}{\lambda}, \frac{\lambda}{k}\}$, $k = 1, 2, \dots,$ we have

$$\begin{aligned} |\Delta_k| \;\;&\leq\;\; 7\sqrt{k}\,\Big(\frac{k-\lambda}{\lambda}\Big)^2\,\lambda_2\min\big\{1,\rho^{-1/2}\big\}\;\mathbb{P}\{Z = k\} \\ &\quad + 21\,k\sqrt{k}\,\frac{\lambda_2}{\lambda}\,\min\big\{1,\rho^{-3/2}\big\}\;\mathbb{P}\{Z = k\}. \end{aligned} \tag{1.11}$$

Let us clarify the meaning of the last bound, assuming that $\lambda_2 \leq \kappa\lambda$ with some constant $\kappa \in (0,1)$. If $k \leq 2\lambda$ and $\lambda \geq \lambda_0 > 0$, then with some $c = c_{\kappa,\lambda_0} > 0$, it gives

$$|\Delta_k| \,\leq\, c\left(\frac{(k-\lambda)^2}{\lambda} + 1\right)\frac{\lambda_2}{\lambda}\,\mathbb{P}\{Z = k\},$$

while for $k \geq \lambda \geq \lambda_0$, we also have

$$|\Delta_k| \,\leq\, c\left(\frac{k}{\lambda}\right)^3 \lambda_2\,\mathbb{P}\{Z = k\}.$$

Since $|k-\lambda|$ is of order at most $\sqrt{\lambda}$ on a sufficiently large part of $\mathbb{Z}$ measured by P_λ, these non-uniform bounds explain the possibility of upper bounds in Theorem 1.2.

Let us finally mention one application of Theorem 1.2 to the problem of estimation of the difference of entropies

$$H(W||Z) = H(Z) - H(W), \tag{1.12}$$

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 remarkable property that $H(W||Z)$ is positive represents a consequence of the assertion, recently proved by Hillion and Johnson [11], that $H(p) \equiv H(W) = H(P_W)$ is a concave function of the vector $p = (p_1,\dots,p_n)$. Indeed, since also $H(p)$ is invariant under permutations of the coordinates p_j, this entropy attains its maximum on the simplex

$$\{p \in \mathbb{R}^n : p_j \geq 0,\ p_1 + \dots + p_n = \lambda\}$$

at the point where all the coordinates coincide, that is, for $p_j = \lambda/n$. But in that case, P_W represents the binomial law $B(n,\lambda/n)$ whose entropy is dominated by $H(Z)$, as was earlier shown by Harremoës [7].

Thus, the difference of entropies in this particular discrete model may be viewed as kind of informational distance. Sason proposed to bound $H(W||Z)$ for equal p_j's by means of the so-called maximal coupling, cf. [23]. 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 leads to the following estimate.

Corollary 1.4. *We have*

$$H(W||Z) \,\leq\, C_\lambda \frac{\lambda_2}{\lambda}, \tag{1.13}$$

where C_λ depends only on λ. If $\lambda_2 \leq \frac{1}{2}\,\lambda$, one may take $C_\lambda = C\log(2+\lambda)$ with an absolute constant C.

The paper is organized as follows. First we describe several general bounds involving the relative entropy and the Pearson distance, together with upper bounds on the probability function of the Poisson law (Section 2). In Sections 3, we consider the deviations Δ_k and prove Theorem 1.2 in case $\lambda \leq 1/2$. Sections 4-5 are devoted to non-uniform bounds and the proof of Theorem 1.3, which is used to complete the proof of Theorem 1.2 for large λ in the non-degenerate case. Uniform bounds for large λ are discussed in Section 7. There we shall demonstrate that in a typical situation, namely when the ratio λ_2/λ is small, the Poisson approximation considerably improves the rate of normal approximation described by the Berry-Esseen bound in the central limit theorem. The remaining part of the paper is devoted to the proof of Theorem 1.2 in the degenerate case (Section 8-11) and of Corollary 1.4 (Section 12). Thus, the paper is structured as follows.

Contents

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 $w_k = \mathbb{P}\{W = \omega_k\}$ and $v_k = \mathbb{P}\{Z = \omega_k\}$, 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 2.1. *We have*

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

Moreover,

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

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

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

Here

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

thus proving the first assertion. Similarly, we have a second identity

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

Adding the two identities, we get

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

which is the desired inequality (2.2). □

Proposition 2.2. *Let W_1 and W_2 be independent, non-negative, integer-valued random variables with finite means, and let Z_1 and Z_2 be independent Poisson random variables with $\mathbb{E}Z_1 = \mathbb{E}W_1$ and $\mathbb{E}Z_2 = \mathbb{E}W_2$. Then*

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

In addition,

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

For the proof, we refer to Johnson [13], pp. 133–134. Let us only mention that (2.5) is obtained in [13] in the more general form

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

with arbitrary $\alpha \geq 1$, 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 the relation (2.3).

When bounding the Poisson probabilities

$$v_k = f(k) = \mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad k = 0, 1, \dots,$$

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

$$\sqrt{2\pi}\, k^{k+\frac{1}{2}}\, e^{-k} \leq k! \leq \ e\, k^{k+\frac{1}{2}}\, e^{-k} \qquad (k \geq 1). \tag{2.5}$$

In particular, it implies the following Gaussian type estimates.

Lemma 2.3. *For all $k \geq 1$,*

$$f(k) \,\leq\, \frac{1}{\sqrt{2\pi k}}. \tag{2.6}$$

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

$$\frac{1}{e\sqrt{k}}\, e^{-\frac{(k-\lambda)^2}{\lambda}} \,\leq\, f(k) \,\leq\, \frac{1}{\sqrt{2\pi k}}\, e^{-\frac{(k-\lambda)^2}{3\lambda}}. \tag{2.7}$$

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

$$f(k) \,\geq\, \frac{1}{e\sqrt{k}}\, e^{-\frac{(k-\lambda)^2}{2\lambda}}. \tag{2.8}$$

Proof. Applying the lower estimate in (2.5), we get

$$\begin{aligned} f(k) \ &\leq \ \frac{1}{\sqrt{2\pi k}}\, e^{k-\lambda} \left(\frac{\lambda}{k}\right)^k \qquad\qquad (2.9)\\ &= \ \frac{1}{\sqrt{2\pi k}}\, e^{k-\lambda}\ e^{-k\log(1+\frac{k-\lambda}{\lambda})} \ = \ \frac{1}{\sqrt{2\pi k}}\, e^{\lambda h(\theta)}, \qquad \theta = \frac{k-\lambda}{\lambda}, \end{aligned}$$

where

$$h(\theta) = \theta - (1+\theta)\log(1+\theta).$$

This function is concave in $\theta \geq -1$, with $h(0) = h'(0) = 0$. Hence, $h(\theta) \leq 0$ for all θ, thus proving the first assertion (2.6).

Assuming that $1 \leq k \leq 2\lambda$ (with $\lambda \geq \frac{1}{2}$), we necessarily have $|\theta| \leq 1$. In this interval, consider the function $T_c(\theta) = h(\theta) + c\theta^2$ with parameter $c > 0$. The second derivative

$$T_c''(\theta) = -\frac{1}{1+\theta} + 2c \qquad (-1 < \theta \leq 1)$$

may change the sign at most at one point, say θ_0, while $T_c''(-1) = -\infty$. Since $T_c(0) = T_c'(0) = 0$, this means that either T_c is concave on $[-1,1]$ and therefore non-positive, or it is concave on $[-1,\theta_0]$ and convex on $[\theta_0,1]$. In the second case, $T_c(\theta) \leq 0$ for all $\theta \in [-1,1]$, if and only if this inequality is fulfilled at $\theta = 1$. But $T_c(1) = 1 - 2\log 2 + c$, so the optimal value is $c = 2\log 2 - 1 = 0.387... > 1/3$. Hence, $h(\theta) \leq -\frac{1}{3}\,\theta^2$, and we arrive at the upper bound in (2.7).

Similarly, applying the upper estimate in (2.5), we get

$$f(k) \,\geq\, \frac{1}{e\sqrt{k}}\, e^{k-\lambda} \left(\frac{\lambda}{k}\right)^k \,=\, \frac{1}{e\sqrt{k}}\, e^{\lambda h(\theta)}, \qquad \theta = \frac{k-\lambda}{\lambda}.$$

Choosing $c = 1$, consider the function $T(\theta) = h(\theta) + \theta^2$ in the interval $|\theta| \leq 1$. Since $T''(-\frac{1}{2}) = 0$, it is concave on $[-1,-\frac{1}{2}]$ and is convex on $[-\frac{1}{2},1]$. Since $T(0) = T'(0) = 0$ and $T(-1) = 0$, this means that $\theta = 0$ is the point of local and thus global minimum of T. Therefore, $T(\theta) \geq 0$, that is, $h(\theta) \geq -\theta^2$ for all $\theta \in [-1,1]$.

Finally, to get the refinement in the region $k \geq \lambda$, consider the function $T(\theta) = h(\theta) + \frac{1}{2}\,\theta^2$ for $\theta \geq 0$. Since $T(0) = 0$ and $T'(\theta) = \theta - \log(1+\theta) \geq 0$, this function is increasing. Therefore, $T(\theta) \geq 0$, that is, $h(\theta) \geq -\frac{1}{2}\,\theta^2$ for all $\theta \geq 0$.

$\square$

3. Elementary Upper Bounds

We keep the same notations as before; in particular,

$$\mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad k = 0, 1, \dots,$$

while

$$\mathbb{P}\{W = k\} \ = \ \sum p_1^{\varepsilon_1}\, (1 - p_1)^{1-\varepsilon_1} \dots p_n^{\varepsilon_n}\, (1 - p_n)^{1-\varepsilon_n}$$

with summation over all 0-1 sequences $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ such that $\varepsilon_1 + \dots + \varepsilon_n = k$. Clearly, $\mathbb{P}\{W = k\} = 0$ for $k > n$. To eliminate this condition, one may always assume that n is arbitrary, by extending the sequence $(X_1, \dots, X_n)$ to $(X_1, \dots, X_k)$ in case $n < k$ with $p_{n+1} = \dots = p_k = 0$. Then the value W does not change.

First, let us consider the values $k = 0$ and $k = 1$.

Lemma 3.1. *If* $\max_j p_j \le \frac{1}{2}$, *then*

$$\begin{aligned} 0 \ &\le \ \mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} \ \le \ 0.8\, \lambda_2\, e^{-\lambda}, \\ 0 \ &\le \ \mathbb{P}\{W = 1\} - \mathbb{P}\{Z = 1\} \ \le \ 2\lambda_2\, e^{-\lambda}. \end{aligned}$$

Proof. Expanding the function $p \to -\log(1-p)$ near zero according to the Taylor formula as in the previous section, write

$$\mathbb{P}\{W = 0\} = \prod_{j=1}^n (1 - p_j) = e^{-\lambda - S}, \qquad S = \sum_{s=2}^\infty \frac{1}{s}\, \lambda_s. \tag{3.1}$$

Using $\lambda_s \le (\max_j p_j)^{s-2}\, \lambda_2 \le 2^{-(s-2)}\, \lambda_2$ for $s \ge 2$, we have

$$S \ \le \ \lambda_2 \sum_{s=2}^\infty \frac{2^{-(s-2)}}{s} \ = \ (4 \log 2 - 2)\, \lambda_2 \ \le \ 0.8\, \lambda_2. \tag{3.2}$$

Hence

$$\mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} = e^{-\lambda}\, (1 - e^{-S}) \le e^{-\lambda}\, S,$$

proving the first inequality.

Next, using the simple representation $\frac{p_j}{1-p_j} = p_j + 2\theta_j p_j^2$ with $0 \le \theta_j \le 1$, we have

$$\begin{aligned} \mathbb{P}\{W = 1\} \ &= \ \prod_{j=1}^n (1 - p_j)\ \sum_{j=1}^n \frac{p_j}{1 - p_j} \\ &\le \ e^{-\lambda - S}\, (\lambda + 2\lambda_2) \ \le \ e^{-\lambda}\, (\lambda + 2\lambda_2) \ = \ \mathbb{P}\{Z = 1\} + 2\lambda_2\, e^{-\lambda}, \end{aligned}$$

which yields the second inequality. $\square$

Note that the condition of Lemma 3.1 is fulfilled automatically, if $\lambda \le 1/2$. In that case, the upper bounds of the lemma may easily be reversed up to numerical factors, for example, in the form

$$\begin{aligned} \mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} \ &\ge \ 0.47\, \lambda_2\, e^{-\lambda}, \\ \mathbb{P}\{W = 1\} - \mathbb{P}\{Z = 1\} \ &\ge \ 0.42\, \lambda_2\, e^{-\lambda}. \end{aligned}$$

Moreover, if $\lambda \le 1/8$, then also

$$\mathbb{P}\{Z=2\} - \mathbb{P}\{W=2\} \ \ge \ \frac{17}{49}\,\lambda_2\,e^{-\lambda}.$$

Here, the value $k=2$ turns out to be most essential for obtaining lower bounds, since it immediately yields $d(W,Z) \ge c\lambda_2$ and $D(W||Z) \ge c\,(\frac{\lambda_2}{\lambda})^2$ with some absolute constant $c>0$.

Returning to upper bounds, in order to involve the values $k \ge 2$, we will need the following:

Lemma 3.2. *If* $\max_j p_j \le 1/2$, *then for any* $k \ge 2$,

$$\big|\mathbb{P}\{W=k\} - \mathbb{P}\{Z=k\}\big| \ \le \ \lambda_2\,\Big(\frac{\lambda^k}{k!} + \frac{e^\lambda - 1}{\lambda}\,\frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda^{k-2}}{(k-2)!}\Big)\,e^{-\lambda}.$$

Proof. Representing the Poisson random variable $Z \sim P_\lambda$ as $Z = Z_1 + \dots + Z_n$ with independent summands $Z_j \sim P_{p_j}$, we have that, for any $k = 0,1,\dots$,

$$\mathbb{P}\{Z=k\} = e^{-\lambda} \sum_{\varepsilon_1+\dots+\varepsilon_n=k} \frac{p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}}{\varepsilon_1!\dots\varepsilon_n!},$$

where the summation is running over all integers $\varepsilon_j \ge 0$ such that $\varepsilon_1 + \dots + \varepsilon_n = k$. Hence, with this assumption, we may start with the formula

$$\mathbb{P}\{Z=k\} - \mathbb{P}\{W=k\} \ = \ e^{-\lambda} \sum_{\varepsilon_1+\dots+\varepsilon_n=k} \frac{1}{\varepsilon_1!\dots\varepsilon_n!}\,U_\varepsilon \quad - \sum_{\varepsilon_1+\dots+\varepsilon_n=k,\,\varepsilon_j\le 1} U_\varepsilon V_\varepsilon,$$

where

$$U_\varepsilon = p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}, \qquad V_\varepsilon = (1-p_1)^{1-\varepsilon_1}\dots(1-p_n)^{1-\varepsilon_n}.$$

For a 0-1 sequence $\varepsilon = (\varepsilon_1,\dots,\varepsilon_n)$ put

$$L_\varepsilon = \varepsilon_1 p_1 + \dots + \varepsilon_n p_n$$

By the Taylor formula once more,

$$V_\varepsilon^{-1} \ = \ e^{S_\varepsilon}, \qquad S_\varepsilon = \sum_{s=1}^\infty \ \frac{1}{s}\ \sum_{j=1}^n (1-\varepsilon_j)\,p_j^s.$$

Similarly to (3.1)-(3.2), we have

$$S_\varepsilon \ = \ \lambda - L_\varepsilon + \sum_{s=2}^\infty \ \frac{1}{s}\ \sum_{j=1}^n (1-\varepsilon_j)\,p_j^s \ = \ \lambda - L_\varepsilon + \theta\lambda_2, \qquad 0 \le \theta \le 1.$$

Therefore,

$$e^\lambda\,V_\varepsilon = e^{L_\varepsilon - \theta\lambda_2} \ge 1 + (L_\varepsilon - \theta\lambda_2) \ge 1 + L_\varepsilon - \lambda_2.$$

Moreover, since $L_\varepsilon \le \lambda$, we have $\frac{e^{L_\varepsilon}-1}{L_\varepsilon} \le \frac{e^\lambda-1}{\lambda} \equiv c_\lambda$, which in turn implies $e^\lambda\,V_\varepsilon \le e^{L_\varepsilon} \le 1 + c_\lambda\,L_\varepsilon$. The two bounds give $L_\varepsilon - \lambda_2 \le e^\lambda\,V_\varepsilon - 1 \le c_\lambda\,L_\varepsilon$, so that

$$\big|U_\varepsilon - e^\lambda\,U_\varepsilon V_\varepsilon\big| \ \le \ \lambda_2\,U_\varepsilon + c_\lambda\,U_\varepsilon L_\varepsilon.$$

Next, applying the multinomial formula, we have

$$\sum_{\varepsilon_1+\dots+\varepsilon_n=k,\ \varepsilon_j\le 1} U_\varepsilon \ \le \ \sum_{\varepsilon_1+\dots+\varepsilon_n=k} \frac{p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}}{\varepsilon_1!\dots\varepsilon_n!} \ = \ \frac{\lambda^k}{k!}$$

and

$$\begin{aligned}
\sum_{\varepsilon_1+\dots+\varepsilon_n=k,\ \varepsilon_j\le 1} U_\varepsilon\, L_\varepsilon \ &= \ \sum_{i=1}^n \ \sum_{\varepsilon_1+\dots+\varepsilon_n=k,\ \varepsilon_j\le 1} \varepsilon_i\ p_1^{\varepsilon_1}\dots p_{i-1}^{\varepsilon_{i-1}}\, p_i^{\varepsilon_i+1}\, p_{i+1}^{\varepsilon_{i+1}}\dots p_n^{\varepsilon_n} \\
&= \ \sum_{i=1}^n p_i^2 \sum_{\varepsilon_1+\dots+\varepsilon_n=k,\ \varepsilon_i=1,\ \varepsilon_j\le 1} p_1^{\varepsilon_1}\dots p_{i-1}^{\varepsilon_{i-1}}\, p_{i+1}^{\varepsilon_{i+1}}\dots p_n^{\varepsilon_n} \\
&\le \ \sum_{i=1}^n p_i^2\, \frac{1}{(k-1)!}\,(\lambda-p_i)^{k-1} \ \le \ \lambda_2\, \frac{\lambda^{k-1}}{(k-1)!}.
\end{aligned}$$

Thus,

$$\sum_{\varepsilon_1+\dots+\varepsilon_n=k,\ \varepsilon_j\le 1} |U_\varepsilon - e^\lambda\, U_\varepsilon V_\varepsilon| \ \le \ \lambda_2\left(\frac{\lambda^k}{k!} + c_\lambda\, \frac{\lambda^{k-1}}{(k-1)!}\right).$$

For the remaining terms participating in $\mathbb{P}(Z=k)$, we have

$$\begin{aligned}
\sum_{\varepsilon_1+\dots+\varepsilon_n=k,\,\varepsilon_n\ge 2} \frac{p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}}{\varepsilon_1!\dots\varepsilon_n!} \ &= \ \sum_{m=2}^k \frac{p_n^m}{m!} \sum_{\varepsilon_1+\dots+\varepsilon_{n-1}=k-m} \frac{p_1^{\varepsilon_1}\dots p_{n-1}^{\varepsilon_{n-1}}}{\varepsilon_1!\dots\varepsilon_{n-1}!} \\
&\le \ \sum_{m=2}^k \frac{p_n^m}{m!}\ \frac{(\lambda-p_n)^{k-m}}{(k-m)!} \\
&\le \ p_n^2 \sum_{m=2}^k \frac{p_n^{m-2}}{(m-2)!}\ \frac{(\lambda-p_n)^{k-m}}{(k-m)!} \ = \ p_n^2\, \frac{\lambda^{k-2}}{(k-2)!},
\end{aligned}$$

and similarly, for any $i=1,\dots,n$,

$$\sum_{\varepsilon_1+\dots+\varepsilon_n=k,\,\varepsilon_i\ge 2} \frac{p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}}{\varepsilon_1!\dots\varepsilon_n!} \ \le \ p_i^2\, \frac{\lambda^{k-2}}{(k-2)!}.$$

Hence, summing over $i\le n$, we then get

$$\sum_{\varepsilon_1+\dots+\varepsilon_n=k,\,\max\varepsilon_j\ge 2} \frac{p_1^{\varepsilon_1}\dots p_n^{\varepsilon_n}}{\varepsilon_1!\dots\varepsilon_n!} \ \le \ \lambda_2\, \frac{\lambda^{k-2}}{(k-2)!}.$$

□

The obtained estimates are sufficient to establish Theorem 1.2 in the non-degenerate case, where λ is not large. To compare the lower and upper bounds, we recall the lower bound (1.7) of Harremoës, Johnson and Kontoyiannis [9].

Proposition 3.3. *If* $\max_j p_j \le \frac{1}{2}$, *then*

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

where C_λ depends on $\lambda \geq 0$ as an increasing continuous function with $C_0 = 2$. In particular, if $\lambda \leq 1/2$, then

$$\chi^2(W,Z) \ \leq \ 16 \Big(\frac{\lambda_2}{\lambda}\Big)^2.$$

Proof. Applying Lemmas 3.1-3.2, we get

$$\lambda_2^{-2}\, e^{\lambda}\, \chi^2(W,Z) \ \leq \ 0.64 + \frac{4}{\lambda} + \sum_{k=2}^\infty \frac{k!}{\lambda^k} \Big(\frac{\lambda^k}{k!} + c_\lambda\, \frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda^{k-2}}{(k-2)!}\Big)^2,$$

where $c_\lambda = \frac{e^\lambda - 1}{\lambda}$. Opening the brackets, the above sum is equal to

$$\sum_{k=2}^\infty \frac{k!}{\lambda^k} \Big(\frac{\lambda^{2k}}{k!^{\,2}} + \frac{2c_\lambda\, \lambda^{2k-1}}{k!\,(k-1)!} + \frac{c_\lambda^2\, \lambda^{2k-2}}{(k-1)!^{\,2}} + \frac{2\lambda^{2k-2}}{k!\,(k-2)!} + \frac{2c_\lambda\, \lambda^{2k-3}}{(k-1)!\,(k-2)!} + \frac{\lambda^{2k-4}}{(k-2)!^{\,2}}\Big)$$

$$\begin{aligned} = \ & \sum_{k=2}^\infty \frac{\lambda^k}{k!} + 2c_\lambda \sum_{k=2}^\infty \frac{\lambda^{k-1}}{(k-1)!} + c_\lambda^2 \sum_{k=2}^\infty k\, \frac{\lambda^{k-2}}{(k-1)!} + 2 \sum_{k=2}^\infty \frac{\lambda^{k-2}}{(k-2)!} \\ & + 2c_\lambda \sum_{k=2}^\infty k\, \frac{\lambda^{k-3}}{(k-2)!} + \sum_{k=2}^\infty k(k-1)\, \frac{\lambda^{k-4}}{(k-2)!}, \end{aligned}$$

which is the same as

$$\begin{aligned} & 3e^\lambda - 1 - \lambda + 2c_\lambda\, (e^\lambda - 1) + c_\lambda^2 \sum_{k=1}^\infty (k+1)\, \frac{\lambda^{k-1}}{k!} \\ & \qquad + 2c_\lambda \sum_{k=0}^\infty (k+2)\, \frac{\lambda^{k-1}}{k!} + \sum_{k=0}^\infty (k+1)(k+2)\, \frac{\lambda^{k-2}}{k!} \\ & = \ 3e^\lambda - 1 - \lambda + 2c_\lambda\, (e^\lambda - 1) + 2c_\lambda e^\lambda\, \frac{2+\lambda}{\lambda} + \frac{2 + 4\lambda + \lambda^2}{\lambda^2}\, e^\lambda. \end{aligned}$$

Multiplying by λ^2, this gives the desired inequality

$$\lambda^2 \lambda_2^{-2}\, \chi^2(W,Z) \ \leq \ C_\lambda \ = \ (0.64\, \lambda^2 + 4\lambda) + B_\lambda$$

with

$$\begin{aligned} B_\lambda \ &= \ \lambda^2\, (3e^\lambda - 1 - \lambda) + 2\lambda\, (e^\lambda - 1)^2 + 2\, (2+\lambda)\, e^\lambda\, (e^\lambda - 1) + (2 + 4\lambda + \lambda^2)\, e^\lambda \\ &= \ \lambda\, (2 - \lambda - \lambda^2) - 2\, (1 + \lambda - 2\lambda^2)\, e^\lambda + 4\, (1+\lambda)\, e^{2\lambda}. \end{aligned}$$

It is easy to check that $\frac{d}{d\lambda}\, B_\lambda > 0$, so that this function is increasing in λ. In addition, $C_0 = B_0 = 2$ and $C_{1/2} = 2.16 + \frac{5}{8} - 2\sqrt{e} + 6\, e < 16$. $\square$

4. **Generating functions**

The probability function $f(k) = \mathbb{P}\{Z = k\}$ of the Poisson random variable $Z \sim P_\lambda$ satisfies the equation $\lambda f(k-1) = k f(k)$ in integers $k \geq 1$, which immediately implies

$$\lambda\, \mathbb{E}\, h(Z+1) = \mathbb{E}\, Z h(Z)$$

for any function h on $\mathbb{Z}$ (as long as the expectations exist). This identity was emphasized by Chen [5] who proposed to consider an approximate equality

$$\lambda\,\mathbb{E}\,h(X+1) \sim \mathbb{E}\,Xh(X)$$

as a characterization of a random variable X being almost Poisson with parameter λ. This idea was inspired by a similar approach by Stein to the problems for normal approximation on the basis of an approximate equality $\mathbb{E}\,h'(X) \sim \mathbb{E}\,Xh(X)$.

Another natural approach to the Poisson approximation is based on the comparison of characteristic functions. Since the random variables W and Z take non-negative integer values, one may equivalently consider the associated generating functions.

The generating function for the Poisson law P_λ with parameter $\lambda > 0$ is given by

$$\varphi(w) = \mathbb{E}\,w^Z = \sum_{k=0}^{\infty} \mathbb{P}\{Z=k\}\,w^k = e^{\lambda(w-1)} = \prod_{j=1}^{n} e^{p_j(w-1)}, \tag{4.1}$$

which is an entire function of the complex variable w. Correspondingly, the generating function for the distribution of the random variable $W = X_1 + \dots + X_n$ in (1.1) is

$$g(w) = \mathbb{E}\,w^W = \sum_{k=0}^{\infty} \mathbb{P}\{W=k\}\,w^k = \prod_{j=1}^{n} (q_j + p_j w), \tag{4.2}$$

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

$$\mathbb{P}\{W=k\} - \mathbb{P}\{Z=k\} = \int_{|w|=r} w^{-k}\,(g(w) - \varphi(w))\,d\mu_r(w), \tag{4.3}$$

where μ_r is the uniform probability measure on the circle $|w| = r$ of an arbitrary radius $r > 0$.

Note that for $w = e^{it}$ with real t, the generating functions φ and g become the characteristic functions of Z and W, respectively. Hence, closeness of the distributions of these random variables may be studied as a problem of the closeness of the generating functions on the unit circle.

Let us now describe first steps based on the application of the formula (4.3). Given complex numbers a_j, b_j $(1 \le j \le n)$, we have an identity

$$a_1 \dots a_n - b_1 \dots b_n = \sum_{j=1}^{n} (a_j - b_j) \prod_{l<j} b_l \prod_{l>j} a_l \tag{4.4}$$

with the convention that $\prod_{l<j} b_l = 1$ for $j = 1$ and $\prod_{l>j} a_l = 1$ for $j = n$. It implies

$$\Big|\prod_{j=1}^{n} a_j - \prod_{b=1}^{n} b_j\Big| \,\le\, \sum_{j=1}^{n} |a_j - b_j| \prod_{l<j} |b_l| \prod_{l>j} |a_l|.$$

According to the product representations (4.1)-(4.2) to be used in (4.3), one should choose here $a_j = q_j + p_j w$ and $b_j = e^{p_j(w-1)}$ with $|w| = r$. Then

$$|a_j| \,\le\, q_j + p_j r \,\le\, e^{p_j(r-1)}, \qquad |b_j| = e^{p_j(\mathrm{Re}\,w - 1)} \le e^{p_j(r-1)}. \tag{4.5}$$

Therefore

$$\begin{aligned} |g(w) - \varphi(w)| &\leq \sum_{j=1}^n |a_j - b_j| \prod_{l \neq j} e^{p_l(r-1)} \\ &= e^{\lambda(r-1)} \sum_{j=1}^n |a_j - b_j|\, e^{-p_j(r-1)}. \end{aligned} \tag{4.6}$$

To estimate the terms in this sum, consider the function

$$\xi(u) = 1 + u - e^u = -u^2 \int_0^1 e^{tu}\,(1-t)\,dt, \qquad u \in \mathbb{C}, \tag{4.7}$$

where the Taylor integral formula is applied in the second representation. If $\mathrm{Re}\,u \leq 0$, then $|u^2\,e^{tu}| = |u|^2\,\exp\{t\,\mathrm{Re}\,u\} \leq |u|^2$, so,

$$|\xi(u)| \leq \frac{1}{2}\,|u|^2, \qquad \mathrm{Re}\,u \leq 0. \tag{4.8}$$

In particular, for $u = p_j(w-1)$ with $w = \cos\theta + i\sin\theta$, we have

$$|w-1|^2 = (\cos\theta - 1)^2 + \sin^2\theta = 2(1-\cos\theta),$$

hence $|\xi(u)| \leq p_j^2\,(1-\cos\theta)$, and (4.6) yields

$$|g(w) - \varphi(w)| \,\leq\, \sum_{j=1}^n |\xi(p_j(w-1))| \,\leq\, (1-\cos\theta)\sum_{j=1}^n p_j^2 \,\leq\, (1-\cos\theta)\,\lambda_2.$$

Integrating over the unit circle in (4.3), we then arrive at the uniform bound:

Proposition 4.1. *We have*

$$\sup_{k\geq 0}\, |\,\mathbb{P}\{W=k\} - \mathbb{P}\{Z=k\}| \,\leq\, \lambda_2. \tag{4.9}$$

This is a weakened variant of Le Cam's bound $|\mathbb{P}\{W\in A\} - \mathbb{P}\{Z \in A\}| \,\leq\, \lambda_2$, specialized to the one-point set $A = \{k\}$. In order to get a similar bound with arbitrary sets, or develop applications to stronger distances, we need sharper forms of (4.9), with the right-hand side properly depending on k.

5. Proof of Theorem 1.3

Applying (4.4) with $a_j = q_j + p_j w$ and $b_j = e^{p_j(w-1)}$ in (4.3), one may write this formula as

$$\Delta_k \,\equiv\, \mathbb{P}\{W=k\} - \mathbb{P}\{Z=k\} \,=\, \sum_{j=1}^n T_j(k), \qquad k = 0,1,\dots, \tag{5.1}$$

with

$$T_j(k) \,=\, \int_{|w|=r} w^{-k}\,(a_j - b_j)\,\prod_{l<j} b_l \prod_{l>j} a_l\;d\mu_r(w), \tag{5.2}$$

where the integration is performed over the uniform probability measure μ_r on the circle $|w| = r$. Let us write $w = r(\cos\theta + i\sin\theta)$, $|\theta| < \pi$, and estimate $|T_j(k)|$ by inserting the absolute value sign inside the integral. Then, using (4.5), we get

$$\begin{aligned}
|T_j(k)| &\leq r^{-k} \int_{|w|=r} |a_j - b_j| \prod_{l<j} |e^{p_l\,(w-1)}| \prod_{l>j} |q_l + p_l w|\; d\mu_r(w) \\
&= r^{-k} \int_{|w|=r} |a_j - b_j|\, \exp\Big\{(r\cos\theta - 1) \sum_{l=1}^{j-1} p_l\Big\} \prod_{l=j+1}^{n} |q_l + p_l w|\; d\mu_r(w) \\
&= r^{-k} e^{(r-1)\sum_{l=1}^{j-1} p_l} \int_{|w|=r} |a_j - b_j|\, \exp\Big\{ -2r\sin^2\frac{\theta}{2} \sum_{l=1}^{j-1} p_l\Big\} \prod_{l=j+1}^{n} |q_l + p_l w|\; d\mu_r(w).
\end{aligned}$$

Here, in order to estimate $|a_j - b_j|$, let us return to the function $\xi(u)$ introduced in (4.7), which we need at the values $u_j = p_j(w-1)$ with $|w| = r$.

Case 1: $r \geq 1$. Since $\mathrm{Re}\, u_j \leq p_j(r-1)$, we have, for any $t \in (0,1)$,

$$|u_j^2\, e^{tu_j}| = |u_j|^2\, e^{t\,\mathrm{Re}\, u_j} \leq |u_j|^2\, e^{p_j t(r-1)} \leq |u_j|^2\, e^{p_j(r-1)},$$

so, by (4.7),

$$|a_j - b_j| = |\xi(u_j)| \leq \frac{1}{2}\, p_j^2\, |w-1|^2\; e^{p_j(r-1)}.$$

Case 2: $0 < r < 1$. Then $\mathrm{Re}\, u_j \leq 0$, so, by (4.8),

$$|a_j - b_j| = |\xi(u_j)| \leq \frac{1}{2}\, p_j^2\, |w-1|^2.$$

Since $|w-1|^2 = (r-1)^2 + 4r\,\sin^2(\theta/2)$, we therefore obtain from (5.2) that

$$|T_j(k)| \leq \frac{1}{2}\, p_j^2\, R_j(r)\, r^{-k} \Big((r-1)^2\, I_{j0}(r) + 4r I_{j2}(r)\Big), \tag{5.3}$$

where

$$R_j(r) = \begin{cases} \exp\big\{(r-1)\sum_{l=1}^{j} p_l\big\} \prod_{l=j+1}^{n} (q_l + p_l r) & \text{for } \; r \geq 1, \\ \exp\big\{(r-1)\sum_{l=1}^{j-1} p_l\big\} \prod_{l=j+1}^{n} (q_l + p_l r) & \text{for } \; r < 1, \end{cases}$$

and

$$I_{jm}(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Big|\sin\frac{\theta}{2}\Big|^m \exp\Big\{ -2r\sin^2\frac{\theta}{2} \sum_{l=1}^{j-1} p_l\Big\} \prod_{l=j+1}^{n} \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r}\; d\theta.$$

In order to estimate the last integrals, which we need with $m = 0$ and $m = 2$, let us first note that

$$\begin{aligned}
|q_l + p_l r e^{i\theta}|^2 &= q_l^2 + p_l^2 r^2 + 2p_l q_l\, r\cos\theta \;=\; (q_l + p_l r)^2 - 4q_l p_l\, r\sin^2\frac{\theta}{2} \\
&= (q_l + p_l r)^2 \Big(1 - \frac{4q_l p_l r}{(q_l + p_l r)^2}\, \sin^2\frac{\theta}{2}\Big).
\end{aligned}$$

Hence, using $1 - x \le e^{-x}$ $(x \in \mathbb{R})$, we have

$$\begin{aligned}
\prod_{l=j+1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r} &= \prod_{l=j+1}^n \Big(1 - \frac{4q_l p_l r}{(q_l + p_l r)^2} \sin^2 \frac{\theta}{2}\Big)^{1/2} \\
&\le \exp\Big\{ - 2 \sin^2 \frac{\theta}{2} \sum_{l=j+1}^n \frac{q_l p_l\, r}{(q_l + p_l r)^2}\Big\}, \qquad (5.4)
\end{aligned}$$

so that

$$\begin{aligned}
I_{jm}(r) &\le \frac{1}{2\pi} \int_{-\pi}^{\pi} \Big|\sin \frac{\theta}{2}\Big|^m \exp\Big\{ - 2\gamma_j(r) \sin^2 \frac{\theta}{2}\Big\}\, d\theta \\
&\le \frac{1}{2\pi}\, 2^{-m} \int_{-\pi}^{\pi} |\theta|^m \exp\Big\{ - \frac{2}{\pi^2}\, \gamma_j(r)\, \theta^2\Big\}\, d\theta. \qquad (5.5)
\end{aligned}$$

Here we applied the inequalities $\frac{2}{\pi}\, t \le \sin t \le t$ $(0 \le t \le \frac{\pi}{2})$ and used the notation

$$\gamma_j(r) = r\,\Big(\sum_{l=1}^{j-1} p_l + \sum_{l=j+1}^n \frac{q_l p_l}{(q_l + p_l r)^2}\Big).$$

Thus, we need to bound γ_j from below. If $r \ge 1$, then $q_l + p_l r \le r$, so

$$\sum_{l=j+1}^n \frac{q_l p_l\, r}{(q_l + p_l r)^2} \ge \frac{1}{r} \sum_{l=j+1}^n q_l p_l.$$

This gives

$$\begin{aligned}
\gamma_j(r) &\ge r \sum_{l=1}^{j-1} p_l + \frac{1}{r} \sum_{l=j+1}^n q_l p_l \\
&= r \sum_{l=1}^{j-1} p_l + \frac{1}{r} \sum_{l=1}^n (p_l - p_l^2) - \frac{1}{r} \sum_{l=1}^j (p_l - p_l^2) \\
&= \Big(r - \frac{1}{r}\Big) \sum_{l=1}^{j-1} p_l + \frac{1}{r} \sum_{l=1}^{j-1} p_l^2 + \frac{1}{r} \sum_{l=1}^n (p_l - p_l^2) - \frac{1}{r}\, q_j p_j \ \ge\ \frac{1}{r}\, (\lambda - \lambda_2 - q_j p_j).
\end{aligned}$$

In case $r \le 1$, we use $q_l + p_l r \le 1$, implying that

$$\sum_{l=j+1}^n \frac{q_l p_l}{(q_l + p_l r)^2} \ge \sum_{l=j+1}^n q_l p_l.$$

Therefore in this range we have a similar lower bound, namely

$$\begin{aligned}
\gamma_j(r) &\ge r \sum_{l=1}^{j-1} p_l + r \sum_{l=j+1}^n q_l p_l \\
&= r \sum_{l=1}^{j-1} p_l + r \sum_{l=1}^n (p_l - p_l^2) - r \sum_{l=1}^j (p_l - p_l^2) \\
&= -r p_j + r \sum_{l=1}^j p_l^2 + r \sum_{l=1}^n (p_l - p_l^2) \ \ge\ r\, (\lambda - \lambda_2 - q_j p_j).
\end{aligned}$$

Since $q_j p_j \le \frac{1}{4}$, both lower bounds yield

$$\gamma_j(r) \,\ge\, \psi(r) - \frac{1}{4}, \qquad \psi(r) = \min\{r, 1/r\}\,(\lambda - \lambda_2).$$

As a result, (5.5) is simplified to

$$\begin{aligned} I_{jm}(r) &\le \frac{1}{2\pi}\, 2^{-m}\sqrt{e} \int_{-\pi}^{\pi} |\theta|^m \exp\Big\{-\frac{2}{\pi^2}\,\psi(r)\,\theta^2\Big\}\,d\theta \\ &= \sqrt{e}\,\frac{\pi^m}{4^{m+1}}\,\psi(r)^{-\frac{m+1}{2}} \int_{-2\sqrt{\psi(r)}}^{2\sqrt{\psi(r)}} |x|^m\, e^{-\frac{1}{2}x^2}\,dx. \end{aligned}$$

The last integral may be extended to the whole real line, which makes sense for large values of $\psi(r)$, or one may bound the exponential term in the integrand by 1, which makes sense for small values of $\psi(r)$. These two ways of estimation lead to

$$\begin{aligned} I_{jm}(r) &\le \sqrt{e}\,\frac{\pi^m}{4^{m+1}}\,\psi(r)^{-\frac{m+1}{2}} \min\Big\{\sqrt{2\pi}\,\mathbb{E}\,|\xi|^m, \frac{2^{m+2}}{m+1}\,\psi(r)^{\frac{m+1}{2}}\Big\} \\ &\le \sqrt{e}\,\frac{\pi^m}{4^{m+1}}\,\max\Big\{\sqrt{2\pi}\,\mathbb{E}\,|\xi|^m, \frac{2^{m+2}}{m+1}\Big\}\,\min\Big\{1, \psi(r)^{-\frac{m+1}{2}}\Big\}, \end{aligned}$$

where ξ is a standard normal random variable. In particular, we get the upper bounds

$$I_{j0}(r) \,\le\, \sqrt{e}\,\min\big\{1, \psi(r)^{-1/2}\big\}, \qquad I_{j2}(r) \,\le\, \frac{\sqrt{e}\,\pi^2}{12}\,\min\big\{1, \psi(r)^{-3/2}\big\}.$$

In view of $q_l + p_l r \le e^{(r-1)p_l}$, from the definition of $R_j(r)$ we also have the bound

$$R_j(r) \,\le\, \exp\Big\{(r-1)\sum_{l=1}^n p_l\Big\} \,=\, e^{\lambda(r-1)}$$

in case $r \ge 1$, while for $r \le 1$

$$R_j(r) \,\le\, \exp\Big\{(r-1)\sum_{l\neq j} p_l\Big\} \,=\, e^{\lambda(r-1)}\,e^{-p_j(r-1)} \,\le\, e^{\lambda(r-1)+1}.$$

Applying these bounds in (5.3), we therefore obtain that $|T_j(k)|$ may be bounded from above by

$$\frac{\delta_r}{2}\,p_j^2\,e^{\lambda(r-1)+\frac{1}{2}}\,r^{-k}\Big((r-1)^2\min\big\{1, \psi(r)^{-1/2}\big\} + \frac{\pi^2}{3}\,r\min\big\{1, \psi(r)^{-3/2}\big\}\Big),$$

where $\delta_r = 1$ in case $r \ge 1$ and $\delta_r = e$ for $r < 1$. Summing over $j \le n$ and recalling (5.1), one can estimate $|\Delta_k|$ from above by

$$\lambda_2\,\delta_r\,e^{\lambda(r-1)}\,r^{-k}\,\Big(\frac{\sqrt{e}}{2}\,(r-1)^2\,\min\big\{1, \psi(r)^{-1/2}\big\} + \frac{\sqrt{e}\,\pi^2}{6}\,r\,\min\big\{1, \psi(r)^{-3/2}\big\}\Big). \quad (5.6)$$

Letting $r \to 1$ ($r > 1$), (5.6) leads to

$$|\Delta_k| \le \frac{\sqrt{e}\,\pi^2}{6}\,\lambda_2\,e^{-\lambda} < 3\lambda_2\,e^{-\lambda},$$

which gives the inequality in (1.10). In case $k \ge 1$, one may also use (5.5) with $r = \frac{k}{\lambda}$ and apply $k! \le e\,k^{k+\frac{1}{2}}\,e^{-k}$, cf. (2.5), giving

$$e^{\lambda(r-1)}\,r^{-k} = \Big(\frac{e\lambda}{k}\Big)^k e^{-\lambda} \le e\sqrt{k}\,f(k), \qquad f(k) = \frac{\lambda^k}{k!}\,e^{-\lambda}.$$

To simplify the numerical constants, note that $\frac{1}{2}\, e^{5/2} < 6.1$ and $\frac{1}{6}\, e^{5/2}\, \pi^2 < 20.1$. Recalling that $\psi(r) = \rho$ for $r = k/\lambda$, we finally get the second inequality (1.11),

$$|\Delta_k| \, \leq \, \lambda_2 \sqrt{k}\, f(k) \, \Big(7 \, \Big(\frac{k-\lambda}{\lambda}\Big)^2 \, \min\big\{1, \rho^{-1/2}\big\} + 21\, \frac{k}{\lambda}\, \min\big\{1, \rho^{-3/2}\big\}\Big). \tag{5.7}$$

$\square$

6. Consequences of Theorem 1.3

Under the natural requirement that λ_2 is bounded away from λ, the bound (5.7) on $\Delta_k = \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}$ may be simplified. As before, we use the notations

$$f(k) = \mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!}\, e^{-\lambda}, \qquad \lambda = p_1 + \dots + p_n, \quad \lambda_2 = p_1^2 + \dots + p_n^2.$$

Note that $\lambda_2 \leq \lambda$ and recall that $\rho = (\lambda - \lambda_2)\, \min\{\frac{k}{\lambda}, \frac{\lambda}{k}\}$.

Corollary 6.1. *If $\lambda_2 \leq \kappa\lambda$, $\kappa \in (0,1)$, then for any integer $k \geq 0$,*

$$|\Delta_k| \, \leq \, \frac{7}{(1-\kappa)^{3/2}}\, \Big(\frac{(k-\lambda)^2}{\lambda} + 3\Big)\, \frac{\lambda_2}{\lambda}\, \max\Big\{\Big(\frac{k}{\lambda}\Big)^3, 1\Big\} f(k). \tag{6.1}$$

In particular, if $k \leq 2\lambda$, then

$$|\Delta_k| \, \leq \, \frac{56}{(1-\kappa)^{3/2}}\, \Big(\frac{(k-\lambda)^2}{\lambda} + 3\Big)\, \frac{\lambda_2}{\lambda}\, f(k). \tag{6.2}$$

If $k \geq \lambda \geq 1/2$, we also have

$$|\Delta_k| \, \leq \, \frac{49}{(1-\kappa)^{3/2}}\, \Big(\frac{k}{\lambda}\Big)^3\, \lambda_2 f(k). \tag{6.3}$$

Proof. The assumption $\lambda_2 \leq \kappa\lambda$ ensures that $\rho \geq (1-\kappa)\lambda\, \min\{\frac{k}{\lambda}, \frac{\lambda}{k}\}$.

If $1 \leq k \leq K\lambda$ $(K \geq 1)$, then $\frac{k}{\lambda} \leq K^2 \frac{\lambda}{k}$ and $\rho \geq \frac{1-\kappa}{K^2}\, k$, so that the right-hand side of (5.6) may be bounded from above by

$$\lambda_2 \sqrt{k}\, f(k) \, \Big(7 \, \Big(\frac{k-\lambda}{\lambda}\Big)^2 \, \frac{K}{\sqrt{(1-\kappa)\, k}} + 21\, \frac{k}{\lambda}\, \frac{K^3}{(1-\kappa)^{3/2}\, k^{3/2}}\Big).$$

Choosing $K = \max\{\frac{k}{\lambda}, 1\}$, this expression does not exceed the right-hand side of (6.1). Thus, the inequality (1.11) yields (6.1), which in turn immediately implies (6.2).

In case $k = 0$, we apply the inequality (1.10). Since $\frac{(k-\lambda)^2}{\lambda} + 3 \geq \lambda$ for $k = 0$, the right-hand side of (1.10) is dominated by the right-hand side of (6.1). Thus, we obtain (6.1) without any constraints on k, and (6.2) for all $k \leq 2\lambda$.

In case $k \geq \lambda$, necessarily $\rho \geq (1-\kappa)\, \lambda^2/k$. Hence, the right-hand side of (5.6) may be bounded from above by

$$\lambda_2 \sqrt{k}\, f(k) \, \Big(7 \, \Big(\frac{k-\lambda}{\lambda}\Big)^2 \, \frac{\sqrt{k}}{\lambda\sqrt{1-\kappa}} + 21\, \frac{k}{\lambda} \cdot \frac{k^{3/2}}{\lambda^3\, (1-\kappa)^{3/2}}\Big).$$

Using $(\frac{k-\lambda}{\lambda})^2 \leq \frac{k^2}{\lambda^2}$ to bound the first term in the brackets and $\frac{k}{\lambda} \leq 2k$ to bound the second term (using $\lambda \geq 1/2$), we obtain the bound (6.3). $\square$

We are now prepared to extend Proposition 3.3 to larger values of λ under the additional assumption that λ_2/λ is bounded away from 1. The next assertion, being combined with Proposition 3.3, yields Theorem 1.2 in the non-degenerate case.

Proposition 6.2. *If $\lambda \geq 1/2$ and $\lambda_2 \leq \kappa\lambda$ with $\kappa \in (0,1)$, then*

$$\frac{1}{4}\Big(\frac{\lambda_2}{\lambda}\Big)^2 \, \leq \, D(W||Z) \, \leq \, \chi^2(W,Z) \, \leq \, c_\kappa \Big(\frac{\lambda_2}{\lambda}\Big)^2. \tag{6.4}$$

where one may take $c_\kappa = c\,(1-\kappa)^{-3}$ with some absolute constant, e.g. $c = 7 \cdot 10^6$.

Proof. The left lower bound in (6.4) is added according to (1.7) (using the Pinsker inequality, it also follows with some constant from Barbour-Hall's lower bound in Theorem 1.1). Hence, one may only focus on the right upper bound in (6.4). Write

$$\chi^2(W,Z) \, = \, \sum_{k=0}^\infty \frac{\Delta_k^2}{f(k)} \, = \, S_1 + S_2 \, = \, \bigg(\sum_{k=0}^{[2\lambda]} + \sum_{k=[2\lambda]+1}^\infty\bigg)\frac{\Delta_k^2}{f(k)}.$$

In the range $0 \leq k \leq [2\lambda]$, we apply the inequality (6.2) which gives

$$\Delta_k^2 \, \leq \, \frac{56^2}{(1-\kappa)^3}\,\Big(\frac{(k-\lambda)^4}{\lambda^2} + 6\,\frac{(k-\lambda)^2}{\lambda^2} + 9\Big)\Big(\frac{\lambda_2}{\lambda}\Big)^2 f(k)^2.$$

Hence

$$S_1 \, \leq \, \frac{56^2}{(1-\kappa)^3}\,\Big(\frac{\mathbb{E}\,(Z-\lambda)^4}{\lambda^2} + 6\,\frac{\mathbb{E}\,(Z-\lambda)^2}{\lambda} + 9\Big)\Big(\frac{\lambda_2}{\lambda}\Big)^2.$$

Using the moment formula $\mathbb{E}\,Z^m = \lambda(\lambda+1)\dots(\lambda+m-1)$, we have $\mathbb{E}\,(Z-\lambda)^2 = \lambda$ and $\mathbb{E}\,(Z-\lambda)^4 = 3\lambda(\lambda+2)$, so that

$$\begin{aligned} S_1 \, &\leq \, \frac{56^2}{(1-\kappa)^3}\,\Big(\frac{3\,(\lambda+2)}{\lambda^2} + 15\Big)\Big(\frac{\lambda_2}{\lambda}\Big)^2 \\ &= \, \frac{18\,816}{(1-\kappa)^3}\,(\lambda^{-1}+3)\Big(\frac{\lambda_2}{\lambda}\Big)^2 \, \leq \, \frac{C_1}{(1-\kappa)^3}\Big(\frac{\lambda_2}{\lambda}\Big)^2 \end{aligned} \tag{6.5}$$

with $C_1 = 94\,080$ (where we used the assumption $\lambda \geq 1/2$ on the last step).

In order to estimate S_2, we use the following elementary bound

$$\sum_{k=k_0}^\infty k^d f(k) \, \leq \, k_0^d\, f(k_0)\,\bigg(1 - \frac{\lambda}{k_0}\,\Big(\frac{k_0+1}{k_0}\Big)^{d-1}\bigg)^{-1}, \tag{6.6}$$

which holds for any $d = 1,2,\dots$ as long as $k_0^d/(k_0+1)^{d-1} > \lambda$. For the proof, write

$$\sum_{k=k_0}^\infty k^d f(k) \, = \, k_0^d f(k_0)\,\big(1 + \theta_1 + \theta_1\theta_2 + \dots + \theta_1\dots\theta_m + \dots\big),$$

where

$$\theta_m = \Big(\frac{k_0+m}{k_0+m-1}\Big)^d\,\frac{\lambda}{k_0+m}, \qquad m = 1,2,\dots$$

Since the function $(x+1)^{d-1}\,x^{-d}$ is decreasing in $x > 0$, we have $1 > \theta_1 > \theta_2 > \dots$ This gives

$$\sum_{k=k_0}^\infty k^d f(k) \, \leq \, k_0^d f(k_0)\,\Big(1 + \sum_{m=1}^\infty \theta_1^m\Big),$$

that is, (6.6). In particular, for $k_0 = [2\lambda] + 1$ and $\lambda \geq 8$ (with $d = 6$),

$$\Big(1 - \frac{\lambda}{k_0}\,\Big(\frac{k_0+1}{k_0}\Big)^5\Big)^{-1} < \Big(1 - \frac{1}{2}\,\Big(\frac{2\lambda+1}{2\lambda}\Big)^5\Big)^{-1} < 3.1.$$

So, by (6.6), and using $[2\lambda] + 1 \leq \frac{17}{8}\lambda$ for the chosen range of λ, we have

$$\sum_{k=[2\lambda]+1}^\infty k^6 f(k) \, \leq \, 3.1\,([2\lambda]+1)^6\, f([2\lambda]+1) \, \leq \, 3.1 \cdot (17\lambda/8)^6\, f([2\lambda]+1).$$

Hence, by (6.3),

$$S_2 \, = \sum_{k=[2\lambda]+1}^\infty \frac{|\Delta_k|^2}{f(k)} \, \leq \, \frac{49^2}{(1-\kappa)^3} \sum_{[2\lambda]+1}^\infty \Big(\frac{k}{\lambda}\Big)^6 \lambda_2^2\, f(k) \, \leq \, \frac{C_2\,\lambda_2^2}{(1-\kappa)^3}\, f([2\lambda]+1) \qquad (6.7)$$

with $C_2 = 49^2 \cdot 3.1 \cdot (17/8)^6 < 685\,343$. Asymptotically with respect to large λ, this bound is much better than (6.4). Applying $f(k) \leq \frac{1}{\sqrt{2\pi k}}\, e^{k-\lambda}\,(\frac{\lambda}{k})^k$ as in (2.9) with $k = [2\lambda] + 1$ and using $2\lambda \leq k \leq 2\lambda + 1$, we have

$$f([2\lambda]+1) \, \leq \, \frac{e}{2\sqrt{\lambda\pi}}\,(e/4)^\lambda \, \leq \, \frac{e}{2\sqrt{\pi}}\ 8^{3/2}\,\Big(\frac{e}{4}\Big)^8\,\frac{1}{\lambda^2} \, < \, \frac{1}{\lambda^2}.$$

This gives

$$S_2 \leq \frac{C_2}{(1-\kappa)^3}\,\Big(\frac{\lambda_2}{\lambda}\Big)^2.$$

As a result, we arrive at the desired upper bound in (6.4).

Finally, let us estimate S_2 for the range $\frac{1}{2} \leq \lambda \leq 8$. Returning to (6.6), we have

$$S_2 \, \leq \, \frac{28^2}{(1-\kappa)^3} \sum_{k=1}^\infty \Big(\frac{k}{\lambda}\Big)^6 \lambda_2^2\, f(k) \, \leq \, \frac{28^2}{(1-\kappa)^3}\,\lambda^{-6}\,\lambda_2^2\,\mathbb{E}Z^6 \, \leq \, \frac{C_2'}{(1-\kappa)^3}\,\Big(\frac{\lambda_2}{\lambda}\Big)^2,$$

where $C_2' = 49^2\, \sup_{\frac{1}{2}\leq\lambda\leq 4} \psi(\lambda)$, $\psi(\lambda) = \lambda^{-4}\,\mathbb{E}Z^6$. Here

$$\psi(\lambda) = \frac{(\lambda+1)\dots(\lambda+5)}{\lambda^3} = \psi_1(\lambda)\psi_2(\lambda)\psi_3(\lambda)$$

with $\psi_1(\lambda) = 5 + \lambda + \frac{4}{\lambda}$, $\psi_2(\lambda) = 7 + \lambda + \frac{10}{\lambda}$, $\psi_3(\lambda) = 1 + \frac{3}{\lambda}$. All these three functions are convex, while ψ_3 is decreasing. In addition, $\psi_i(1/2) \geq \psi_i(8)$ for $i = 1, 2$. Hence $\psi(\lambda) \leq \psi(1/2) = \frac{1}{4} \cdot 11!!$ It follows that $C_2' = 49^2 \cdot \frac{1}{4} \cdot 11!! < 6\,239\,560$, and thus $c = C_1 + C_2''$ is the resulting constant in (6.4). $\square$

Remark 6.3. Assume that $\lambda \geq 1/2$, and let us recall that χ^2 is a stronger distance than the total variation in view of the general relation $d(W,Z)^2 \leq \frac{1}{2}\, D(W,Z)$. Hence, the upper bound in (6.4) implies the inequality $d(W,Z) \leq c_\kappa \lambda_2/\lambda$ (like in Chen's result), provided that $\lambda_2 \leq \kappa\lambda$. But, in the other case $\lambda_2 \geq \kappa\lambda$, there is nothing to prove, since $d(W,Z) \leq 1$.

Also let us recall that, for $\lambda \leq 1/2$, the correct upper bound on the total variation distance is of the form $d(W,Z) \leq C\lambda_2$. It may be obtained by elementary methods, as already illustrated in Proposition 3.3.

7. Uniform Bounds. Comparison with Normal Approximation

A different choice of the parameter r in the proof of Theorem 1.3 may provide various uniform bounds in the Poisson approximation, like in the next assertion. Using the $L^\infty(\mu)$-norm with respect to the counting measure μ on $\mathbb{Z}$, let us focus on the deviations of the densities of W and Z and the deviations of their distribution functions. These distances are thus given by

$$\begin{aligned} M(W,Z) &= \sup_{k\geq 0} \, |\mathbb{P}\{W=k\} - \mathbb{P}\{Z=k\}|, \\ K(W,Z) &= \sup_{k\geq 0} \, |\mathbb{P}\{W\leq k\} - \mathbb{P}\{Z\leq k\}|. \end{aligned}$$

Putting $r=1$ in (5.6), we arrive at the next assertion which sharpens Proposition 4.1.

Theorem 7.1. *We have*

$$M(W,Z) \,\leq\, \frac{\sqrt{e}\,\pi^2}{6}\,\lambda_2 \min\big\{1,(\lambda-\lambda_2)^{-3/2}\big\}. \tag{7.1}$$

This uniform bound is not new; with a non-explicit numerical factor, it corresponds to Theorem 3.1 in Cekanavicius [4], p. 53. For $\lambda \leq 1$, this relation is simplified to

$$M(W,Z) \,\leq\, \frac{\sqrt{e}\,\pi^2}{6}\,\lambda_2,$$

which cannot be improved (modulo a numerical factor) in view of the lower bounds on $|\Delta_k|$ with $k=0,1,2$ mentioned in Section 3. We also have a similar bound for the Kolmogorov distance, $K(W,Z) \leq C\lambda_2$, which follows from the upper bound for the stronger total variation distance as in Theorem 1.1.

When, however, λ is large (and say all $p_j \leq 1/2$), it is commonly believed that it will be more accurate, if we replace the Poisson approximation for P_W by the normal law $N(\lambda,\lambda)$ with mean λ and variance λ. Indeed, suppose, for example, that $p_j = 1/2$, so that W has a binomial distribution with parameters $(n,1/2)$, while the approximating Poisson distribution has parameter $\lambda = n/2$ with $\lambda_2 = n/4$. Here the inequality (1.2) only yields $d(W,Z) \sim 1$, which means that there is no Poisson approximation with respect to the total variation! Nevertheless, the approximation is still meaningful in a weaker sense in terms of the Kolmogorov distance K, as well as in terms of M. In this case, both P_W and P_λ are almost equal to $N(\lambda,\lambda)$, and the Berry-Esseen theorem provides a correct bound $K(W,Z) \leq \frac{c}{\sqrt{n}}$ via the triangle inequality for K. Since $M \leq 2K$ (which holds true for all probability distributions on $\mathbb{Z}$), we also have $M(W,Z) \leq \frac{c}{\sqrt{n}}$. Note that this inequality also follows from Theorem 7.1. Indeed, when $\lambda_2 \leq \frac{1}{2}\,\lambda$, (7.1) is simplified to

$$M(W,Z) \,\leq\, \frac{\sqrt{2e}\,\pi^2}{3}\,\frac{\lambda_2}{\lambda^{3/2}}, \tag{7.2}$$

which yields a correct order for growing n. Thus, the two approaches are equivalent for this particular (i.i.d.) example.

To realize whether or not the normal approximation is better or worse than the Poisson approximation in the general non-i.i.d. situation (that is, with different p_j's), let us evaluate the corresponding Lyapunov ratio in the central limit theorem and apply the Berry-Esseen bound $K(W,N_\lambda) \leq cL_3$, where the random variable N_λ is distributed according to $N(\lambda,\lambda)$. Since $\mathrm{Var}(W) = \sum_{j=1}^n p_j q_j = \lambda - \lambda_2$, the Lyapunov ratio for the sequence $X_1,\dots,X_n$ is given by

$$\begin{aligned} L_3 \ &= \ \frac{1}{\mathrm{Var}(W)^{3/2}} \sum_{j=1}^n \mathbb{E}\,|X_j - \mathbb{E}X_j|^3 \\ &= \ \frac{1}{(\lambda-\lambda_2)^{3/2}} \sum_{j=1}^n (p_j^2 + q_j^2)\, p_j q_j \ \leq \ \frac{1}{\sqrt{\lambda-\lambda_2}} \end{aligned}$$

(note that $\frac{1}{2} \leq p_j^2 + q_j^2 \leq 1$). Hence $K(W,N_\lambda) \leq \frac{c}{\sqrt{\lambda-\lambda_2}}$, up to some absolute constant $c>0$. A similar bound holds for Z as well when representing W as the sum of n independent Poisson random variables Z_j with parameters p_j. Namely, for the sequence $Z_1,\dots,Z_n$, we have

$$L_3 \ = \ \frac{1}{\mathrm{Var}(Z)^{3/2}} \sum_{j=1}^n \mathbb{E}\,|Z_j - \mathbb{E}Z_j|^3 \ \leq \ \frac{c}{\lambda^{3/2}} \sum_{j=1}^n p_j \ = \ \frac{c}{\sqrt{\lambda}}.$$

Therefore, $K(Z,N_\lambda) \leq \frac{c}{\sqrt{\lambda}}$ and hence, by the triangle inequality, $K(W,Z) \leq \frac{c}{\sqrt{\lambda-\lambda_2}}$. In particular, in a typical situation where $\lambda_2 \leq \frac{1}{2}\,\lambda$, the normal approximation yields

$$M(W,Z) \ \leq \ \frac{c}{\sqrt{\lambda}} \tag{7.3}$$

with some absolute constant c. But, this bound is surprisingly worse than (7.2) as long as $\lambda_2 = o(\lambda)$.

Consider as an example $p_j = 1/(2\sqrt{j})$ for $j=1,\dots,n$. Then $\lambda \sim \sqrt{n}$, $\lambda_2 \sim \log n$, and we get $M(W,Z) \leq cn^{-3/4}\log n$ in (7.2), while (7.3) only yields $M(W,Z) \leq cn^{-1/4}$. This example is also illustrative when comparing Theorem 1.2 with (1.5). The first one provides a correct asymptotic $D(W,Z) \sim \frac{\log^2 n}{n}$ (within absolute factors), while (1.5) only gives $D(W,Z) \leq c$.

8. Upper Bounds on D and χ^2 in the Degenerate Case

We now turn to Theorem 1.2 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 involving the quantity

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

Proposition 8.1. *For $\lambda \geq 1/2$, we have*

$$\chi^2(W,Z) \ \leq \ 19\sqrt{Q}, \tag{8.1}$$

$$D(W||Z) \ \leq \ 23\,\log(eQ). \tag{8.2}$$

These bounds turn out to be sharp when $\lambda_2 \geq \kappa\lambda$, cf. Propositions 9.1 and 10.1.

Proof. Setting $g(w) = \prod_{l=1}^n (q_l + p_l w)$ as before, we exploit the representation

$$\mathbb{P}\{W = k\} = \int_{|w|=r} w^{-k} g(w)\, d\mu_r(w) = \frac{1}{2\pi}\, r^{-k} \int_{-\pi}^{\pi} g(re^{i\theta})\, e^{-ik\theta}\, d\theta,$$

which is valid for all $r > 0$. Like in Section 6, we have an upper bound

$$\mathbb{P}\{W = k\} \leq R_k(r)\, I(r) \qquad \text{with} \qquad R_k(r) = r^{-k} \prod_{l=1}^n (q_l + p_l r)$$

and

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

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

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

Moreover, applying $(\frac{e}{k})^k \leq e\sqrt{k}\, \frac{1}{k!}$ as in (2.5), the above is simplified to

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

where $f(k)$ is the density of $Z \sim P_\lambda$ with respect to the counting measure on non-negative integers $k = 0, 1, \dots$

On the other hand, repeating the arguments from Section 5, or just applying (5.4) with $j = 0$, we have for all $|\theta| \leq \pi$,

$$\begin{aligned}
\prod_{l=1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r} \ &= \ \prod_{l=1}^n \Big(1 - \frac{4 q_l p_l r}{(q_l + p_l r)^2} \sin^2 \frac{\theta}{2}\Big)^{1/2} \\
&\leq \ \exp\Big\{-2 \sin^2 \frac{\theta}{2} \sum_{l=1}^n \frac{q_l p_l\, r}{(q_l + p_l r)^2}\Big\} \ \leq \ \exp\Big\{-\frac{2\theta^2}{\pi^2} \sum_{l=1}^n \frac{q_l p_l\, r}{(q_l + p_l r)^2}\Big\}.
\end{aligned}$$

Here

$$\sum_{l=1}^n \frac{q_l p_l\, r}{(q_l + p_l r)^2} \ \geq \ \frac{1}{r} \sum_{l=1}^n q_l p_l = \frac{1}{r}\, (\lambda - \lambda_2) \quad \text{in case} \ \ r \geq 1$$

and

$$\sum_{l=1}^n \frac{q_l p_l\, r}{(q_l + p_l r)^2} \ \geq \ r \sum_{l=1}^n q_l p_l = r\, (\lambda - \lambda_2) \quad \text{in case} \ \ r \leq 1.$$

These right-hand sides have the form

$$\psi(r) = \min\{r, 1/r\}\, (\lambda - \lambda_2),$$

and we get

$$\begin{aligned}
I(r) \ &\leq \ \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\Big\{-\frac{2}{\pi^2}\, \psi(r)\, \theta^2\Big\}\, d\theta \ = \ \frac{1}{4\, \psi(r)^{1/2}} \int_{-2\sqrt{\psi(r)}}^{2\sqrt{\psi(r)}} e^{-\frac{1}{2} x^2}\, dx \\
&\leq \ \frac{1}{4\, \psi(r)^{1/2}} \min\big\{\sqrt{2\pi}, 4\, \psi(r)^{1/2}\big\} \ \leq \ \min\big\{1, \psi(r)^{-1/2}\big\}.
\end{aligned}$$

First, we consider the region $\frac{1}{4}\,\lambda \leq k \leq 4\lambda$, in which case $\frac{1}{4} \leq r \leq 3$ and $\psi(r) \geq \frac{1}{4}\,(\lambda - \lambda_2)$ and thus

$$I(r) \leq \min\Big\{1, \frac{2}{\sqrt{\lambda - \lambda_2}}\Big\} \,\leq\, 2\sqrt{Q_0}, \qquad Q_0 = 1/\max\{1, \lambda - \lambda_2\}.$$

Hence

$$\mathbb{P}\{W = k\} \,\leq\, 2e\sqrt{Q_0}\,\sqrt{k}\,f(k). \tag{8.3}$$

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

$$\mathbb{P}\{W = k\} \,\leq\, \Big(\frac{e\lambda}{k}\Big)^k e^{-\lambda} \,\leq\, e\sqrt{k}\,f(k). \tag{8.4}$$

Now, recall that $\mathbb{P}\{W = 0\} \leq f(0)$ (according to Lemma 3.1) and write

$$\begin{aligned}
\chi^2(W,Z) \ &= \ \sum_{k=0}^\infty \frac{\mathbb{P}\{W = k\}^2}{f(k)} - 1 \,\leq\, S_1 + S_2 + S_3 \\
&= \ \bigg(\sum_{1 \leq k < \frac{1}{4}\,\lambda} + \sum_{\frac{1}{4}\,\lambda \leq k \leq 4\lambda} + \sum_{k > 4\lambda}\bigg)\,\frac{\mathbb{P}\{W = k\}^2}{f(k)}.
\end{aligned}$$

By (8.3),

$$S_2 \,\leq\, 2e\sqrt{Q_0} \sum_{\frac{1}{4}\,\lambda \leq k \leq 4\lambda} \sqrt{k}\ \mathbb{P}\{W = k\} \,\leq\, 4e\sqrt{Q} \sum_{\frac{1}{4}\,\lambda \leq k \leq 4\lambda} \mathbb{P}\{W = k\} \,\leq\, 4e\sqrt{Q}.$$

To estimate S_1, first note that $S_1 = 0$ for $\lambda < 4$. For $\lambda \geq 4$, using the property that the function $k \rightarrow (\frac{e\lambda}{k})^k$ is increasing for $k < \lambda$, we obtain from (8.4) that

$$\begin{aligned}
S_1 \ &\leq \ e^{-\lambda+1} \sum_{k < \frac{1}{4}\,\lambda} \sqrt{k}\,\Big(\frac{e\lambda}{k}\Big)^k \,\leq\, \frac{1}{2}\sqrt{\lambda}\ e^{-\lambda+1} \sum_{1 \leq k < \frac{1}{4}\,\lambda} \Big(\frac{e\lambda}{k}\Big)^k \\
&\leq \ \frac{1}{2}\sqrt{\lambda}\ e^{-\lambda+1} \sum_{1 \leq k < \frac{1}{4}\,\lambda} (4e)^{\lambda/4} \,\leq\, e\,\Big(\frac{\lambda}{4}\Big)^{3/2} \Big(\frac{4}{e^3}\Big)^{\lambda/4} \\
&\leq \ e\,\Big(\frac{3}{2e\log(e^3/4)}\Big)^{3/2} \,<\, 0.544.
\end{aligned}$$

Here we applied the inequality

$$x^p\,c^x \,\leq\, \Big(\frac{p}{e\log(1/c)}\Big)^p \qquad p, x > 0, \quad 0 < c < 1, \tag{8.5}$$

with $p = 3/2$ and $c = 4/e^3$.

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

$$\begin{aligned}
u(x) \ &= \ \log\Big(\sqrt{x}\,\Big(\frac{e\lambda}{x}\Big)^x\Big) - \log(b^x) \\
&= \ \frac{1}{2}\,\log x + x + x\log\lambda - x\log x - x\log b, \qquad k \geq 4\lambda.
\end{aligned}$$

We have

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

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

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

where on the last step we choose $b = 3/4$ and applied (8.5) with $p = 1/2$ and $c = e/3$. Thus, putting $k_0 = [4\lambda] + 1$ and noting that $k_0 \ge 2$, we get

$$\begin{aligned} S_3 \ &\le\ e^{-\lambda+1}\sum_{k>4\lambda}\sqrt{k}\,\Big(\frac{e\lambda}{k}\Big)^{k} \,\le\, \sqrt{e}\sum_{k\ge k_0} A\,\Big(\frac{3}{4}\Big)^{k} \\ &=\ 4A\sqrt{e}\,\Big(\frac{3}{4}\Big)^{k_0} \,\le\, \frac{9}{4}\,A\sqrt{e} \,<\, 5.067. \end{aligned}$$

Finally, using $Q = \lambda Q_0 \ge 1/2$ (due to $\lambda \ge 1/2$), we get $S_1 + S_3 < 5.611 \le 5.611\sqrt{2Q}$. This gives $S_1 + S_2 + S_3 < (5.611\sqrt{2} + 4e)\sqrt{Q} < 18.81\sqrt{Q}$, so (8.1) follows.

Turning to the second assertion and using $\mathbb{P}\{W = 0\} \le f(0)$, write similarly

$$\begin{aligned} D(W||Z) \ &=\ \sum_{k=0}^{\infty}\mathbb{P}\{W = k\}\,\log\frac{\mathbb{P}\{W = k\}}{\mathbb{P}\{Z = k\}} \,=\, T_1 + T_2 + T_3 \\ &\le\ \Big(\sum_{1\le k<\frac{1}{4}\lambda} + \sum_{\frac{1}{4}\lambda\le k\le 4\lambda} + \sum_{k>4\lambda}\Big)\,\mathbb{P}\{W = k\}\,\log\frac{\mathbb{P}\{W = k\}}{f(k)}. \end{aligned}$$

For the region $\frac{1}{4}\,\lambda \le k \le 4\lambda$, we can apply the bound (8.3) again, which gives

$$\mathbb{P}\{W = k\} \,\le\, 2\sqrt{Q_0}\;e\sqrt{k}\,f(k) \,\le\, 4e\sqrt{Q}\,f(k),$$

and therefore, using $Q \ge 1/2$,

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

Using also (8.4) together with the inequality $\log(et) \le t$ ($t > 0$), we obtain, similarly to the derivation of the bound on T_1 in the χ^2-case, that

$$\begin{aligned} T_1 \ &\le\ e^{-\lambda}\sum_{1\le k<\frac{1}{4}\lambda}\Big(\frac{e\lambda}{k}\Big)^{k}\log(e\sqrt{k}) \,\le\, e^{-\lambda}\log(e\sqrt{\lambda/4})\sum_{1\le k<\frac{1}{4}\lambda}\Big(\frac{e\lambda}{k}\Big)^{k} \\ &\le\ \Big(\frac{\lambda}{4}\Big)^{3/2}\Big(\frac{4}{e^3}\Big)^{\lambda/4} \,\le\, \Big(\frac{3}{2e\log(e^3/4)}\Big)^{3/2} \,<\, 0.2. \end{aligned}$$

Putting again $k_0 = [3\lambda] + 1$ similarly to the derivation of the bound on S_3 in the χ^2-case, we also get

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

Hence, $T_1 + T_3 < 5.087 < 16.578\,\log(eQ)$, and (8.2) follows as well. $\square$

9. Lower Bound on χ^2 in the Degenerate Case

Here, we complement Proposition 8.1 with a similar lower bound about χ^2-distance in terms of the same quantity $Q = \lambda/\max\{1, \lambda - \lambda_2\}$. Put $c_0 = 2.5 \cdot 10^{-6}$.

Proposition 9.1. *If $\lambda \geq 1/2$, then with some absolute constant $c \in [c_0, 1)$*

$$1 + \chi^2(W, Z) \geq c\sqrt{Q}. \tag{9.1}$$

Moreover,

$$\chi^2(W, Z) \geq \frac{c}{9}\sqrt{Q} \tag{9.2}$$

as long as $\lambda_2 \geq (1 - \frac{c^2}{4})\,\lambda$.

Suppose that $\lambda_2 \geq (1 - \frac{c^2}{4})\,\lambda$. To derive the second inequality of Proposition 9.1 from the first one, it is sufficient to require that $c\sqrt{Q} \geq 2$, since then $c\sqrt{Q} - 1 \geq \frac{c}{2}\sqrt{Q}$. This condition is fulfilled, as long as $\lambda \geq \lambda_0 = \frac{4}{c^2}$ and then we obtain (9.2). In the remaining case $\frac{1}{2} \leq \lambda \leq \lambda_0$, the inequality (9.2) follows from Harremoës' lower bound $\chi^2(W, Z) \geq \frac{1}{4}\,(\frac{\lambda_2}{\lambda})^2$. Indeed, in this case, $\lambda - \lambda_2 \leq \frac{c^2}{4}\,\lambda \leq 1$, so that $Q = \lambda \leq \frac{4}{c^2}$, and thus $\frac{c}{9}\sqrt{Q} \leq \frac{2}{9}$, while $\frac{1}{4}\,(\frac{\lambda_2}{\lambda})^2 \geq \frac{1}{4}\,(1 - \frac{c^2}{4})^2$.

Thus, we may focus on the first inequality (9.1). First we prove it, assuming that $\lambda - \lambda_2$ is sufficiently large. As in Section 9, for any fixed $r > 0$, we apply the Cauchy theorem and write

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

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

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

and

$$I_k(r) \ = \ \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{l=1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r}\, \exp\Big\{ - ik\theta + i \sum_{l=1}^n \mathrm{Im}\big(\log(q_l + p_l r\, e^{i\theta})\big)\Big\}\, d\theta.$$

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

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

where

$$I_{k1}(r) \ = \ \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \prod_{l=1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r}\, \exp\Big\{ - ik\theta + i \sum_{l=1}^n \mathrm{Im}\big(\log(q_l + p_l r\, e^{i\theta})\big)\Big\}\, d\theta,$$

$$I_{k2}(r) \ = \ \frac{1}{2\pi} \int_{\frac{\pi}{2} < |\theta| < \pi} \prod_{l=1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r}\, \exp\Big\{ - ik\theta + i \sum_{l=1}^n \mathrm{Im}\big(\log(q_l + p_l r e^{i\theta})\big)\Big\}\, d\theta.$$

Here we choose the radius $r = r(k) > 0$ by the condition $R_k'(r) = 0$, or equivalently

$$F(r) \equiv \sum_{l=1}^n \frac{p_l r}{q_l + p_l r} = k. \tag{9.3}$$

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 9.2. *For any $k = 0, \dots, n-1$, the solution $r = r(k)$ to the equation* (9.3) *satisfies*

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

Moreover, in case $|k - \lambda| \leq \frac{1}{6}(\lambda - \lambda_2)$, we have $\frac{5}{6} \leq r \leq \frac{6}{5}$, and actually with some $0 \leq b_i \leq 1$

$$\begin{aligned} r &= 1 + \Big(\frac{6}{5}\Big)^2 b_1 \frac{k - \lambda}{\lambda - \lambda_2} \\ &= 1 + \frac{k - \lambda}{\lambda - \lambda_2} + \Big(\frac{6}{5}\Big)^9 b_2 \frac{\lambda_2 - \lambda_3}{\lambda - \lambda_2} \Big(\frac{k - \lambda}{\lambda - \lambda_2}\Big)^2. \end{aligned}$$

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

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

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

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

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

implying $r(k) \geq \frac{5}{6}$. For $r \geq 1$, one may use $q_l + p_l r \leq r$, which gives $F'(r) \geq \frac{1}{r^2}\,(\lambda - \lambda_2)$ and $F(r) - F(1) \geq (1 - \frac{1}{r})\,(\lambda - \lambda_2)$. Hence, again by the assumption,

$$\frac{1}{6}\,(\lambda - \lambda_2) \geq k - \lambda = F(r(k)) - F(1) \geq \Big(1 - \frac{1}{r(k)}\Big)\,(\lambda - \lambda_2),$$

implying $r(k) \leq \frac{6}{5}$. In both cases, $\frac{5}{6} \leq r(k) \leq \frac{6}{5}$, proving the second assertion of the lemma.

Now, in the interval $\frac{5}{6} \leq r \leq \frac{6}{5}$, we necessarily have $\frac{5}{6} \leq q_l + p_l r \leq \frac{6}{5}$, so

$$\Big(\frac{5}{6}\Big)^2\,(\lambda - \lambda_2) \leq F'(r) \leq \Big(\frac{6}{5}\Big)^2\,(\lambda - \lambda_2).$$

In addition,

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

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

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

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

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

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 $\frac{1}{F'(r_1)} \le (\frac{6}{5})^2\,\frac{1}{\lambda-\lambda_2}$, while

$$\frac{1}{2\,F'(r_2)^3}\,|F''(r_2)| \,\le\, \frac{1}{2\,(\frac{5}{6})^6\,(\lambda-\lambda_2)^3}\cdot 2\cdot\Big(\frac{6}{5}\Big)^3 (\lambda_2-\lambda_3) \,=\, \Big(\frac{6}{5}\Big)^9\,\frac{\lambda_2-\lambda_3}{(\lambda-\lambda_2)^3}.$$

Note that $(\frac{6}{5})^2 = 1.44$ and $(\frac{6}{5})^9 < 5.16$. □

Lemma 9.3. *Let $r = r(k)$ be the solution of* (9.3) *for $0 \le \lambda - k \le \frac{1}{6}\,(\lambda-\lambda_2)$. Then*

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

Proof. The function

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

is vanishing at $r = 1$ and has 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, $F(a) - F(b) \le F'(b)\,(a-b)$ whenever $a \ge b > 0$. In particular, in the interval $r(k) \le r \le 1$, we have

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

which implies

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

By Lemma 9.2, $\frac{5}{6} \le r(k) \le \frac{6}{5}$ and $1 - r(k) \le (\frac{6}{5})^2 \frac{k-\lambda}{\lambda-\lambda_2}$. Moreover, as was shown in the proof, $F'(r(k)) \le (\frac{6}{5})^2 (\lambda - \lambda_2)$. Hence

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

Here the constant $(\frac{6}{5})^7 < 3.6$. $\square$

Lemma 9.4. *Let $\lambda - \lambda_2 \ge 100$. Then, for $0 \le \lambda - k \le \frac{1}{6}\,(\lambda - \lambda_2)$,*

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

Proof. By Lemma 9.2, $1 \ge r(k) \ge \frac{5}{6}$. Recalling (6.4) which is needed with $j = 0$, note that, for $r > 0$ and $-\pi \le \theta \le \pi$,

$$\begin{aligned}\prod_{l=1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r} \ &= \ \prod_{l=1}^n \Big(1 - \frac{4q_lp_lr}{(q_l+p_lr)^2}\,\sin^2\frac{\theta}{2}\Big)^{1/2} \\ &\le \ \exp\Big\{-2\sum_{l=1}^n \frac{q_lp_lr}{(q_l+p_lr)^2}\,\sin^2\frac{\theta}{2}\Big\}.\end{aligned}$$

For $\frac{5}{6} \le r \le 1$, necessarily $q_l + p_l r \le 1$ and

$$\sum_{l=1}^n \frac{q_lp_lr}{(q_l+p_lr)^2} \ \ge\ \sum_{l=1}^n q_lp_lr \ =\ (\lambda-\lambda_2)\,r \ \ge\ \frac{5}{6}\,(\lambda-\lambda_2).$$

Hence

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

Let us now estimate I_{k1}. Using $4q_lp_lr \le (q_l+p_lr)^2$ (since $(q_l-p_lr)^2 \ge 0$), we have, for $|\theta| \le \pi/2$,

$$\frac{4q_lp_lr}{(q_l+p_lr)^2}\,\sin^2\frac{\theta}{2} \ \le\ \frac{1}{2}, \quad l = 1,\dots,n.$$

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

$$\prod_{l=1}^n \frac{|q_l+p_lr\,e^{i\theta}|}{q_l+p_lr} \ \ge\ \exp\Big\{-\log 2\sum_{l=1}^n \frac{4q_lp_lr}{(q_l+p_lr)^2}\,\sin^2(\theta/2)\Big\}.$$

But, any function $w_l(r) = \frac{r}{(q_l+p_lr)^2}$ is increasing in $0 < r \le r_l \equiv q_l/p_l$ and decreasing in $r \ge r_l$. Hence, if $r_l \ge 1$, then $\max_{\frac{5}{6}\le r\le 1} w_l(r) = w(1) = 1$. If $r_l \le \frac{5}{6}$, that is, when $p_l \ge \frac{6}{11}$, we have

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

Finally, if $\frac{5}{6} \leq r_l \leq 1$, which is equivalent to $\frac{1}{2} \leq p_l \leq \frac{6}{11}$, we have

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

Thus, in all cases, $w(r) \leq \frac{6}{5}$ on the interval $\frac{5}{6} \leq r \leq 1$, so that

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

and thus

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

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

$$\begin{aligned}
\frac{1}{2\pi\sqrt{\frac{6}{5}\,\log 4}} \int_{-\frac{\pi}{2}\sqrt{\frac{6}{5}\,(\log 4)\,(\lambda-\lambda_2)}}^{\frac{\pi}{2}\sqrt{\frac{6}{5}\,(\log 4)\,(\lambda-\lambda_2)}} e^{-\frac{1}{2}\,x^2}\, dx \quad &\geq \quad \frac{1}{2\pi\sqrt{\frac{6}{5}\,\log 4}} \int_{-5\pi\sqrt{\frac{6}{5}\,\log 4}}^{5\pi\sqrt{\frac{6}{5}\,\log 4}} e^{-\frac{1}{2}\,x^2}\, dx \\
&= \frac{1}{\sqrt{2\pi\,\frac{6}{5}\,\log 4}}\,\mathbb{P}\Big\{|\xi| \leq 5\pi\sqrt{\frac{6}{5}\,\log 4}\Big\} \,>\, 0.3093,
\end{aligned}$$

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

$$\begin{aligned}
\frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \prod_{l=1}^n \frac{|q_l + p_l r\, e^{i\theta}|}{q_l + p_l r}\,\theta^6\, d\theta \quad &\leq \quad \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \exp\Big\{-\frac{5}{3}\,(\lambda - \lambda_2)\,\sin^2\frac{\theta}{2}\Big\}\,\theta^6\, d\theta \\
&\leq \quad \frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \exp\Big\{-\frac{10}{3\pi^2}\,(\lambda - \lambda_2)\,\theta^2\Big\}\,\theta^6\, d\theta \\
&\leq \quad \frac{1}{\pi}\Big(\frac{20}{3\pi^2}\,(\lambda - \lambda_2)\Big)^{-7/2} \int_{-\infty}^{\infty} e^{-x^2/2}\, x^6\, dx \\
&= \quad \pi^{\frac{13}{2}}\Big(\frac{3}{20}\Big)^{7/2}\, 15\sqrt{2}\,\frac{1}{(\lambda - \lambda_2)^{7/2}} \,<\, \frac{48}{(\lambda - \lambda_2)^{7/2}}.
\end{aligned}$$

The assumption (9.1) may be rewritten as

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

Note that the functions $\mathrm{Im}\big(\log(q_l + p_l r\, e^{i\theta})\big)$ are odd, so their 2nd derivatives are vanishing at zero. We now apply the Taylor formula up to the cubic term to the function

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

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

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

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

$$h(v) = \log(q + pr\, e^{iv}), \qquad p,q,r > 0.$$

We have

$$\begin{aligned}
h'(v) &= \frac{pr\, ie^{iv}}{q + pr\, e^{iv}} \;=\; i\Big(1 - \frac{q}{q + pr\, e^{iv}}\Big) \;=\; i - iq\,(q + pr\, e^{iv})^{-1},\\
h''(v) &= -pqr\ e^{iv}\,(q + pr\, e^{iv})^{-2},\\
h'''(v) &= -pqr\ \Big(ie^{iv}\,(q + pr\, e^{iv})^{-2} - 2i\,pr\, e^{2iv}\,(q + pr\, e^{iv})^{-3}\Big).
\end{aligned}$$

Therefore,

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

implying that

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

But, for $\frac{5}{6} \le r \le 1$ and $|\theta| \le \frac{\pi}{2}$,

$$\begin{aligned}
|q_l + p_l r e^{i\theta}|^2 &= (q_l + p_l r)^2 (1 - \frac{4 q_l p_l r}{(q_l + p_l r)^2}\,\sin^2\frac{\theta}{2}\Big)\\
&\ge (q_l + p_l r)^2 - 2 q_l p_l r \;=\; q_l^2 + p_l^2 r^2.
\end{aligned}$$

Hence

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

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

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

From this,

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

so that

$$\Big|\Big(\mathrm{Im}\,\sum_{l=1}^n \log(q_l + p_l r\, e^{i\theta})\Big)'''\Big| \,\le\, \frac{121}{60}\sum_{l=1}^n p_l q_l + 2\,\Big(\frac{121}{60}\Big)^{3/2}\sum_{l=1}^n q_l p_l^2 \,\le\, c_0\,(\lambda-\lambda_2)$$

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

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

Now, as we mentioned before, the function A_k is odd in θ, so that

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

Hence, using

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

from the previous estimates we may deduce the lower bound

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

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

$$\begin{aligned} I_k(r) &\ge 0.1093\,\frac{1}{\sqrt{\lambda-\lambda_2}} - \frac{1}{2}\,e^{-\frac{5}{6}\,(\lambda-\lambda_2)} \\ &\ge \big(0.1093 - 5\,e^{-\frac{500}{6}}\big)\,\frac{1}{\sqrt{\lambda-\lambda_2}} \,>\, \frac{0.1}{\sqrt{\lambda-\lambda_2}}. \end{aligned}$$

Thus, Lemma 9.4 is proved. $\square$

Proof of Proposition 9.1. We conclude from Lemmas 9.3 and 9.4 that

$$\mathbb{P}\{W = k\} \,\ge\, \frac{1}{10\sqrt{\lambda-\lambda_2}}\,e^{-4\,\frac{(\lambda-k)^2}{\lambda-\lambda_2}} \tag{9.4}$$

for $0 \le \lambda - k \le \frac{1}{6}\,(\lambda-\lambda_2)$ under the assumption $\lambda - \lambda_2 \ge 100$.

On the other hand, according to Lemma 2.3, $f(k) = \mathbb{P}\{Z = k\} \,\le\, \frac{1}{\sqrt{2\pi k}}$. Since $k \ge \lambda - \frac{1}{6}\,(\lambda-\lambda_2) \ge \frac{5}{6}\,\lambda$, we have

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

As a consequence,

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

In order to clarify the last inequality, note that the condition $\lambda-\lambda_2\geq 100$ implies that $\lambda>100$. The above summation is performed over all integers k from the interval $\lambda-\frac{1}{6}\sqrt{\lambda-\lambda_2}\leq x\leq\lambda$ 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 $h=\frac{1}{6}\sqrt{\lambda-\lambda_2}$. Moreover,

$$\begin{aligned}
\sum_{0\leq\lambda-k\leq h} e^{-8\,\frac{(\lambda-k)^2}{\lambda-\lambda_2}} &\geq \sum_{[\lambda-h]+1\leq k\leq[\lambda]} \int_{\lambda-k}^{\lambda-k+1} e^{-\frac{8\,x^2}{\lambda-\lambda_2}}\,dx \\
&= \int_{\lambda-[\lambda]}^{\lambda-[\lambda-h]} e^{-\frac{8x^2}{\lambda-\lambda_2}}\,dx \ \geq\ \frac{1}{4}\sqrt{\lambda-\lambda_2}\int_{2/5}^{2/3} e^{-y^2/2}\,dy \\
&= \frac{\sqrt{2\pi}}{4}\sqrt{\lambda-\lambda_2}\,\big(\Phi(2/3)-\Phi(2/5)\big)\ \geq\ 0.056\sqrt{\lambda-\lambda_2}.
\end{aligned}$$

Here, we used the bounds $4\,\frac{\lambda-[\lambda]}{\sqrt{\lambda-\lambda_2}}\leq\frac{2}{5}$ and $4\,\frac{\lambda-[\lambda-h]}{\sqrt{\lambda-\lambda_2}}\geq 4\,\frac{\lambda-[\lambda-10/6]}{10}\geq\frac{2}{3}$, together with $\Phi(2/3)-\Phi(2/5)>0.09$.

In order to treat the region $\lambda-\lambda_2\leq 100$, we apply Proposition 2.2. Let $W_1=W$ and $W_2=Y_1+\dots+Y_m$, where $Y_1,\dots Y_m$ are independent Bernoulli random variables taking the 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 $\tilde\lambda-\tilde\lambda_2\geq 100$.

Denote by Z_2 a Poisson random variable with $\mathbb{E}Z_2=m/2$ which is independent of $Z_1=Z$. By the previous step and the inequality (2.4) of Proposition 2.2,

$$\begin{aligned}
0.001\sqrt{\frac{\tilde\lambda}{\tilde\lambda-\tilde\lambda_2}} &\leq \chi^2(W_1+W_2,Z_1+Z_2)+1 \\
&\leq (\chi^2(W_1,Z_1)+1)(\chi^2(W_2,Z_2)+1).
\end{aligned}$$

Here, by (8.1), $\chi^2(W_2,Z_2)\leq 19\sqrt{2}$. Moreover, since $\lambda-\lambda_2\leq 100$, we have

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

It follows that

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

Hence, Proposition 9.1 holds in the case $\lambda-\lambda_2\leq 100$ as well.

10. Lower Bound for D in the Degenerate Case

An analogue of Proposition 9.1 is the following statement about the relative entropy. Recall that $Q = \lambda/\max\{1, \lambda - \lambda_2\}$.

Proposition 10.1. *If $\lambda_2 \geq \kappa_0 \lambda$ and $\lambda \geq \lambda_0$, then*

$$D(W||Z) \, \geq \, c_0 \log(eQ). \tag{10.1}$$

where $\kappa_0 = 1 - \exp\{-10^9\}$, $\lambda_0 = \exp\{2 \cdot 10^7\}$, and $c_0 = e^{-14}$.

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

$$w_k \, \equiv \, \mathbb{P}\{W = k\} \, \geq \, \frac{1}{10\sqrt{\lambda - \lambda_2}} \, e^{-4\,\frac{(\lambda-k)^2}{\lambda-\lambda_2}}, \qquad v_k \, \equiv \, \mathbb{P}\{Z = k\} \, \leq \, \frac{1}{\sqrt{2\pi k}},$$

The first one is valid under the conditions $0 \leq \lambda - k \leq \frac{1}{6}\,(\lambda - \lambda_2)$ and $\lambda - \lambda_2 \geq 100$, cf. (9.4). Clearly, they are fulfilled if $0 \leq \lambda - k \leq \frac{5}{3}\sqrt{\lambda - \lambda_2}$ and $\lambda - \lambda_2 \geq 100$, and if additionally $\lambda_2 \geq \kappa\lambda$, $0 < \kappa < 1$, then we have

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

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

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

In order that $w_k \geq v_k$, it is therefore sufficient to require that $\frac{1}{10\sqrt{1-\kappa}}\, e^{-100/9} \geq \frac{1}{\sqrt{5\pi/3}}$, that is, $1 - \kappa \leq \frac{\pi}{60}\, e^{-200/9}$. We have, moreover,

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

Now, applying the general inequality (2.1) of Proposition 2.1, we get

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

Note that, if $\lambda - \lambda_2 \geq 100$, the x-interval $0 \leq \lambda - x \leq \frac{5}{3}\sqrt{\lambda - \lambda_2}$ has length at least $50/3$, so, the total number of integer points in this interval is also at least $50/3$.

Hence, the last sum can be bounded from below by

$$\frac{50/3}{10\sqrt{\lambda-\lambda_2}}\,e^{-100/9}\sum_{0\le\lambda-k\le\frac{5}{3}\sqrt{\lambda-\lambda_2}}1\;\ge\;\frac{5}{3}\,e^{-100/9}\;>\;e^{-11}.$$

Thus,

$$D(W||Z)\;\ge\;\frac{1}{2}\,e^{-11}\log\frac{e\lambda}{\lambda-\lambda_2}-15\;\ge\;\frac{1}{4}\,e^{-11}\,\log\frac{e\lambda}{\lambda-\lambda_2},$$

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

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

if $1-\kappa\le\exp\{-60\,e^{11}\}$. Let us assume that.

The proposition is thus proved under the conditions $\lambda-\lambda_2\ge 100$ and $\lambda_2\ge\kappa\lambda$. It remains to eliminate the first condition, assuming that $\lambda-\lambda_2<100$ and again that $\lambda_2\ge\kappa\lambda$. To this aim, we appeal to Proposition 2.2 again like on the last step of the proof of Proposition 9.1. Namely, using the same notations and assumptions, from the inequality (2.3) and using the previous step, we obtain that

$$\begin{aligned}\frac{1}{4}\,e^{-11}\,\log\frac{e\lambda}{\max\{1,\lambda-\lambda_2\}}\;&\le\;D(W_1+W_2||Z_1+Z_2)\\&\le\;D(W_1||Z_1)+D(W_2||Z_2),\end{aligned}\tag{10.2}$$

where $W_1=W$ and $Z_1=Z$. It holds, as long as $\tilde\lambda_2\ge\kappa\tilde\lambda$, i.e., $\lambda_2+m/4\;\ge\;\kappa\,(\lambda+m/2)$. Since $\lambda-\lambda_2<100$, the latter would follow from

$$\lambda-100+m/4\;\ge\;\kappa\,(\lambda+m/2)$$

which is solved as $\lambda\ge 50\,\frac{\kappa}{1-\kappa}$.

Moreover, by (9.2), we have $D(W_2||Z_2)\le 23\,\log(2e)$. This bound may be used in (10.2), which gives

$$\begin{aligned}D(W||Z)\;&\ge\;\frac{1}{4}\,e^{-11}\,\log\frac{e\lambda}{\max\{1,\lambda-\lambda_2\}}-23\,\log(2e)\\&\ge\;\frac{1}{8}\,e^{-11}\,\log\frac{e\lambda}{\max\{1,\lambda-\lambda_2\}},\end{aligned}$$

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

$$\frac{1}{8}\,e^{-11}\log\frac{e\lambda}{\lambda-\lambda_2}\;\ge\;\frac{1}{8}\,e^{-11}\log\frac{1}{1-\kappa}\;\ge\;23\,\log(2e),$$

if $1-\kappa\le\exp\{-8\cdot 23\cdot\log(2e)\cdot e^{11}\}$. Since the product in the exponent is smaller than 18 700 000, we may choose $\kappa=\kappa_1=1-\exp\{-18\,700\,000\}$. In this case,

$$D(W||Z)\;\ge\;c_1\,\log\frac{e\lambda}{\lambda-\lambda_2},\qquad c_1=\frac{1}{8}\,e^{-11},$$

assuming that $\lambda\ge 50\,\frac{\kappa_1}{1-\kappa_1}$. It remains to note $50\,\frac{\kappa_1}{1-\kappa_1}<\lambda_0$, $\kappa_1<\kappa_0$, $c_1>c_0$. $\square$

11. Summarizing Remarks. Proof of Theorem 1.2

Let us summarize. Using an additional quantity

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

the obtained results on Poisson approximation for different regions of λ and λ_2 are united in Theorem 1.2 in the form of two-sided bounds

$$c_1 \left(\frac{\lambda_2}{\lambda}\right)^2 (1 + \log F) \,\leq\, D(W||Z) \,\leq\, c_2 \left(\frac{\lambda_2}{\lambda}\right)^2 (1 + \log F), \tag{11.1}$$

$$c_1 \left(\frac{\lambda_2}{\lambda}\right)^2 \sqrt{F} \,\leq\, \chi^2(W, Z) \,\leq\, c_2 \left(\frac{\lambda_2}{\lambda}\right)^2 \sqrt{F}, \tag{11.2}$$

which are valid up to some absolute positive constants c_1 and c_2. Let us give final comments on the proof of Theorem 1.2 and indicate these constants explicitly. As we will see, (11.1)-(11.2) hold with $c_1 = 10^{-7}$ and $c_2 = 5.6 \cdot 10^7$.

An upper bound in (12.1).

If $\lambda \leq 1/2$, these bounds are simplified and may be precised as

$$\frac{1}{4} \left(\frac{\lambda_2}{\lambda}\right)^2 \,\leq\, D(W||Z) \,\leq\, \chi^2(W, Z) \,\leq\, 16 \left(\frac{\lambda_2}{\lambda}\right)^2. \tag{11.3}$$

Here, the left inequality is true for all λ and λ_2, cf. [H-J-K], while the right inequality is proved in Proposition 3.3. Note that $\lambda \leq 1/2$ implies $\lambda_2 \leq \frac{1}{2}\,\lambda$.

In the case where $\lambda \geq 1/2$ and $\lambda_2 \leq \frac{1}{2}\,\lambda$, we have, by Proposition 6.2,

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

so that

$$D(W||Z) \,\leq\, 56 \cdot 10^6 \left(\frac{\lambda_2}{\lambda}\right)^2 (1 + \log F) \tag{11.4}$$

In the case where $\lambda \geq 1/2$ and $\lambda_2 > \frac{1}{2}\,\lambda$, one may apply (8.2) which gives

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

Here, the right-hand side contains a better numerical constant in comparison with (11.4), and we finally get (11.1) with constant $c_2 = 56 \cdot 10^6$.

A lower bound in (12.1).

If $\lambda \leq 1$, then $F = 1$, so that the lower bound in (11.3) yields (11.1) with $c_1 = 1/4$.

Assume that $\lambda \geq 1$. The inequality (11.4) may be reversed by virtue of (10.1), which gives

$$D(W||Z) \,\geq\, c_0(1 + \log F) \,\geq\, c_0 \left(\frac{\lambda_2}{\lambda}\right)^2 (1 + \log F) \tag{11.5}$$

with $c_0 = e^{-14}$, provided that $\lambda_2 \geq \kappa_0 \lambda$ and $\lambda \geq \lambda_0$, where $\kappa_0 = 1 - \exp\{-2 \cdot 10^7\}$ and $\lambda_0 = \exp\{2 \cdot 10^7\}$.

But, the remaining regions belong to the non-degenerate case, where F is bounded by a quantity which depends on κ_0 or λ_0, and then the lower bound in (11.3) is asymptotically optimal. Indeed, if $\lambda_2 \leq \kappa_0 \lambda$, then $\log F \leq -\log(1 - \kappa_0) = 2 \cdot 10^7$, so,

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

This means that the left inequality in (11.1) holds true with constant $c_1 = \frac{1}{4\,(1+2\cdot 10^7)}$ which is better than c_0 in the analogous inequality (11.5). Similarly, if $1 \leq \lambda < \lambda_0$, then $F \leq \lambda < \lambda_0$, and we get, by the lower bound in (11.3),

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

This means that the left inequality in (11.1) holds true with the same constant c_1 as above. Thus, the lower bound in (11.1) holds with constant c_0 ($> 10^{-7}$).

An upper bound in (11.2).

If $\lambda \leq 1/2$, we have (11.3), which implies (11.2) with $c_2 = 14$. In the case $\lambda \geq 1/2$ and $\lambda_2 \leq \frac{1}{2}\,\lambda$, a stronger version of (11.4) is still provided by Proposition 6.2, which gives

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

so that (12.2) holds true with $c_2 = 56\cdot 10^6$. In the case where $\lambda \geq 1/2$ and $\lambda_2 > \frac{1}{2}\,\lambda$, one may apply (8.1) which gives

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

Here, the right-hand side contains a better numerical constant, and we finally get (11.2) with the same constant c_2 as in (11.1).

A lower bound in (11.2).

If $\lambda \leq 1$, then $F = 1$, so that the lower bound in (11.3) yields (11.1) with $c_1 = 1/4$.

Assume that $\lambda \geq 1$, in which case $F = Q = \lambda/\max(1,\lambda-\lambda_2)$. By Proposition 9.1, cf. (9.2), we have

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

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

$$\chi^2(W,Z) \,\geq\, \frac{c_0}{9}\Big(\frac{\lambda_2}{\lambda}\Big)^2\sqrt{F}, \tag{11.6}$$

and we obtain the left inequality in (11.2) with $c_1 = c_0/9 > 10^{-7}$.

The remaining region belongs to the non-degenerate case, where F is bounded, and then the lower bound in (11.3) is asymptotically optimal. Indeed, if $\lambda_2 \leq \kappa_0\lambda$, then $1/\sqrt{F} \geq \sqrt{1-\kappa_0} = \frac{c_0}{2} = 0.8\,\cdot 10^{-6}$, so that, by the left inequality in (11.3),

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

This means that the left inequality in (11.1) holds true with constant $c_1 = 2\,\cdot 10^{-7}$ which is slightly better than the constant in the analogous inequality (10.6). Thus, the lower bound in (11.2) holds true with constant $c_1 = 10^{-7}$. $\square$

12. **Difference of Entropies**

For the proof of Corollary 1.4, we shall use another functional

$$H_2(Z) = \big(\mathbb{E}\,(\log v(Z))^2\big)^{1/2} = \Big(\sum_k v_k\,(\log v_k)^2\Big)^{1/2},$$

where Z is an integer-valued random variable with probability function $v_k = \mathbb{P}\{Z = k\}$, $k \in \mathbb{Z}$. Thus, while the Shannon entropy $H(Z) = -\mathbb{E}\,\log v(Z)$ describes the average of the informational content $-\log v(Z)$, the informational quantity $H_2(Z)$ represents the 2nd moment of this random variable.

An application of Theorem 1.2 is based upon the following elementary relation.

Proposition 12.1. *For all integer-valued random variables W and Z with finite entropies, we have*

$$H(W||Z) \leq \chi^2(W,Z) + H_2(Z)\sqrt{\chi^2(W,Z)}. \tag{12.1}$$

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 $w_k = \mathbb{P}\{W = k\}$. Define $t_k = w_k/v_k$ in case $v_k > 0$. Recalling the definition (1.12), 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 $t\log t \leq (t-1) + (t-1)^2$ $(t \geq 0)$, obtaining

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

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

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

$\square$

In view of (12.1), we also need:

Proposition 12.2. *If $Z \sim P_\lambda$, then*

$$H_2(Z) \leq \begin{cases} \sqrt{50}\, \log(1+\lambda), & if \quad \lambda \geq 1, \\ 5\sqrt{\lambda}\, \log(e/\lambda), & if \quad \lambda \leq 1. \end{cases}$$

Proof. As before, let $v_k = \mathbb{P}\{Z = k\}$. In particular, $v_0\,(\log v_0)^2 = \lambda^2\, e^{-\lambda}$ and $v_1\,(\log v_1)^2 = \lambda e^{-\lambda}\,(\lambda + \log(1/\lambda))^2$. This shows that the above upper bound for small λ can be reversed up to a constant. For $\lambda \leq 1$, given $k \geq 1$, from

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

we get

$$\sum_{k\geq 1} v_k\, (\log v_k)^2 \leq \mathbb{E}\, Z^4 \log^2\Big(\frac{e}{\lambda}\Big) \leq 24\,\lambda\, \log^2\Big(\frac{e}{\lambda}\Big),$$

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

Now, assuming that $\lambda \ge 1$, let us apply the lower bounds (2.7)-(2.8) from Lemma 2.3, which for all $k \ge 1$ give

$$\log\frac{1}{v_k} \le 1 + \frac{1}{2}\,\log k + \frac{1}{\lambda}\,(k-\lambda)^2 \le \log(ek) + \frac{1}{\lambda}\,(k-\lambda)^2$$

and

$$\log^2\frac{1}{v_k} \le 2\log^2(e(k+1)) + \frac{2}{\lambda^2}\,(k-\lambda)^4.$$

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

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

Hence $H_2(Z) \le Cx$, $x = \log(1+\lambda) \ge \log 2$, with $C^2 = 2\,(1+\frac{1}{x})^2 + \frac{18}{x^2} < 50$.

Applying the upper bound (2.7) from Lemma 2.3, we also see that this upper bound on H_2 can also be reversed up to a constant.

$\square$

Remark 12.3. With similar arguments, it follows that

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

which can also 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 1.4. By Theorem 1.2 with W as in (1.1) and $Z \sim P_\lambda$, we have

$$\chi^2(W,Z)\ \le\ C\,\Big(\frac{\lambda_2}{\lambda}\Big)^2\,\sqrt{2+\lambda}$$

with some absolute constant C. Using this estimate in (13.1) and applying Proposition 13.2, the desired inequality (1.11) immediately follows (in view of $\lambda_2 \le \lambda$).

To derive a more precise inequality illustrating the asymptotic behaviour in λ in the typical case $\lambda_2 \le \frac{1}{2}\,\lambda$, let us apply once more Theorem 1.2 with its sharper bound

$$\chi^2(W,Z)\ \le\ C\,\Big(\frac{\lambda_2}{\lambda}\Big)^2.$$

By Proposition 12.1, this gives

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

and it remains to note that $1 + H_2(Z) \le C\log(2+\lambda)$, according to Proposition 12.2.

$\square$

Acknowledgement. The authors would like to thank Igal Sason for drawaing our attention to additional references related to the Poisson approximation in informational distances.

References

[1] Andersen, K. F. Weighted inequalities for iterated convolutions. Proc. Amer. Math. Soc. 127 (1999), no. 9, 2643–2651.

[2] Barbour, A. D.; Hall, P. On the rate of Poisson convergence. Math. Proc. Cambridge Philos. Soc. 95 (1984), no. 3, 473–480.

[3] Barbour, A. D.; Holst, L.; Janson, S. Poisson approximation. Oxford Studies in Probability, 2. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1992. x+277 pp.

[4] Čekanavicius, V. Approximation methods in Probability Theory. Universitext. Springer (2016), 274 pp.

[5] Chen, L. H. Y. Poisson approximation for dependent trials. Ann. Probability 3 (1975), no. 3, 534–545.

[6] van Erven, T., Harremoës, P. Rényi divergence and Kullback-Leibler divergence. IEEE Trans. Inform. Theory 60 (2014), no. 7, 3797–3820.

[7] Harremoës, P. Binomial and Poisson distributions as maximum entropy distributions. IEEE Trans. Inform. Theory 47 (2001), no. 5, 2039–2041.

[8] Harremoës, P. Convergence to Poisson distribution in information divergence. Preprint 2, Math. Department, University of Copenhagen, Feb. 2003.

[9] Harremoës, P.; Johnson, O.; Kontoyiannis. Thinning and information projections. arXive:1601.04255, Jan. 2016.

[10] Harremoës, P.; Ruzankin, P. S. Rate of convergence to Poisson law in terms of information divergence. IEEE Trans. Inform. Theory 50 (2004), no. 9, 2145–2149.

[11] Hillion, E.; Johnson, O. A proof of the Shepp-Olkin entropy concavity conjecture. Bernoulli 23 (2017), no. 4B, 3638–3649.

[12] Hodges, J. L., Jr.; Le Cam, L. The Poisson approximation to the Poisson binomial distribution. Ann. Math. Statist. 31 (1960), 737–740.

[13] Johnson, O. Information theory and central limit theorem. Imperial College Press, London, 2004, 209 p.

[14] Kerstan, J. Verallgemeinerung eines Satzes von Prochorow und Le Cam. (German) Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 2 (1964) 173–179.

[15] Kontoyiannis, I.; Harremoës, P.; Johnson, O. Entropy and the law of small numbers. IEEE Trans. Inform. Theory 51 (2005), no. 2, 466–472.

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

[17] LeCam, L. M. Asymptotic Methods in Statistical Decision Theory. Springer Series in Statistics. Springer-Verlag, New York, 1986. xxvi+742 pp.

[18] Liese, F.; Vajda, I. Convex statistical distances. With German, French and Russian summaries. Teubner-Texte zur Mathematik [Teubner Texts in Mathematics], 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987, 224 pp.

[19] Prokhorov, Yu. V. Asymptotic behavior of the binomial distribution. (Russian) Uspehi Matem. Nauk (N.S.) 8 (1953), no. 3 (55), 135–142.

[20] Romanowska, M. A note on the upper bound for the distrance in total variation between the binomial and the Poisson distribution. Statistica Neerlandica 31 (1977), no. 3, 127–130.

[21] Roos, B. On the rate of multivariate Poisson convergence. J. Multivariate Anal. 69 (1999), no. 1, 120–134.

[22] Sason, I. Improved lower bounds on the total variation distance for the Poisson approximation. Statist. Probab. Lett. 83 (2013), no. 10, 2422–2431.

[23] Sason, I. Entropy bounds for discrete random variables via maximal coupling. IEEE Trans. Inform. Theory 59 (2013), no. 11, 7118–7131.

[24] Serfling, R. J. A general Poisson approximation theorem. Ann. Probab. 3 (1975), no. 4, 726–731.

[25] Vajda, I. Theory of Statistical Inference and Information. Kluwer Academic Publishers, Dordrecht-Borston-London, 1989, 432 pp.

[26] Vervaat, W. Upper bounds for the distance in total variation between the binomial or negative binomial and the Poisson distribution. Statistica Neerlandica 23 (1969), 79–86.

Sergey G. Bobkov
School of Mathematics, University of Minnesota
127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455 USA
Email address: bobkov@math.umn.edu

Gennadiy P. Chistyakov
Fakultät für Mathematik, Universität Bielefeld
Postfach 100131, 33501 Bielefeld, Germany
Email address: goetze@mathematik.uni-bielefeld.de

Friedrich Götze
Fakultät für Mathematik, Universität Bielefeld
Postfach 100131, 33501 Bielefeld, Germany
Email address: goetze@mathematik.uni-bielefeld.de