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

S. G. BOBKOV¹, G. P. CHISTYAKOV², AND F. GÖTZE²

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 χ²-distance. The results are based on proper non-uniform estimates
for densities.

1. Introduction

Let $X_1, \ldots, X_n$ 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 $W = X_1 + \cdots + X_n$ takes values $k = 0, 1, \ldots, n$
with probabilities

$P\{W = k\} = \sum p_1^{\varepsilon_1}q_1^{1-\varepsilon_1} \cdots p_n^{\varepsilon_n}q_n^{1-\varepsilon_n}$, \hspace{1cm} (1.1)

where the summation runs over all 0-1 sequences $\varepsilon_1, \ldots, \varepsilon_n$ such that $\varepsilon_1 + \cdots + \varepsilon_n = k$. 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. Namely, putting

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

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

$P\{Z = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$, \hspace{1cm} k = 0, 1, \ldots$

It is well-known for a long time that the distribution $P_\lambda$ approximates the distribution
$P_W$ of $W$, if $\max_{j \leq n} p_j$ is small. In particular, denoting the total variation distance
of $W$ and $Z$ by

$d(W, Z) = \frac{1}{2} ||P_W - P_\lambda||_{TV}$

$= \sup_{A \subset \mathbb{Z}} |P\{W \in A\} - P\{Z \in A\}| = \frac{1}{2} \sum_{k=0}^{\infty} |w_k - v_k|$, \hspace{1cm} 1991 Mathematics Subject Classification. Primary 60E, 60F.

Key words and phrases. χ²-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-1612961.

2) Faculty of Mathematics, University of Bielefeld, Germany; research was partially supported by
SFB 1283.
where \( w_k = \mathbb{P}\{W = k\} \) and \( v_k = \mathbb{P}\{Z = k\} \), there is a remarkable two-sided bound on \( d(W, Z) \) based on the Stein-Chen method in terms of
\[
\lambda_2 = p_1^2 + \cdots + p_n^2
\]
due to Barbour and Hall. Namely, we have:

**Theorem 1.1** (Barbour and Hall [B-H]). One has
\[
\frac{1}{32} \min(1, 1/\lambda) \lambda_2 \leq d(W, Z) \leq \frac{1 - e^{-\lambda}}{\lambda} \lambda_2.
\] (1.2)

The parameter \( \lambda_2 \), or more precisely – the ratio \( \lambda_2/\lambda \) (for \( \lambda \) bounded away from zero), thus plays a similar role as the Lyapunov ratio \( L_3 \) in the central limit theorem.

In the i.i.d. case with \( p_j = \frac{\lambda}{n} \) and fixed \( \lambda > 0 \), both sides of (1.2) are of the same order \( \frac{1}{n} \). Note that, in 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 [P] and bounds for the general case by Le Cam [LC1]. In particular, Le Cam obtained the upper bound
\[
d(W, Z) \leq \lambda_2.
\] (1.3)

For large \( \lambda \) Kerstan [K] and respectively Chen [Ch] improved these bounds to
\[
d(W, Z) \leq \frac{1.05}{\lambda} \lambda_2 \quad \text{if} \quad \max_{j \leq n} p_j \leq \frac{1}{4}, \quad \text{respectively} \quad d(W, Z) \leq \frac{5}{\lambda} \lambda_2.
\]

See also [H-LC], [V], [Se], [R], [Ro] and the references therein. A certain refinement of the lower bound in (1.2) was obtained in Sason [S1].

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. Consider for example, the relative entropy
\[
D(W||Z) = \sum_{k=0}^{\infty} w_k \log \frac{w_k}{v_k},
\]
often called Kullback-Leibler distance, or informational divergence of \( P_W \) from \( P_\lambda \). It dominates the total variation distance in view of the general Pinsker-type inequality
\[
D(W||Z) \geq 2d(W, Z)^2.
\]

In this context, Kontoyiannis, Harremoës and Johnson [K-H-J] have derived the following elegant estimate
\[
D(W||Z) \leq \frac{1}{\lambda} \sum_{j=1}^{n} \frac{p_j^3}{1 - p_j}.
\] (1.4)

In the i.i.d. case where \( p_j = \frac{\lambda}{n} \), this bound yields \( D(W||Z) \leq \frac{2\lambda^2}{n^2} \ (\lambda \leq \frac{n}{2}) \), which has the correct decay with respect to \( n \), thus sharpening Le Cam’s inequality (1.3).

Nevertheless, in the general non-i.i.d. case, Pinsker’s inequality suggests that a sharper bound for \( D(W||Z) \), such as
\[
D(W||Z) \leq A\lambda\lambda_2^3,
\] (1.5)
might hold involving the functional $\lambda_2$ appearing in (1.2) and (1.3), rather than the functional $\lambda_3 = \sum_{j=1}^n p_j^3$. To compare both quantities, note that, by Cauchy’s inequality, we have $\lambda_2^2 \leq \lambda \lambda_3$. Hence, (1.5) would be sharper compared to (1.4), although modulo a $\lambda$-dependent constant. As it turns out, the bound (1.5) does hold (at least, in the non-degenerate case); moreover, one can further sharpen this inequality by replacing the relative entropy with the Pearson $\chi^2$-distance, as well as with other Rényi/Tsallis distances. To avoid technical complications, let us restrict ourselves to the $\chi^2$-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 is dominated by the relative entropy via the inequality

$$
\chi^2(W, Z) \geq D(W || Z).
$$

For a general theory of informational distances, we refer interested readers to the recent review by van Erven and Harremoës [E-H]; additional material may be found in the books [LC2], [L-V], [V], [J].

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

**Theorem 1.2.** We have

$$
D(W || Z) \sim \left( \frac{\lambda_2}{\lambda} \right)^2 \log \frac{2 + \lambda}{\max\{1, \lambda - \lambda_2\}} ,
$$

$$
\chi^2(W, Z) \sim \left( \frac{\lambda_2}{\lambda} \right)^2 \left( \frac{2 + \lambda}{\max\{1, \lambda - \lambda_2\}} \right)^{1/2}.
$$

If $\lambda$ is not large (say $\lambda \leq 10$, which is typical for many applications), or if $\lambda_2/\lambda$ is bounded away from 1 (for instance, when $\max_j p_j \leq \frac{1}{2}$), both equivalences may be simplified to

$$
D(W || Z) \sim \chi^2(W, Z) \sim \left( \frac{\lambda_2}{\lambda} \right)^2 .
$$

(1.7)

Hence in the above regime, (1.5) indeed holds with a factor $A_\lambda \sim 1/\lambda^2$, which tends to infinity as $\lambda$ is approaching zero, in contrast with the lower estimate in (1.2).

Moreover, if the above assumptions on $\lambda$ and $\lambda_2$ are violated (which we call the “degenerate case”), both distances are bounded away from zero and can be large for growing $n$, since then

$$
D(W || Z) \sim \log \frac{\lambda}{\max\{1, \lambda - \lambda_2\}} ,
$$

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

For example, in the case where $p_1 = \cdots = p_n = 1$, we have $\lambda_2 = \lambda = n$. Here $\mathbb{P}\{W = n\} = 1$, hence as $n \to \infty$

$$
D(W || Z) = \log \frac{1}{\mathbb{P}\{Z = n\}} = \log \left( \frac{n!}{n^n e^n} \right) \sim \log n,$$
\[ \chi^2(W, Z) = \frac{1}{P\{Z = n\}} - 1 = \frac{n!}{n^n} e^n - 1 \sim \sqrt{2\pi n}. \]

For the study of the asymptotic behavior of \( D \) and \( \chi^2 \) in terms of \( \lambda \) and \( \lambda_2 \), we derive new bounds on the pointwise deviations \( |w_k - v_k| \) for different zones of \( \lambda \)'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

\[ \left| P\{W = k\} - P\{Z = k\} \right| \leq 2.5 \lambda_2 P\{k - 2 \leq Z \leq k\}. \]

It immediately yields sharp upper bounds for all above distances as in Theorems 1.1-1.2 in the case of small \( \lambda \), at least up to numerical factors.

To treat larger values of \( \lambda \), a more sophisticated analysis in the complex plane is involved – using the closeness of the generating functions associated to the sequences \( w_k \) and \( v_k \). In particular, the following statement may be of independent interest.

**Theorem 1.3.** We have

\[ \left| P\{W = 0\} - P\{Z = 0\} \right| \leq 3 \lambda_2 e^{-\lambda}, \quad (1.8) \]

and for \( k = 1, 2, \ldots , \)

\[ \left| P\{W = k\} - P\{Z = k\} \right| \leq 7 \sqrt{k} \left( \frac{k - \lambda}{\lambda} \right)^2 \lambda_2 \min \{1, \rho^{-1/2}\} P\{Z = k\} \]

\[ + 21 k \sqrt{k} \frac{\lambda_2}{\lambda} \min \{1, \rho^{-3/2}\} P\{Z = k\}, \quad (1.9) \]

where

\[ \rho = (\lambda - \lambda_2) \min \left\{ \frac{k}{\lambda}, \frac{\lambda}{k} \right\}. \]

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

\[ \left| P\{W = k\} - P\{Z = k\} \right| \leq c \left( \frac{(k - \lambda)^2}{\lambda} + 1 \right) \lambda_2 P\{Z = k\}, \]

while for \( k \geq \lambda \geq \lambda_0 \), we also have

\[ \left| P\{W = k\} - P\{Z = k\} \right| \leq c \left( \frac{k}{\lambda} \right)^3 \lambda_2 P\{Z = k\}. \]

Since \( |k - \lambda| \) is of order at most \( \sqrt{\lambda} \) on a sufficiently large part of \( Z \) measured by \( P_{\lambda} \), these non-uniform bounds explain the possibility of upper bounds in the equivalence (1.7).

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

\[ H(W||Z) = H(Z) - H(W), \quad (1.10) \]

where \( H \) stands for the Shannon entropy, that is,

\[ H(Z) = - \sum_k v_k \log v_k, \quad H(W) = - \sum_k v_k \log v_k. \]

The remarkable property that \( H(W||Z) \) is positive represents a consequence of the highly non-trivial assertion that \( H(W) \) is a concave function of the vector \( (p_1, \ldots, p_n) \). The latter was only recently proved by Hillon and Johnson [H-J], cf. also [H]. Thus,
the difference of the 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. [S2]. Here, we show that this distance may be controled 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

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

(1.11)

where $C_{\lambda}$ depends only on $\lambda$. 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 3-4, we consider the deviations $P\{W = k\} - P\{Z = k\}$ for the first values $k = 0, 1, 2$, and prove Theorem 1.2 in case of small $\lambda$. Sections 5-6 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 $\lambda$ in the non-degenerate case. Uniform bounds for large $\lambda$ are discussed in Section 8. There we shall demonstrate that in a typical situation, namely when the ratio $\lambda_2/\lambda$ 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 9-12) and of Corollary 1.4 (Section 13). Thus, the paper is structured as follows.

**Contents**

1. Introduction
2. General bounds on relative entropy and $\chi^2$
3. Poisson and Bernoulli convolutions for $k = 0, 1, 2$
4. Upper bounds in case of small $\lambda$
5. Generating functions
6. Proof Theorem 1.3
7. Consequences from Theorem 1.3
8. Uniform bounds. Comparison with normal approximation
9. Upper bounds on $D$ and $\chi^2$ in the degenerate case
10. Lower bound on $\chi^2$ in the degenerate case
11. Lower bound on $D$ in the degenerate case
12. Summarizing remarks. Proof of Theorem 1.2
13. Difference of entropies

2. **General Bounds on Relative Entropy and $\chi^2$**

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

**Proposition 2.1.** We have
\[ -\sum_{w_k < v_k} w_k \log \frac{w_k}{v_k} \leq 1. \quad (2.1) \]
Moreover,
\[ D(W \| Z) \geq \frac{1}{2} \sum_k \frac{(w_k - v_k)^2}{\max\{w_k, v_k\}}. \quad (2.2) \]

**Proof.** Using the Taylor formula for the logarithmic function, write
\[
\sum_{w_k < v_k} w_k \log \frac{w_k}{v_k} = \sum_{w_k < v_k} (w_k - (v_k - w_k)) \log \left(1 - \frac{v_k - w_k}{v_k}\right)
= \sum_{w_k < v_k} (w_k - v_k) + \sum_{w_k < v_k} \sum_{m=1}^{\infty} \frac{1}{m(m+1)} \frac{(v_k - w_k)^{m+1}}{v_k^m}.
\]
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,
\[
\sum_{w_k > v_k} w_k \log \frac{w_k}{v_k} = -\sum_{w_k > v_k} w_k \log \frac{v_k}{w_k}
= -\sum_{w_k > v_k} w_k \log \left(1 - \frac{w_k - v_k}{w_k}\right)
= \sum_{w_k > v_k} (w_k - v_k) + \sum_{w_k > v_k} \sum_{m=2}^{\infty} \frac{1}{m} \frac{(w_k - v_k)^m}{w_k^m - v_k^{m-1}}.
\]
On the other hand, as was noticed above, there is a similar identity
\[
\sum_{w_k < v_k} w_k \log \frac{w_k}{v_k} = \sum_{w_k < v_k} (w_k - v_k) + \sum_{w_k < v_k} \sum_{m=2}^{\infty} \frac{1}{m(m-1)} \frac{(v_k - w_k)^m}{v_k^m - v_k^{m-1}}.
\]
Adding the two identities and using \(w_k \leq 1\) and \(v_k \leq 1\), 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). \(\square\)

**Proposition 2.2.** Let \(W_1\) and \(W_2\) be independent, non-negative, integer-valued random variables with mean \(\lambda_1\) and \(\lambda_2\) respectively, and let \(Z_1\) and \(Z_2\) be independent Poisson random variables with \(\mathbb{E}Z_1 = \lambda_1\) and \(\mathbb{E}Z_2 = \lambda_2\). Then
\[ D(W_1 + W_2 \| Z_1 + Z_2) \leq D(W_1 \| Z_1) + D(W_2 \| Z_2). \quad (2.3) \]
In addition,
\[ \chi^2(W_1 + W_2, Z_1 + Z_2) + 1 \leq (\chi^2(W_1, Z_1) + 1)(\chi^2(W_2, Z_2) + 1). \] (2.4)

For the proof, we refer to Johnson [J], pp. 133–134. Here, let us only mention that (2.4) is obtained in [J] in the more general form
\[ \sum_{k=0}^{\infty} P(W_1 + W_2 = k)\alpha \leq \sum_{k=0}^{\infty} P(Z_1 + Z_2 = k)\alpha \]
with arbitrary \( \alpha \geq 1 \), which represents a Poisson analog of weighted convolution inequalities due to Andersen [A]. 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
\[ v_k = f(k) = P(Z = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots, \]
with a fixed parameter \( \lambda > 0 \), it is convenient to use the well-known Stirling-type two-sided bound:
\[ \sqrt{2\pi} k^{k+\frac{1}{2}} e^{-k} \leq k! \leq e k^{k+\frac{1}{2}} e^{-k} \quad (k \geq 1). \] (2.5)
In particular, it implies the following Gaussian type estimates.

**Lemma 2.3.** For all \( k \geq 1 \),
\[ f(k) \leq \frac{1}{\sqrt{2\pi k}}. \] (2.6)

Moreover, if \( 1 \leq k \leq 2\lambda \) (with necessarily \( \lambda \geq \frac{1}{2} \)), the number \( \theta \) satisfies \( |\theta| \leq 1 \). In this interval, consider the function \( u(\theta) = h(\theta) + c\theta^2 \) with parameter \( c > 0 \). For \(-1 < \theta \leq 1 \), the 2nd derivative
\[ u''(\theta) = -\frac{1}{1+\theta} + 2c \]
may change the sign at most at one point $\theta_0 > 0$, while $u''(-1) = -\infty$. Since $u(0) = u'(0) = 0$, this means that either $u$ is concave on $[-1, 1]$ and therefore non-positive, or it is concave on $[-1, \theta_0]$ and convex on $[\theta_0, 1]$ (when $0 < \theta_0 < 1$). In this case, $u(\theta) \leq 0$ for all $\theta \in [-1, 1]$, if and only if this inequality is fulfilled at $\theta = 1$. But $u(1) = 1 - 2 \log 2 + c$, so the optimal value is $c = 2 \log 2 - 1 = 0.387... > 1/3$. Hence, $h(\theta) \leq -\frac{1}{2} \theta^2$, and we arrive at the upper bound in (2.7).

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

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

Choosing $c = 1$, consider the same function $u(\theta) = h(\theta) + \theta^2$ in the interval $|\theta| \leq 1$. Since $u''(-\frac{1}{2}) = 0$, it is concave on $[-1, -\frac{1}{2}]$ and convex on $[-\frac{1}{2}, 1]$. Since $u(0) = u'(0) = 0$, $u(-1) = 0$, and $u''(0) = 1$, this means that $\theta = 0$ is the point of local and thus global minimum of $u$. Therefore, $u(\theta) \geq 0$, that is, $h(\theta) \geq -\theta^2$ for all $\theta \in [-1, 1]$.

3. Poisson and Bernoulli Convolutions for $k = 0, 1, 2$

In the next preliminary step, let us compare the distribution of $W$ given in (1.1) and of $Z \sim P_\lambda$ for the first values $k = 0, 1, 2$. As we will see, the value $k = 2$ is most important for obtaining lower bounds on $D$ and $\chi^2$.

**Lemma 3.1.** For all $\lambda > 0$,

$$\mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} \geq (1 - e^{-\frac{1}{2} \lambda^2}) e^{-\lambda}.$$

Moreover, if $\lambda \leq \frac{1}{2}$, then

$$\mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} \geq 0.47 \lambda^2 e^{-\lambda},$$

$$\mathbb{P}\{W = 1\} - \mathbb{P}\{Z = 1\} \geq 0.42 \lambda^2 e^{-\lambda},$$

and if $\lambda \leq \frac{1}{8}$, then also

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

**Proof.** Expanding the function $p \rightarrow -\log(1 - p)$ near zero according to the Taylor formula and recalling (1.1), we get the representation

$$\mathbb{P}\{W = 0\} = q_1 \cdots q_n$$

$$= (1 - p_1) \cdots (1 - p_n) = e^{-\lambda S}, \quad S = \frac{1}{2} \lambda_2 + \frac{1}{3} \lambda_3 + \cdots,$$

where $\lambda_s = p_1^s + \cdots + p_n^s$. Applying the inequality $1 - x \leq e^{-x}$ together with $S \geq \frac{1}{2} \lambda_2$, this gives

$$\mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} = e^{-\lambda} (1 - e^{-S}) \geq e^{-\lambda} (1 - e^{-\frac{1}{2} \lambda^2}).$$

Moreover, if $\lambda \leq \lambda_0 = \frac{1}{2}$, then $\lambda_2 \leq \lambda^2 \leq \frac{1}{4}$, so

$$1 - e^{-\frac{1}{2} \lambda^2} \geq \frac{1 - e^{-1/8}}{1/8} \cdot \frac{1}{2} \lambda_2 \geq 0.47 \lambda_2.$$
This proves the first assertion of the lemma about the case \( k = 0 \).

Now, turning to the case \( k = 1 \), we apply \( \lambda_s \leq \lambda_0 \lambda_2, \ s \geq 3 \), to get an upper bound

\[
S \leq \lambda_2 \sum_{s=2}^{\infty} \frac{\lambda_0^{s-2}}{s} = c \lambda_2, \quad c = \frac{-\log(1 - \lambda_0) - \lambda_0}{\lambda_0^2} = 4 \log 2 - 2,
\]

which gives \( q_1 \ldots q_n \geq e^{-\lambda - c \lambda_2} \). Hence, again according to (1.1),

\[
\mathbb{P}\{W = 1\} = q_1 \ldots q_n \sum_{j=1}^{n} \frac{p_j}{1 - p_j} \geq q_1 \ldots q_n \sum_{j=1}^{n} p_j (1 + p_j)
\]

\[
= q_1 \ldots q_n (\lambda + \lambda_2) \geq (\lambda + \lambda_2) e^{-\lambda - c \lambda_2}.
\]

From this

\[
\mathbb{P}\{W = 1\} - \mathbb{P}\{Z = 1\} \geq (\lambda + \lambda_2) e^{-\lambda - c \lambda_2} - \lambda e^{-\lambda} = e^{-\lambda} (\lambda_2 e^{-c \lambda_2} - \lambda (1 - e^{-c \lambda_2})).
\]

Using again \( 1 - x \leq e^{-x} \) and \( \lambda_2 \leq \lambda^2 \), we see that the expression in the brackets is greater than or equal to

\[
\lambda_2 (1 - c \lambda_2) - c \lambda_2 = \lambda_2 (1 - c (\lambda + \lambda_2)) \geq \lambda_2 \left( 1 - \frac{3}{4} c \right) > 0.42 \lambda_2
\]

which proves the second assertion of the lemma.

Now, consider the case \( k = 2 \). If \( n = 1 \), then \( \mathbb{P}\{W = 2\} = 0 \), so

\[
\mathbb{P}\{Z = 2\} - \mathbb{P}\{W = 2\} = \frac{1}{2} \lambda^2 e^{-\lambda} \geq \frac{1}{2} \lambda_2 e^{-\lambda}.
\]

Let \( n \geq 2 \), in which case

\[
2 \mathbb{P}\{W = 2\} = 2 q_1 \ldots q_n \sum_{1 \leq i < j \leq n} \frac{p_i p_j}{q_i q_j}
\]

\[
= q_1 \ldots q_n \left[ \left( \sum_{j=1}^{n} \frac{p_j}{q_j} \right)^2 - \sum_{j=1}^{n} \left( \frac{p_j}{q_j} \right)^2 \right].
\]

Here the last sum may be bounded from below by \( \lambda_2 \). Since \( p_j \leq \lambda \leq \frac{1}{7} \), we also have \( \frac{1}{1 - p_j} \leq 1 + \frac{8}{7} p_j \), so that the first sum is bounded from above by

\[
\sum_{j=1}^{n} p_j \left( 1 + \frac{8}{7} p_j \right) = \lambda + \frac{8}{7} \lambda_2.
\]

Hence, together with the bound \( q_1 \ldots q_n \leq e^{-\lambda} \), we get

\[
2 e^{\lambda} \mathbb{P}\{W = 2\} \leq e^{-\lambda} \left[ \left( \lambda + \frac{8}{7} \lambda_2 \right)^2 - \lambda_2 \right]
\]

and thus

\[
2 e^{\lambda} \mathbb{P}\{Z = 2\} - 2 e^{\lambda} \mathbb{P}\{W = 2\} \geq \lambda^2 - \left[ \left( \lambda + \frac{8}{7} \lambda_2 \right)^2 - \lambda_2 \right]
\]

\[
= \lambda_2 \left( 1 - \frac{8}{7} \left( 2 \lambda + \frac{8}{7} \lambda_2 \right) \right)
\]

\[
\geq \lambda_2 \left( 1 - \frac{8}{7} \left( 2 \lambda + \frac{8}{7} \lambda^2 \right) \right) \geq \frac{34}{49} \lambda_2.
\]
The lemma is proved. Summing the estimates in Lemma 3.1, we get:

**Corollary 3.2.** If \( \lambda \leq \frac{1}{2} \), then
\[
d(W, Z) \geq 0.44 \lambda_2 e^{-\lambda},
\]
and if \( \lambda \leq \frac{1}{8} \), then moreover
\[
d(W, Z) \geq 0.61 \lambda_2 e^{-\lambda}.
\]

This improves the numerical constant in the lower bound of Theorem 1.1 in the above regions of \( \lambda \)'s.

**Corollary 3.3.** If \( \lambda \leq \frac{1}{8} \), then
\[
D(W \| Z) \geq 0.1 \left( \frac{\lambda_2}{\lambda} \right)^2, \quad \chi^2(W, Z) \geq 0.2 \left( \frac{\lambda_2}{\lambda} \right)^2.
\]

Indeed, one may now combine the lower bound (2.2) of Proposition 2.1 with probabilities
\[
u_2 = \mathbb{P}\{W = 2\} \quad \text{and} \quad \nu_2 = \mathbb{P}\{Z = 2\}
\]
and Lemma 4.1 for \( k = 2 \) in which \( w_2 \geq v_2 \). This gives
\[
D(W \| Z) \geq \frac{1}{2} \left( \frac{w_2 - v_2}{w_2} \right)^2 \geq \frac{1}{2} \left( \frac{17}{49} \lambda_2 e^{-\lambda} \right)^2 \frac{2}{\lambda^2} e^\lambda \geq 0.12 e^{-1/8} \left( \frac{\lambda_2}{\lambda} \right)^2 \geq 0.1 \left( \frac{\lambda_2}{\lambda} \right)^2.
\]
Similarly, \( \chi^2(W, Z) \geq \frac{(w_2 - v_2)^2}{w_2} \).

## 4. Upper Bounds in Case of Small \( \lambda \)

Let us now show that the bounds of Lemma 3.1 and Corollary 3.3 are sharp and, at the expense of a suitable numerical factor, they may be reversed in a similar range of \( \lambda \). We keep the same notations. In particular,
\[
\mathbb{P}\{Z = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots,
\]
while
\[
\mathbb{P}\{W = k\} = \sum \beta_1^{\alpha_1}(1 - p_1)^{1-\alpha_1} \ldots p_n^{\alpha_n}(1 - p_n)^{1-\alpha_n}
\]
with summation over all 0-1 sequences \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n) \) such that \( \varepsilon_1 + \cdots + \varepsilon_n = k \). 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 \( (X_1, \ldots, X_n) \) to \( (X_1, \ldots, X_k) \) in case \( n < k \) with \( p_{n+1} = \cdots = p_k = 0 \). Then the value \( W \) does not change.

First, let us reverse the lower bounds of Lemma 3.1 for the values \( k = 0 \) and \( k = 1 \).

**Lemma 4.1.** If \( \lambda \leq \frac{1}{2} \), then
\[
\mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} \leq 0.8 \lambda_2 e^{-\lambda}, \quad \mathbb{P}\{W = 1\} - \mathbb{P}\{Z = 1\} \leq 2\lambda_2 e^{-\lambda}.
\]
Proof. Expanding the function $p \rightarrow -\log(1-p)$ near zero according to the Taylor formula as in the previous section, write
\[
\mathbb{P}\{W = 0\} = \prod_{j=1}^{n} (1 - p_j) = e^{-\lambda - s}, \quad S = \sum_{s=2}^{\infty} \frac{1}{s} \lambda_s.
\]
Using $\lambda_s \leq (\max_j p_j)^{s-2} \lambda_2 \leq \lambda^{s-2} \lambda_2$ for $s \geq 2$ with $\lambda \leq \frac{1}{2}$, we have
\[
S \leq \lambda_2 \sum_{s=2}^{\infty} \frac{2^{-s}}{s} = (4 \log 2 - 2) \lambda_2 \leq 0.8 \lambda_2.
\]
Hence
\[
\mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\} = e^{-\lambda} (1 - e^{-S}) \leq e^{-\lambda} S,
\]
proving the first inequality.

Next, using the simple representation $\frac{p_j}{1-p_j} = p_j + 2 \theta_j p_j^2$ with $0 \leq \theta_j \leq 1$, we have
\[
\mathbb{P}\{W = 1\} = \prod_{j=1}^{n} (1 - p_j) \sum_{j=1}^{n} \frac{p_j}{1-p_j}
\leq e^{-\lambda - S} (\lambda + 2 \lambda_2) \leq e^{-\lambda} (\lambda + 2 \lambda_2) = \mathbb{P}\{Z = 1\} + 2 \lambda_2 e^{-\lambda},
\]
which yields the second inequality. \qed

Lemma 4.2. If $\lambda \leq \frac{1}{2}$, then for any $k \geq 2$,
\[
|\mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}| \leq \lambda_2 \left( \frac{\lambda^k}{k!} + 2.5 \frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda^{k-2}}{(k-2)!} \right) e^{-\lambda}.
\]

Proof. Representing the Poissonian random variable $Z \sim P_\lambda$ as $Z = Z_1 + \cdots + Z_n$ with independent summands $Z_j \sim P_{p_j}$, we have that, for any $k = 0, 1, \ldots$,
\[
\mathbb{P}\{Z = k\} = e^{-\lambda} \sum_{\alpha_1 + \cdots + \alpha_n = k} \frac{p_1^{\alpha_1} \cdots p_n^{\alpha_n}}{\alpha_1! \cdots \alpha_n!},
\]
where the summation is running over all integers $\alpha_j \geq 0$ such that $\alpha_1 + \cdots + \alpha_n = k$.

Hence, with this assumption, we may start with the formula
\[
\mathbb{P}\{Z = k\} - \mathbb{P}\{W = k\} = e^{-\lambda} \sum_{\alpha_1 + \cdots + \alpha_n = k} \frac{p_1^{\alpha_1} \cdots p_n^{\alpha_n}}{\alpha_1! \cdots \alpha_n!} - \sum_{\alpha_1 + \cdots + \alpha_n = k, \alpha_j \leq 1} T_\alpha,
\]
where
\[
T_\alpha = (1 - p_1) \cdots (1 - p_n) \left( \frac{p_1}{1-p_1} \right)^{\alpha_1} \cdots \left( \frac{p_n}{1-p_n} \right)^{\alpha_n}.
\]
In order to study the asymptotic behavior of the latter quantity, again using the Taylor formula, we have as before
\[
\prod_{j=1}^{n} (1 - p_j) = e^{-\lambda - S}, \quad S = \sum_{s=2}^{\infty} \frac{1}{s} \lambda_s \leq \lambda_2.
\]
and

\[ \prod_{j=1}^{n} \frac{1}{(1 - p_j)^{\alpha_j}} = e^{S_n}, \quad S_\alpha = \sum_{s=1}^{\infty} \frac{1}{s} \sum_{j=1}^{n} \alpha_j p_j^s. \]

Here

\[ S_\alpha \leq (\alpha_1 p_1 + \cdots + \alpha_n p_n) \sum_{s=1}^{\infty} \frac{1}{s} 2^{-(s-1)} = (\alpha_1 p_1 + \cdots + \alpha_n p_n) 2 \log 2. \]

Moreover, since \( \alpha_j \leq 1 \), necessarily \( \alpha_1 p_1 + \cdots + \alpha_n p_n \leq \lambda \leq \frac{1}{2} \), so that \( S_\alpha \leq 2 \log 2 \) and

\[ e^{S_\alpha} - 1 \leq \frac{e^{2 \log 2} - 1}{\log 2} = \frac{1}{\log 2} \]

which implies that

\[ e^{S_\alpha} - 1 \leq \frac{1}{\log 2} S_\alpha \leq \frac{1}{\log 2} (\alpha_1 p_1 + \cdots + \alpha_n p_n) 2 \log 2 = 2 (\alpha_1 p_1 + \cdots + \alpha_n p_n). \]

Thus

\[ \left( \frac{p_1}{1 - p_1} \right)^{\alpha_1} \cdots \left( \frac{p_n}{1 - p_n} \right)^{\alpha_n} = p_1^{\alpha_1} \cdots p_n^{\alpha_n} \left( 1 + 2 \theta (\alpha_1 p_1 + \cdots + \alpha_n p_n) \right), \]

where \( \theta \) denotes a quantity such that \(|\theta| \leq 1 \) (which may be different in different places). Therefore,

\[ T_n e^\lambda = A p_1^{\alpha_1} \cdots p_n^{\alpha_n} \]

with

\[ A = (1 + \theta \lambda_2) \left( 1 + 2 \theta (\alpha_1 p_1 + \cdots + \alpha_n p_n) \right) = 1 + \theta \lambda_2 + 2 \theta (1 + \lambda_2) (\alpha_1 p_1 + \cdots + \alpha_n p_n) = 1 + \theta \lambda_2 + 2.5 \theta \left( \alpha_1 p_1 + \cdots + \alpha_n p_n, \right), \]

where we used \( \lambda_2 \leq \lambda^2 \leq \frac{1}{4} \) on the last step. Thus, for any 0-1 sequence \( \alpha = (\alpha_1, \ldots, \alpha_n) \),

\[ |p_1^{\alpha_1} \cdots p_n^{\alpha_n} - T_\alpha e^\lambda| \leq \lambda \theta p_1^{\alpha_1} \cdots p_n^{\alpha_n} + 2.5 p_1^{\alpha_1} \cdots p_n^{\alpha_n} (\alpha_1 p_1 + \cdots + \alpha_n p_n). \]

Next, applying an obvious identity, we have

\[ \sum_{\alpha_1 + \cdots + \alpha_n = k, \ \alpha_j \leq 1} p_1^{\alpha_1} \cdots p_n^{\alpha_n} \leq \sum_{\alpha_1 + \cdots + \alpha_n = k} \frac{p_1^{\alpha_1} \cdots p_n^{\alpha_n}}{\alpha_1! \cdots \alpha_n!} = \frac{\lambda^k}{k!}, \]

where we always assume that \( \alpha_j \) are non-negative integers. Similarly,

\[ \sum_{\alpha_1 + \cdots + \alpha_n = k, \ \alpha_j \leq 1} p_1^{\alpha_1} \cdots p_n^{\alpha_n} (\alpha_1 p_1 + \cdots + \alpha_n p_n) \]

\[ = \sum_{i=1}^{n} \alpha_i \sum_{\alpha_1 + \cdots + \alpha_n = k, \ \alpha_j \leq 1} p_1^{\alpha_1} \cdots p_i^{\alpha_i} p_i^{\alpha_i+1} \cdots p_n^{\alpha_n} \]

\[ = \sum_{i=1}^{n} p_i^{\alpha_i} \sum_{\alpha_1 + \cdots + \alpha_n = k, \ \alpha_i = 1, \ \alpha_j \leq 1} p_1^{\alpha_1} \cdots p_i^{\alpha_i-1} p_i^{\alpha_i+1} \cdots p_n^{\alpha_n} \]

\[ \leq \sum_{i=1}^{n} p_i^{\alpha_i} \frac{1}{(k-1)!} (p_1 + \cdots + p_{i-1} + p_{i+1} + \cdots + p_n)^{k-1} \leq \frac{\lambda^k}{(k-1)!}. \]
Thus,

$$\sum_{\alpha_1+\ldots+\alpha_n=k, \alpha_i \leq 1} |e^{-\lambda} p_1^{\alpha_1} \cdots p_n^{\alpha_n} - T_\alpha| \leq \lambda_2 \left( \frac{\lambda k}{k!} + 2.5 \frac{\lambda^{k-1}}{(k-1)!} \right) e^{-\lambda}.$$  

For the remaining terms participating in $P(Z = k)$, we similarly have

$$\sum_{\alpha_1+\ldots+\alpha_n=k, \max \alpha_j \geq 2} \frac{p_1^{\alpha_1} \cdots p_n^{\alpha_n}}{\alpha_1! \cdots \alpha_n!} \leq \lambda_2 \sum_{\alpha_1+\ldots+\alpha_n=k-2} \frac{p_1^{\alpha_1} \cdots p_n^{\alpha_n}}{\alpha_1! \cdots \alpha_n!} \leq \lambda_2 \frac{\lambda^{k-2}}{(k-2)!}.$$  

The Lemma is proved. $\blacksquare$

The obtained estimates are sufficient to establish Theorem 1.2 for sufficiently small value of $\lambda$. For completeness, we also include a two-sided bound for the total variation distance (as an illustration of the usefulness of elementary methods).

**Proposition 4.3.** If $\lambda \leq \frac{1}{5}$, then

$$0.6 \lambda_2 \leq d(W, Z) \leq 3.6 \lambda_2,$$

and

$$0.1 \left( \frac{\lambda_2}{\lambda} \right)^2 \leq D(W || Z) \leq \chi^2(W, Z) \leq 28 \left( \frac{\lambda_2}{\lambda} \right)^2.$$  

The upper bounds remain valid in the larger interval $\lambda \leq \frac{1}{2}$.

**Proof.** The lower bounds on the total variation distance and the relative entropy have already been obtained (Corollaries 3.2 and 3.3). As for the first upper bound, one may apply Lemmas 4.1-4.2 with $\lambda \leq \frac{1}{2}$, which give

$$2 \lambda_2^{-1} d(W, Z) \leq e^{-\lambda} \left[ 2.8 + \sum_{k=2}^{\infty} \left( \frac{k^2}{k!} + 2.5 \frac{k^{k-1}}{(k-1)!} + \frac{k^{k-2}}{(k-2)!} \right) \right] = 7.3 + (\lambda - 0.7) e^{-\lambda} \leq 7.3 - 0.2\sqrt{e} < 7.2.$$  

It remains to establish the last upper bound. Applying Lemmas 4.1-4.2, we get

$$\lambda_2^{-2} e^{\lambda} \chi^2(W, Z) \leq 1 + \frac{4}{\lambda} + \sum_{k=2}^{\infty} \frac{k!}{k^2} \left( \frac{k^2}{k!} + 2.5 \frac{k^{k-1}}{(k-1)!} + \frac{k^{k-2}}{(k-2)!} \right)^2.$$  

Opening the brackets, the above sum is equal to

$$\sum_{k=2}^{\infty} \frac{k!}{k^2} \left( \frac{\lambda^{2k}}{k!} + 5 \frac{\lambda^{2k-1}}{k!(k-1)!} + 6.25 \frac{\lambda^{2k-2}}{k!(k-1)!^2} + 2 \frac{\lambda^{2k-2}}{k!(k-1)!} + 5 \frac{\lambda^{2k-3}}{k!(k-1)!(k-2)!} + \frac{\lambda^{2k-4}}{(k-2)!^2} \right)$$

$$= \sum_{k=2}^{\infty} \frac{k^2}{k!} \lambda^k + 5 \sum_{k=2}^{\infty} \frac{k^2}{k!(k-1)!} \lambda^{k-1} + 6.25 \sum_{k=2}^{\infty} k \frac{k^{k-1}}{k!(k-1)!} \lambda^{k-2} + 2 \sum_{k=2}^{\infty} \frac{k^2}{(k-2)!} \lambda^{k-2}$$

$$+ 5 \sum_{k=2}^{\infty} \frac{k}{k!(k-2)!} \lambda^{k-3} + \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^{k-4}}{(k-2)!}.$$  

Relative entropy and $\chi^2$ divergence from the Poisson law
which is the same as
\[ 8e^\lambda - 1 - \lambda + 6.25 \sum_{k=1}^{\infty} (k + 1) \frac{\lambda^{k-1}}{k!} + 5 \sum_{k=0}^{\infty} (k + 2) \frac{\lambda^{k-1}}{k!} + \sum_{k=0}^{\infty} (k + 1)(k + 2) \frac{\lambda^{k-2}}{k!} = 8e^\lambda - 1 - \lambda + 6.25. \]
Being multiplied by \( \lambda^2 \), the last expression is increasing in \( \lambda \) and is equal to \( \frac{275}{16} \sqrt{e} - \frac{7}{2} < 25 \) for \( \lambda = \frac{1}{2} \). Hence
\[ \lambda^2 e^\lambda \chi^2 \left( W, Z \right) < 1 + \frac{4}{\lambda} + \frac{25}{\lambda^2} < \frac{28}{\lambda^2}. \]

\[ \blacksquare \]

5. Generating functions

The probability function \( f(k) = \mathbb{P}\{Z = k\} \) of the Poissonian random variable \( Z \sim P_\lambda \) satisfies the equation \( \lambda f(k - 1) = kf(k) \) in integers \( k \geq 1 \), which immediately implies
\[ \lambda \mathbb{E} h(Z + 1) = \mathbb{E} Zh(Z) \]
for any function \( h \) on \( \mathbb{Z} \) (as long as the expectations exist). This identity was emphasized by Chen \[Ch\] 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 Poissonian with parameter \( \lambda \). 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 \( P_\lambda \) with parameter \( \lambda > 0 \) is given by
\[ \varphi(w) = \mathbb{E} w^Z = \sum_{k=0}^{\infty} \mathbb{P}\{Z = k\} w^k = e^{\lambda(w-1)} = \prod_{j=1}^{n} e^{p_j(w-1)}, \]
which is an entire function of the complex variable \( w \). Correspondingly, the generating function for the distribution of the random variable \( W = X_1 + \cdots + X_n \) in (1.1) is
\[ g(w) = \mathbb{E} w^W = \sum_{k=0}^{\infty} \mathbb{P}\{W = k\} w^k = \prod_{j=1}^{n} (q_j + p_j w), \]
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
\[ \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\} = \int_{|w|=r} w^{-k} (g(w) - \varphi(w)) \, d\mu_r(w), \]
where \( \mu_r \) 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 \( \varphi \) 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 (5.3). Given complex numbers \( a_j, b_j \ (1 \leq j \leq n) \), we have an identity

\[
a_1 \ldots a_n - b_1 \ldots b_n = \sum_{j=1}^{n} (a_j - b_j) \prod_{l<j} b_l \prod_{l>j} a_l \tag{5.4}
\]

with the convention that \( \prod_{l<j} b_l = 1 \) in case \( j = 1 \) and \( \prod_{l>j} a_l = 1 \) in case \( j = n \), which implies

\[
\left| \prod_{j=1}^{n} a_j - \prod_{j=1}^{n} b_j \right| \leq \sum_{j=1}^{n} |a_j - b_j| \prod_{l<j} |b_l| \prod_{l>j} |a_l|.
\]

According to the product representations (5.1)-(5.2) to be used in (5.3), one should choose here \( a_j = q_j + p_j w \) and \( b_j = e^{p_j(w-1)} \) with \( |w| = r \). Then

\[
|a_j| \leq q_j + p_j r \leq e^{p_j(r-1)}, \quad |b_j| = e^{p_j(Re w-1)} \leq e^{p_j(r-1)}. \tag{5.5}
\]

Therefore

\[
|g(w) - \varphi(w)| \leq \sum_{j=1}^{n} |a_j - b_j| \prod_{l \neq j} e^{p_l(r-1)} = e^\lambda(r-1) \sum_{j=1}^{n} |a_j - b_j| e^{-p_j(r-1)}. \tag{5.6}
\]

To estimate the terms in this sum, consider the function \( \xi(u) = (1+u) - e^u \), \( u \in \mathbb{C} \), and write down it by virtue of the Taylor integral formula as

\[
\xi(u) = -u^2 \int_{0}^{1} e^{tu} (1-t) \, dt.
\]

If \( \text{Re} \, u \leq 0 \), then

\[
|u^2 e^{tu}| = |u|^2 e^{\text{Re} \, u} \leq |u|^2,
\]

so \( |\xi(u)| \leq \frac{1}{2} |u|^2 \). 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 (5.6) 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 (5.3), we then arrive at the uniform bound:

**Proposition 5.1.** We have

\[
\sup_{k \geq 0} |\mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}| \leq \lambda_2. \tag{5.7}
\]

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 (5.7), with the right-hand side properly depending on \( k \).
6. Proof of Theorem 1.3

Applying (5.4) with \(a_j = q_j + p_j w\) and \(b_j = e^{p_j(w-1)}\) in (5.3), one may write this formula as

\[
P\{W = k\} - P\{Z = k\} = \sum_{j=1}^{n} T_j(k), \quad k = 0, 1, \ldots, \tag{6.1}
\]

with

\[
T_j(k) = \int_{|w|=r} w^{-k} (a_j - b_j) \prod_{l<j} b_l \prod_{l>j} a_l \ d\mu_r(w), \tag{6.2}
\]

where the integration is performed over the uniform probability \(\mu_r\) measure 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 (5.5), we get

\[
|T_j(k)| \leq r^{-k} \int_{|w|=r} |a_j - b_j| \prod_{l<j} |e^{p_l(w-1)}| \prod_{l>j} |q_l + p_l w| \ d\mu_r(w)
\]

\[
= r^{-k} \int_{|w|=r} |a_j - b_j| \exp \left\{ (r \cos \theta - 1) \sum_{l=1}^{j-1} p_l \right\} \prod_{l=j+1}^{n} |q_l + p_l w| \ d\mu_r(w)
\]

\[
= r^{-k} e^{(r-1) \sum_{l=1}^{j-1} p_l} \int_{|w|=r} |a_j - b_j| \exp \left\{ -2r \sin^2 \left( \frac{\theta}{2} \right) \sum_{l=1}^{j-1} p_l \right\} \prod_{l=j+1}^{n} |q_l + p_l w| \ d\mu_r(w).
\]

Here, in order to estimate \(|a_j - b_j|\), let us return to the function

\[
\xi(w) = (1 + u) - e^u = -u^2 \int_0^1 e^{tu} (1 - t) \ dt,
\]

which we need at the values \(u_j = p_j(w-1)\) with \(|w| = r\).

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

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

so

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

**Case** \(0 < r \leq 1\). Then \(\text{Re } u_j \leq 0\) and \(|u_j^2 e^{tu_j}| \leq |u_j|^2\). Hence

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

Since \(|w - 1|^2 = (r - 1)^2 + 4r \sin^2(\theta/2)\), we therefore obtain from (6.2) that

\[
|T_j(k)| \leq \frac{1}{2} p_j^2 R_j(r) r^{-k} \left( (r - 1)^2 I_{j0}(r) + 4r I_{j2}(r) \right), \tag{6.3}
\]

where

\[
R_j(r) = \exp \left\{ (r - 1) \sum_{l=1}^{j} p_l \right\} \prod_{l=j+1}^{n} (q_l + p_l r), \quad r \geq 1,
\]

\[
R_j(r) = \exp \left\{ (r - 1) \sum_{l=1}^{j-1} p_l \right\} \prod_{l=j+1}^{n} (q_l + p_l r), \quad r \leq 1,
\]
Here we applied the inequalities in case

\[ \gamma_j(\theta) = r \left( \sum_{l=1}^{j-1} p_l + \sum_{l=j+1}^{n} \frac{q_lp_l}{(q_l + pr)^2} \right). \]

Thus, we need to bound \( \gamma_j \) from below. If \( r \geq 1 \), then \( q_l + p_l r \leq r \), so

\[ \sum_{l=j+1}^{n} \frac{q_lp_l r}{(q_l + pr)^2} \geq \frac{1}{r} \sum_{l=j+1}^{n} q_lp_l. \]

This gives

\[ \gamma_j(r) \geq r \sum_{l=1}^{j-1} p_l + \frac{1}{r} \sum_{l=j+1}^{n} q_lp_l \]

\[ = r \sum_{l=1}^{j-1} p_l + \frac{1}{r} \sum_{l=1}^{n} (p_l - p_l^2) - \frac{1}{r} \sum_{l=1}^{j} (p_l - p_l^2) \]

\[ = \left( r - \frac{1}{r} \right) \sum_{l=1}^{j-1} p_l + \frac{1}{r} \sum_{l=1}^{j-1} p_l^2 + \frac{1}{r} \sum_{l=1}^{n} (p_l - p_l^2) - \frac{1}{r} q_j p_j \geq \frac{1}{r} (\lambda^2 - \lambda \lambda_2 - q_j p_j). \]

In case \( r \leq 1 \), we use \( q_l + p_l r \leq 1 \), implying that

\[ \sum_{l=j+1}^{n} \frac{q_lp_l}{(q_l + pr)^2} \geq \sum_{l=j+1}^{n} q_lp_l. \]
Therefore in this range we have a similar lower bound, namely

\[ \gamma_j(r) \geq r \sum_{l=1}^{j-1} p_l + r \sum_{l=j+1}^{n} q_l p_l \]

\[ = r \sum_{l=1}^{j-1} p_l + r \sum_{l=1}^{n} (p_l - p_l^2) - r \sum_{l=1}^{j} (p_l - p_l^2) \]

\[ = -r p_j + r \sum_{l=1}^{j} p_l^2 + r \sum_{l=1}^{n} (p_l - p_l^2) \geq r (\lambda - \lambda_2 - q_j p_j). \]

Since \( q_j p_j \leq \frac{1}{4} \), both lower bounds yield

\[ \gamma_j(r) \geq \psi(r) - \frac{1}{4}, \quad \psi(r) = \min\{r, 1/r\} (\lambda - \lambda_2). \]

As a result, (6.4) is simplified to

\[ I_{jm}(r) \leq \frac{1}{2\pi} 2^{-m} \sqrt{e} \int_{-\pi}^{\pi} |\theta|^m \exp \left\{ -\frac{2}{\pi^2} \psi(r) \theta^2 \right\} d\theta \]

\[ = \sqrt{e} \frac{\pi^m}{4^{m+1}} \psi(r)^{-\frac{m+1}{2}} \min\left\{ \sqrt{2\pi E |\xi|^m}, \frac{2^{m+2}}{m+1} \psi(r)^{\frac{m+1}{2}} \right\} \]

\[ \leq \sqrt{e} \frac{\pi^m}{4^{m+1}} \max\left\{ \sqrt{2\pi E |\xi|^m}, \frac{2^{m+2}}{m+1} \right\} \min\left\{ 1, \psi(r)^{-\frac{m+1}{2}} \right\}, \]

where \( \xi \) is a standard normal random variable. In particular, we get the upper bounds

\[ I_{j0}(r) \leq \sqrt{e} \min\left\{ 1, \psi(r)^{-1/2} \right\}, \quad I_{j2}(r) \leq \frac{\sqrt{e} \pi^2}{12} \min\left\{ 1, \psi(r)^{-3/2} \right\}. \]

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

\[ R_j(r) \leq \exp \left\{ (r - 1) \sum_{l=1}^{n} p_l \right\} = e^{\lambda(r-1)} \]

in case \( r \geq 1 \), while for \( r \leq 1 \)

\[ R_j(r) \leq \exp \left\{ (r - 1) \sum_{l \neq j} p_l \right\} = e^{\lambda(r-1)} e^{-p_j(r-1)} \leq e^{\lambda(r-1)+1}. \]

Applying these bounds in (6.3), we therefore obtain that

\[ |T_j(k)| \leq \frac{\delta r}{2} p_j^2 e^{\lambda(r-1)+\frac{1}{2}} r^{-k} \left( (r - 1)^2 \min\left\{ 1, \psi(r)^{-1/2} \right\} + \frac{\pi^2}{3} r \min\left\{ 1, \psi(r)^{-3/2} \right\} \right), \]
where \( \delta_r = 1 \) in case \( r \geq 1 \) and \( \delta_r = e \) for \( r < 1 \). Summing over \( j \leq n \) and recalling (6.1), one can estimate \( |\mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}| \) from above by

\[
\lambda_2 \delta_r e^{\lambda(r-1)} r^{-k} \left( \frac{\sqrt{\pi} (r-1)^2}{2} \min\{1, \psi(r)^{-1/2}\} + \frac{\sqrt{\pi} \pi^2}{6} r \min\{1, \psi(r)^{-3/2}\} \right). \tag{6.5}
\]

**Case** \( k = 0 \). Letting \( r \to 0 \), we then get

\[
|\mathbb{P}(W = 0) - \mathbb{P}(Z = 0)| \leq \frac{e^{\beta/2}}{2} \lambda_2 e^{-\lambda}
\]

where the constant factor is smaller than 3. This gives the desired inequality (1.8).

**Case** \( k \geq 1 \). We use (6.5) with \( r = \frac{k}{\lambda} \) and apply \( k! \leq e^{k+\frac{1}{3}} e^{-k} \), cf. (2.5), giving

\[
e^{\lambda(r-1)} r^{-k} = \left( \frac{e\lambda}{k} \right)^k e^{-\lambda} \leq e\sqrt{k} f(k), \quad f(k) = \frac{\lambda^k}{k!} e^{-\lambda}.
\]

To simplify the numerical constants, note that \( e^{5/2} < 6.1 \) and \( e^{5/2} \pi^2 < 20.1 \). Recalling that \( \psi(r) = \rho \) for \( r = k/\lambda \), we can bound (6.5) by

\[
\lambda_2 \sqrt{k} f(k) \left( 7 \left( \frac{k-\lambda}{\lambda} \right)^2 \min\{1, \rho^{-1/2}\} + 21 \frac{k}{\lambda} \min\{1, \rho^{-3/2}\} \right), \tag{6.6}
\]

which proves the second inequality (1.9). \( \square \)

### 7. Consequences of Theorem 1.3

To clarify the meaning of (1.9), let us simplify this bound under the natural requirement that \( \lambda_2 \) is bounded away from \( \lambda \), and assuming that \( \lambda \) is bounded away from zero. As before, we use the notations

\[
f(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \lambda = p_1 + \cdots + p_n,
\]

and \( \lambda_2 = p_1^2 + \cdots + p_n^2 \). Note that \( \lambda_2 \leq \lambda \).

**Corollary 7.1.** Suppose that \( \lambda_2 \leq \kappa \lambda \) with some constant \( \kappa \in (0, 1) \). Then, for any integer \( k \geq 0 \),

\[
|\mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}| \leq \frac{21}{(1-\kappa)^{3/2}} \left( \frac{(k-\lambda)^2}{\lambda} + 1 \right) \frac{\lambda_2}{\lambda} \max\left\{ \left( \frac{k}{\lambda} \right)^3, 1 \right\} f(k). \tag{7.1}
\]

In particular, if \( k \leq 2\lambda \) and \( \lambda \geq 1/8 \), then with some absolute constants \( c_j \),

\[
|\mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}| \leq \frac{c_1}{(1-\kappa)^{3/2}} \left( \frac{(k-\lambda)^2}{\lambda} + 1 \right) \frac{\lambda_2}{\lambda} f(k) \tag{7.2}
\]

\[
\leq \frac{c_2}{(1-\kappa)^{3/2}} \lambda_2 f(k). \tag{7.3}
\]

If \( k \geq \lambda \geq 1/8 \), we also have

\[
|\mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\}| \leq \frac{c_3}{(1-\kappa)^{3/2}} \left( \frac{k}{\lambda} \right)^3 \lambda_2 f(k). \tag{7.4}
\]

**Proof.** Indeed, the assumption \( \lambda_2 \leq \kappa \lambda \) ensures that \( \rho \geq (1-\kappa)\lambda \min\{\frac{k}{\lambda}, \frac{\lambda}{k}\} \).
First suppose that $1 \leq k \leq K\lambda$ ($K \geq 1$). Then $\frac{k}{\lambda} \leq K^2 \frac{\lambda}{k}$, so that $\rho \geq \frac{1-\kappa}{\lambda^2} k$. This gives $\rho^{-1/2} \leq K \frac{1}{\sqrt{(1-\kappa)k}}$ and $\rho^{-3/2} \leq K^3 \frac{1}{(1-\kappa)^{3/2} k^{3/2}}$. Choosing $K = \max\{\frac{k}{\lambda}, 1\}$, the expression (6.6) does not exceed the right-hand side of (7.1), so the inequality (1.9) yields (7.1), which in turn immediately implies (7.2)-(7.3).

In case $k = 0$, we apply the inequality (1.8). Since $\frac{(k-\lambda)^2}{\lambda^2} + 1 \geq \lambda$, the right-hand side of (1.8) is also dominated by the right-hand side of (7.1). Thus, we obtain (7.1) without any constraints on $k$, and (7.2)-(7.3) for all $k \leq 2\lambda$.

In case $k \geq \lambda$, we have $\rho \geq (1-\kappa) \frac{\lambda^2}{k}$ which gives $\rho^{-1/2} \leq \frac{\sqrt{k}}{\lambda \sqrt{1-\kappa}}$ and $\rho^{-3/2} \leq \frac{k^{3/2}}{\lambda^3 (1-\kappa)^{3/2}}$. Using also $(\frac{k-\lambda}{\lambda})^2 \leq \frac{k^2}{\lambda^2}$, (7.1) yields the last bound (7.4).

We are now prepared to extend Proposition 4.3 to larger values of $\lambda$ under the additional assumption that $\lambda_2/\lambda$ is bounded away from 1. The next assertion yields, when combined with Proposition 4.3, Theorem 1.2 in the non-degenerate case.

**Proposition 7.2.** If $\lambda \geq 1/8$ and $\lambda_2 \leq \kappa \lambda$ with $\kappa \in (0, 1)$, then

$$c_0 \left( \frac{\lambda_2}{\lambda} \right)^2 \leq D(W, Z) \leq \chi^2(W, Z) \leq c\kappa \left( \frac{\lambda_2}{\lambda} \right)^2.$$  \hspace{1cm} (7.5)

where $c_0 > 0$ is a universal constant, while the constant $c_\kappa > 0$ depends on $\kappa$, only.

**Proof.** In view of the Pinsker inequality, the left lower bound in (7.5) follows from Barbour-Hall’s lower bound in Theorem 1.1. Hence, one may only focus on the right upper bound in (7.5). Write

$$\chi^2(W, Z) = \sum_{k=0}^{\infty} \frac{\left( \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\} \right)^2}{\mathbb{P}\{Z = k\}} = S_1 + S_2$$

$$= \left( \sum_{k=0}^{[2\lambda]} + \sum_{[2\lambda]}^{\infty} \right) \frac{\left( \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\} \right)^2}{\mathbb{P}\{Z = k\}}.$$

In the case $0 \leq k \leq [2\lambda]$ we apply the inequality (7.2). Up to some constant $c = c(\kappa) > 0$, it gives

$$c \left( \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\} \right)^2 \leq \left( \frac{(k-\lambda)^4}{\lambda^2} + 1 \right) \left( \frac{\lambda_2}{\lambda} \right)^2 f(k)^2.$$  \hspace{1cm} (7.6)

Hence

$$c S_1 \leq \left( \frac{\lambda_2}{\lambda} \right)^2 \left( 1 + \sum_{k=0}^{[2\lambda]} \frac{(k-\lambda)^4}{\lambda^2} \right) f(k) = \left( \frac{\lambda_2}{\lambda} \right)^2 \left( 1 + \frac{\mathbb{E}(Z-\lambda)^4}{\lambda^2} \right).$$

But, using the moment formula for the Poisson distribution

$$\mathbb{E} Z^m = \lambda (\lambda + 1) \ldots (\lambda + m - 1),$$

we easily find that $\mathbb{E} (Z-\lambda)^4 = 3\lambda(\lambda + 2)$, so that $S_1 \leq C \left( \frac{\lambda_2}{\lambda} \right)^2$, where the constant $C$ depends on $\kappa$ only.

For the values $k \geq 2\lambda$, we apply the inequality (7.4), which gives

$$\left| \mathbb{P}\{W = k\} - \mathbb{P}\{Z = k\} \right|^2 \leq C k^6 f(k)^2 \frac{\lambda_2^2}{\lambda^6}.$$
Applying $f(k) \leq \frac{1}{\sqrt{2\pi k}} e^{k-\lambda} \left(\frac{\lambda}{k}\right)^k \leq e^{k-\lambda} \left(\frac{\lambda}{k}\right)^k$ as in (2.8), we get

$$S_2 \leq C \lambda^2 \lambda^{-6} e^{-\lambda} \sum_{k=\lfloor 2\lambda \rfloor+1}^{\infty} k^6 \left(\frac{e^{\lambda}}{k}\right)^k \leq C' \lambda^2 e^{-\lambda} \sum_{k=\lfloor 2\lambda \rfloor+1}^{\infty} k^6 \left(\frac{\lambda}{k}\right)^k \leq C'' \lambda^2 \lambda^{-6}$$

for some constants $C, C', C'' > 0$ depending on $\kappa$, only. It remains to collect together the two bounds for $S_1$ and $S_2$.

**Remark 7.3.** Assume that $\lambda \geq 1/8$, and let us recall that the $\chi^2$ 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 (7.5) implies the inequality $d(W, Z) \leq c \lambda \lambda^{-2}$ up to a numerical factor $c$ (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 $\lambda \leq 1/8$, 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 4.3.

**8. 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 $\mu$ on $Z$, 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

$$M(W, Z) = \sup_{k \geq 0} |P\{W = k\} - P\{Z = k\}|,$$

$$K(W, Z) = \sup_{k \geq 0} |P\{W \leq k\} - P\{Z \leq k\}|.$$

Putting $r = 1$ in (6.9), we arrive at the next assertion which sharpens Proposition 5.1.

**Theorem 8.1.** We have

$$M(W, Z) \leq \sqrt{\frac{e}{\pi}} \frac{\pi^2}{6} \lambda_2 \min\{1, (\lambda - \lambda_2)^{-3/2}\}. \quad (8.1)$$

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

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

which cannot be improved in view of Lemma 3.1 (modulo a numerical factor). 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, $\lambda$ 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 $\lambda$ and variance $\lambda$. 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 $P_N$ and $P_\lambda$ 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
8.1. Indeed, when $\lambda_2 \leq \frac{1}{2} \lambda$, (8.1) is simplified to
\[ M(W, Z) \leq \frac{\sqrt{2e} \pi^2}{3} \frac{\lambda_2}{\lambda^{3/2}}, \tag{8.2} \]
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 $N_\lambda$ is
distributed according to $N(\lambda, \lambda)$. Since $\text{Var}(W) = \sum_{j=1}^n p_j q_j = \lambda - \lambda_2$, the Lyapunov
ratio for the sequence $X_1, \ldots, X_n$ is given by
\[ L_3 = \frac{1}{\text{Var}(W)^{3/2}} \sum_{j=1}^n \mathbb{E}|X_j - \mathbb{E}X_j|^3 \]
\[ = \frac{1}{(\lambda - \lambda_2)^{3/2}} \sum_{j=1}^n (p_j^2 + q_j^2) p_j q_j \leq \frac{1}{\sqrt{\lambda - \lambda_2}} \]
(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 $Z_1, \ldots, Z_n$, we have
\[ L_3 = \frac{1}{\text{Var}(Z)^{3/2}} \sum_{j=1}^n \mathbb{E}|Z_j - \mathbb{E}Z_j|^3 \leq \frac{c}{\lambda^{3/2}} \sum_{j=1}^n p_j = \frac{c}{\sqrt{\lambda}}. \]
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
\[ M(W, Z) \leq \frac{c}{\sqrt{\lambda}} \tag{8.3} \]
with some absolute constant $c$. But, this bound is surprisingly worse than (8.2) as
long as $\lambda_2 = o(\lambda)$.

Consider as an example $p_j = 1/(2\sqrt{j})$ for $j = 1, \ldots, n$. Then $\lambda \sim \sqrt{n}$, $\lambda_2 \sim \log n$, and
we get $M(W, Z) \leq cn^{-3/4} \log n$ in (8.2), while (8.3) only yields $M(W, Z) \leq cn^{-1/4}$. This example is also illustrative when comparing the inequalities (7.5) and
(1.4). The first one provides a correct asymptotic $D(W, Z) \sim \log^2 \frac{n}{n}$ (within absolute factors), while (1.4) only gives $D(W, Z) \leq c$. 
9. Upper Bounds on $D$ and $\chi^2$ in the Degenerate Case

We now turn to Theorem 1.2 in the degenerate case, where the optimal bounds on the relative entropy and $\chi^2$ have a different behavior. As an intermediate step, let us derive the following upper bounds for the $\chi^2$-distance and the relative entropy.

**Proposition 9.1.** For $\lambda \geq 1/8$, we have
\[
\chi^2(W, Z) \leq c \sqrt{\lambda / \max\{1, \lambda - \lambda_2\}},
\]
\[
D(W || Z) \leq c \log \frac{e\lambda}{\max\{1, \lambda - \lambda_2\}}
\]
with some absolute constant $c > 0$.

**Proof.** Setting $g(w) = \prod_{l=1}^{n} (q_l + p_l w)$ as before, we exploit the representation
\[
P\{W = k\} = \int_{|w|=r} w^{-k} g(w) d\mu_r(w) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(re^{i\theta}) e^{-ik\theta} d\theta,
\]
which is valid for all $r > 0$. Like in Section 6, we have an upper bound
\[
P\{W = k\} \leq R_k(r) I(r) \quad \text{with} \quad R_k(r) = r^{-k} \prod_{l=1}^{n} (q_l + p_l r)
\]
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 $q_j + p_j r \leq e^{p_j (r-1)}$,
\[
R_k(r) \leq r^{-k} \prod_{j=1}^{n} (q_j + p_j r) \leq e^{(r-1) - k \log r} = \left( \frac{e\lambda}{k} \right)^k e^{-\lambda}.
\]
Moreover, applying $(\frac{\lambda}{k})^k \leq e \sqrt{k \frac{\lambda}{k}}$ 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 the values $k = 0, 1, \ldots$

On the other hand, repeating the arguments from Section 6, for all $|\theta| \leq \pi$,
\[
\prod_{l=1}^{n} \frac{|q_l + p_l r e^{i\theta}|}{q_l + p_l r} = \prod_{l=1}^{n} \left( 1 - \frac{4q_l p_l r}{(q_l + p_l r)^2} \sin^2 \frac{\theta}{2} \right)^{1/2} \leq \exp \left\{ -2 \sin^2 \frac{\theta}{2} \sum_{l=1}^{n} \frac{q_l p_l r}{(q_l + p_l r)^2} \right\} \leq \exp \left\{ -\frac{2\theta^2}{\pi^2} \sum_{l=1}^{n} \frac{q_l p_l r}{(q_l + p_l r)^2} \right\}.
\]
Here
\[
\sum_{l=1}^{n} \frac{q_l p_l r}{(q_l + p_l r)^2} \geq \frac{1}{r} \sum_{l=1}^{n} q_l p_l = \frac{1}{r} (\lambda - \lambda_2) \quad \text{in case} \quad r \geq 1
\]
and
\[
\sum_{l=1}^{n} \frac{q_l p_r}{(q_l + p_r)^2} \geq r \sum_{l=1}^{n} q_l p_l = r (\lambda - \lambda_2) \quad \text{in case } r \leq 1.
\]
These right-hand sides have the form \( \psi(r) = \min\{r, 1/r\} (\lambda - \lambda_2) \), and we get
\[
I(r) \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp \left\{ -\frac{2}{\pi^2} \psi(r) \theta^2 \right\} d\theta = \frac{1}{4 \psi(r)^{1/2}} \int_{-2\sqrt{\psi(r)}}^{2\sqrt{\psi(r)}} e^{-\frac{1}{2} x^2} dx
\]
\[
\leq \frac{1}{4 \psi(r)^{1/2}} \min\{\sqrt{2\pi}, 4(\psi(r)^{1/2})\} \leq \min\{1, (\psi(r)^{-1/2})\}.
\]
First, we consider the region \( \frac{1}{2} \lambda \leq k \leq 3 \lambda \), in which case \( \frac{1}{2} \leq r \leq 3 \) and \( \psi(r) \geq \frac{1}{3} (\lambda - \lambda_2) \) and thus
\[
I(r) \leq \min\left\{1, \frac{\sqrt{3}}{\sqrt{\lambda - \lambda_2}}\right\} \leq \sqrt{\frac{3}{\max\{1, \lambda - \lambda_2\}}}.
\]
Hence
\[
P\{W = k\} \leq \sqrt{\frac{3}{\max\{1, \lambda - \lambda_2\}}} e^{\sqrt{k} f(k)}. \tag{9.3}
\]
As for the regions \( 0 \leq k \leq \frac{1}{2} \lambda \) and \( k > 3 \lambda \), we use the property \( |I(r)| \leq 1 \), which yields simpler upper bounds
\[
P\{W = k\} \leq \left(\frac{e\lambda}{k}\right)^{k} e^{-\lambda} \leq e\sqrt{k} f(k). \tag{9.4}
\]
Write
\[
\chi^2(W, Z) = \sum_{k=0}^{\infty} \frac{\mathbb{P}\{W = k\}^2}{\mathbb{P}\{Z = k\}} - 1 = S_1 + S_2 + S_3 - 1
\]
\[
= \left( \sum_{0 \leq k \leq \frac{1}{2} \lambda} + \sum_{\frac{1}{2} \lambda < k < 3 \lambda} + \sum_{k \geq 3 \lambda} \right) \frac{\mathbb{P}\{W = k\}^2}{\mathbb{P}\{Z = k\}} - 1.
\]
By (9.3), we see that
\[
S_2 \leq e \sqrt{\frac{3}{\max\{1, \lambda - \lambda_2\}}} \sum_{\frac{1}{2} \lambda < k < 3 \lambda} \sqrt{k} \mathbb{P}\{W = k\}
\]
\[
\leq e \sqrt{\frac{3\lambda}{\max\{1, \lambda - \lambda_2\}}} \sum_{\frac{1}{2} \lambda < k < 3 \lambda} \mathbb{P}\{W = k\},
\]
where the last sum does not exceed 1. Also, by (9.4), we obtain

$$S_1 \leq e^{-\lambda+1} \sum_{k<\frac{1}{2}\lambda} \sqrt{k} \left( \frac{e\lambda}{k} \right)^k$$

$$\leq \sqrt{\frac{1}{2}} \lambda e^{-\lambda+1} \sum_{k<\frac{1}{2}\lambda} \left( \frac{e\lambda}{k} \right)^k$$

$$\leq \sqrt{\frac{1}{2}} \lambda e^{-\lambda+1} \sum_{k<\frac{1}{2}\lambda} (2e)^{\lambda/2} \leq (\lambda/2)^{3/2} e^{-\lambda+1} (2e)^{\lambda/2} < e^{-c\lambda}$$

for some absolute constant $c > 0$ uniformly over all $\lambda \geq 1/8$. Here we used the property that the function $k \to (\frac{e\lambda}{k})^k$ is increasing for $k < \lambda$. Similarly,

$$S_3 \leq e^{-\lambda+1} \sum_{k>3\lambda} \sqrt{k} \left( \frac{e\lambda}{k} \right)^k < e^{-\lambda}.$$  

In both last cases, $e^{-c\lambda} \leq \frac{e^{\sqrt{\lambda}}}{\sqrt{\lambda}} \leq \frac{e^{\sqrt{\lambda}}}{\sqrt{\lambda}}$, so (9.1) holds.

Turning to the second assertion, write similarly

$$D(W, Z) = \sum_{k=0}^{\infty} \mathbb{P}\{W = k\} \log \frac{\mathbb{P}\{W = k\}}{\mathbb{P}\{Z = k\}} = S_1 + S_2 + S_3$$

$$= \left( \sum_{k=0}^{[\frac{1}{2}\lambda]} + \sum_{[\frac{1}{2}\lambda]+1}^{[3\lambda]} + \sum_{[3\lambda]+1}^{\infty} \right) \mathbb{P}\{W = k\} \log \frac{\mathbb{P}\{W = k\}}{\mathbb{P}\{Z = k\}}.$$

For the region $\frac{1}{2} \lambda \leq k \leq 3\lambda$, we can apply the bound (9.3) again, which gives

$$\mathbb{P}\{W = k\} \leq \sqrt{\frac{3}{\max\{1, \lambda - \lambda_2\}}} e^{\sqrt{k} f(k)} \leq 3e^{\sqrt{\frac{\lambda}{\max\{1, \lambda - \lambda_2\}}} f(k)},$$

and therefore

$$S_2 \leq \log(3e) + \frac{1}{2} \log \frac{\lambda}{\max\{1, \lambda - \lambda_2\}}.$$  

Using also (9.4), we obtain that

$$S_1 \leq e^{-\lambda+1} \sum_{k=0}^{[\frac{1}{2}\lambda]} \left( \frac{e\lambda}{k} \right)^k \log(e^{\sqrt{k}})$$

$$\leq e^{-\lambda+1} \log(e^{\sqrt{\lambda/2}}) \sum_{k=0}^{[\frac{1}{2}\lambda]} \left( \frac{e\lambda}{k} \right)^k \leq \frac{e\lambda}{2} \log(e^{\sqrt{\lambda/2}}) e^{-\frac{1}{2} \lambda \log \frac{2}{e}}$$

and

$$S_3 \leq e^{-\lambda+1} \sum_{k=([3\lambda]+1)}^{\infty} \left( \frac{e\lambda}{k} \right)^k \log(e^{\sqrt{k}}) \leq 100 e^{-\lambda+1}.$$

The last two upper bounds are majorized by a constant, hence, by the right-hand side of the inequality (9.2) uniformly over all $\lambda \geq 1/8$. □
10. Lower Bound on $\chi^2$ in the Degenerate Case

Here we complement Proposition 9.1 with a similar lower bound about $\chi^2$-distance.

Proposition 10.1. If $\lambda \geq 1/8$, then with some absolute constant $c \in (0, 1)$

$$1 + \chi^2(W, Z) \geq c \sqrt{\lambda/} \max\{1, \lambda - \lambda_2\}. $$

In particular, there exist absolute constants $\kappa' \in (0, 1)$ and $c' \in (0, 1)$ such that

$$\chi^2(W, Z) \geq c' \sqrt{\lambda/} \max\{1, \lambda - \lambda_2\}$$

as long as $\lambda_2 \geq \kappa' \lambda$.

To derive the second inequality from the first one, suppose that the first inequality holds true with some constants $c \in (0, 1)$ and $\kappa \in (0, 1)$. To obtain the second inequality with constants $c' = c/2$ and some $\kappa'$, it is sufficient to require that

$$Q \equiv \lambda/ \max\{1, \lambda - \lambda_2\} \geq \frac{2}{c'},$$

since then $cQ - 1 \geq \frac{c}{2} Q$. This condition is fulfilled, if $\lambda \geq \lambda_0 = \frac{1}{16}$, $\lambda_2 \geq (1 - \frac{c^2}{4}) \lambda$, and then we obtain the second inequality with $\kappa' = \max\{\kappa, 1 - \frac{c^2}{4}\}$. In the remaining case, where $\frac{1}{8} \leq \lambda \leq \lambda_0$ and $\lambda_2 \geq \kappa' \lambda$, one may use the same value $\kappa'$, by applying the first inequality of Lemma 3.1. It implies that

$$\chi^2(W, Z) \geq \frac{(\mathbb{P}\{Z = 0\} - \mathbb{P}\{W = 0\})^2}{\mathbb{P}\{Z = 0\}} \geq (1 - e^{-\frac{1}{2} \lambda_2})^2 e^{-\lambda} \geq (1 - e^{-\frac{1}{2} \kappa' \lambda_0})^2 e^{-\lambda_0} \geq (1 - e^{-\frac{1}{2} \kappa' \lambda_0})^2 e^{-\lambda_0},$$

while $Q \leq \sqrt{\lambda_0}$. Then, the second inequality of Proposition 10.1 will hold true with $c' = \frac{1}{\sqrt{\lambda_0}} (1 - e^{-2\kappa}) e^{-\lambda_0}$, which should be decreased to $\frac{1}{2}$ if necessary, so as to unite both cases. Thus, the second assertion of Proposition 10.1 follows from the first one.

First we prove the proposition (1st assertion) assuming that $\lambda - \lambda_2$ is sufficiently large. As in Section 9, for any fixed $r > 0$, we apply the Cauchy theorem and write

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

with integration over the uniform distribution $\mu_r$ 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

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

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

$$\mathbb{P}\{W = k\} = R_k(r) I_k(r) = R_k(r) (I_{k1}(r) + I_{k2}(r)).$$
Relative entropy and $\chi^2$ divergence from the Poisson law

where

$$I_{k1}(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{i=1}^{n} \frac{|q_l + pr e^{i\theta}|}{q_l + pr} \exp \left\{ -ik\theta + i \sum_{i=1}^{n} \text{Im}(\log(q_l + pr e^{i\theta})) \right\} d\theta,$$

$$I_{k2}(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{i=1}^{n} \frac{|q_l + pr e^{i\theta}|}{q_l + pr} \exp \left\{ -ik\theta + i \sum_{i=1}^{n} \text{Im}(\log(q_l + pr e^{i\theta})) \right\} d\theta.$$

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

$$F(r) \equiv \sum_{l=1}^{n} \frac{pr}{q_l + pr} = k. \quad (10.1)$$

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 $p_l$ 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 10.2.** For any $k = 0, \ldots, n - 1$, the solution $r = r(k)$ to the equation (10.1) satisfies

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

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

$$r = 1 + \left(\frac{6}{\pi}\right)^2 b_1 \frac{k - \lambda}{\lambda - \lambda_2},$$

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

**Proof.** We have

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

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

$$F^{-1}(s) \geq F^{-1}(\lambda) + (F^{-1})'(\lambda) (s - \lambda)$$

$$= F^{-1}(\lambda) + \frac{1}{F'(F^{-1}(\lambda))} (s - \lambda) = 1 + \frac{1}{\lambda - \lambda_2} (s - \lambda).$$

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

Now, since $q_l + pr \leq 1$ for $r \leq 1$, we conclude that $F'(r) \geq \sum_{l=1}^{n} p_lq_l = \lambda - \lambda_2$ 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 $F(r) - F(1) \geq (1 - \frac{1}{r^{2}}) (\lambda - \lambda_{2})$. Hence, again by the assumption,

$$
\frac{1}{6} (\lambda - \lambda_{2}) \geq k - \lambda = F(r(k)) - F(1) \geq \left(1 - \frac{1}{r(k)^{2}}\right) (\lambda - \lambda_{2}),
$$

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

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

In addition,

$$
-F''(r) = 2 \sum_{l=1}^{n} \frac{p_{l}^{2}q_{l}}{(q_{l} + p_{l}r)^{3}} \leq 2 \cdot \left(\frac{6}{5}\right)^{3} \sum_{l=1}^{n} p_{l}^{2}q_{l} = 2 \cdot \left(\frac{6}{5}\right)^{3} (\lambda_{2} - \lambda_{3}).
$$

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

$$
F^{-1}(s) = 1 + \frac{1}{F'(s_{1})} (s - \lambda) = 1 + \frac{1}{F(1)} (s - \lambda) - \frac{1}{2 F'(1)^{2}} F''(s_{2}) (s - \lambda)^{2},
$$

where the points $s_{1}$ and $s_{2}$ lie between $\lambda$ and $s$. Putting $r = F^{-1}(s)$ and $r_{i} = F^{-1}(s_{i})$, the above is simplified as

$$
r = 1 + \frac{1}{F'(r_{1})} (s - \lambda)
= 1 + \frac{1}{\lambda - \lambda_{2}} (s - \lambda) - \frac{1}{2 F'(r_{2})^{2}} F''(r_{2}) (s - \lambda)^{2},
$$

where $r_{1}$ and $r_{2}$ lie between 1 and $r$. It remains to apply these equalities with $s = k$, that is, $r = r(k)$, and note that $\frac{1}{F'(r_{1})} \leq \left(\frac{6}{5}\right)^{2} \frac{1}{\lambda - \lambda_{2}}$, while

$$
\frac{1}{2 F'(r_{1})^{2}} |F''(r_{1})| \leq \frac{1}{2} \left(\frac{6}{5}\right)^{3} (\lambda_{2} - \lambda_{3}) = \left(\frac{6}{5}\right)^{9} \frac{\lambda_{2} - \lambda_{3}}{(\lambda - \lambda_{2})^{2}}.
$$

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

\[\textbf{Lemma 10.3.}\] Let $r = r(k)$ be the solution of (9.1) for $0 \leq \lambda - k \leq \frac{1}{5} (\lambda - \lambda_{2})$. Then

$$
R_{k}(r) = r^{-k} \prod_{l=1}^{n} (q_{l} + p_{l}r) \geq \exp \left\{ -4 \frac{(\lambda - k)^{2}}{(\lambda - \lambda_{2})} \right\}.
$$

\[\textbf{Proof.}\] The function

$$
\psi_{k}(r) = \log R_{k}(r) = \sum_{l=1}^{n} \log (q_{l} + p_{l}r) - k \log r, \quad r > 0,
$$
is vanishing at \( r = 1 \) and has derivative

\[
\psi'_k(r) = \sum_{l=1}^{n} \frac{p_l}{q_l + pr} - \frac{k}{r} = \frac{F(r) - k}{r} = \frac{F(r) - F(r(k))}{r}.
\]

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

\[
\psi'_k(r) \leq \frac{F'(r(k))}{r} (r - r(k)) \leq \frac{F'(r(k))}{r(k)} (1 - r(k)),
\]

which implies

\[
\psi_k(r(k)) = \psi_k(r(k)) - \psi_k(1) \geq - \frac{F'(r(k))}{r(k)} (1 - r(k))^2.
\]

By Lemma 10.2, \( \frac{5}{6} \leq r(k) \leq \frac{6}{5} \) and \( 1 - r(k) \leq \frac{(\frac{6}{5})^2}{\lambda - \lambda_2} \). Moreover, as was shown in the proof, \( F'(r(k)) \leq \frac{(\frac{6}{5})^2}{\lambda - \lambda_2} \). Hence

\[
\frac{F'(r(k))}{r(k)} (1 - r(k))^2 \leq \frac{(\frac{6}{5})^2}{5/6} \frac{(\frac{6}{5})^2}{\lambda - \lambda_2} = \frac{6}{5} \frac{(\lambda - \lambda_2)^2}{\lambda - \lambda_2}.
\]

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

**Lemma 10.4.** Let \( \lambda - \lambda_2 \geq 100 \). Then, for \( 0 \leq \lambda - k \leq \frac{1}{6} \lambda - \lambda_2 \),

\[
I_k(r(k)) \geq \frac{1}{10 \sqrt{\lambda - \lambda_2}}.
\]

**Proof.** By Lemma 10.2, \( 1 \geq r(k) \geq \frac{5}{6} \). Note that, for \( r > 0 \) and \( -\pi \leq \theta \leq \pi \),

\[
\prod_{l=1}^{n} \left| \frac{q_l + pr e^{i\theta}}{q_l + pr} \right| = \prod_{l=1}^{n} \left( 1 - \frac{4q_l p_r e^{i\theta}}{(q_l + pr)^2} \sin^2 \frac{\theta}{2} \right)^{1/2} \leq \exp \left\{ -2 \sum_{l=1}^{n} \frac{q_l p_r}{(q_l + pr)^2} \sin^2 \frac{\theta}{2} \right\}.
\]

For \( \frac{5}{6} \leq r \leq 1 \), necessarily \( q_l + pr \leq 1 \) and

\[
\sum_{l=1}^{n} \frac{q_l p_r}{(q_l + pr)^2} \geq \sum_{l=1}^{n} q_l p_r = (\lambda - \lambda_2) r \geq \frac{5}{6} (\lambda - \lambda_2).
\]

Hence

\[
I_{k2}(r) \leq \frac{1}{2 \pi} \int_{\frac{\pi}{2} \leq |\theta| \leq \pi} \prod_{l=1}^{n} \frac{|q_l + pr e^{i\theta}|}{q_l + pr} d\theta \leq \frac{1}{2 \pi} \int_{\frac{\pi}{2} \leq |\theta| \leq \pi} \exp \left\{ -\frac{5}{3} (\lambda - \lambda_2) \sin^2 \frac{\theta}{2} \right\} d\theta \leq \frac{1}{2} e^{-\frac{5}{3} (\lambda - \lambda_2)}.
\]

Let us now estimate \( I_{k1} \). Using \( 4q_l p_r \leq (q_l + pr)^2 \) (since \( (q_l - pr)^2 \geq 0 \)), we have, for \( |\theta| \leq \pi/2 \),

\[
\frac{4q_l p_r}{(q_l + pr)^2} \sin^2 \frac{\theta}{2} \leq \frac{1}{2}, \quad l = 1, \ldots, n.
\]
In the region $0 \leq \varepsilon \leq \varepsilon_0 < 1$, there is a lower bound $1 - \varepsilon \geq e^{-c\varepsilon}$ 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 $|\theta| \leq \frac{\pi}{2}$,

$$
\prod_{l=1}^{n} \frac{|q_l + pr e^{i\theta}|}{q_l + pr} \geq \exp \left\{ - \log 2 \sum_{l=1}^{n} \frac{4q_lr}{(q_l + pr)^2} \sin^2(\theta/2) \right\}.
$$

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

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

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

$$
\max_{\frac{5}{6} \leq r \leq 1} w_l(r) = w_l(r_l) = \frac{1}{4p_lr} \leq \frac{1}{4 \cdot \frac{6}{11} \cdot \frac{7}{11}} = \frac{121}{120}.
$$

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

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

and thus

$$
\frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \prod_{l=1}^{n} \frac{|q_l + pr e^{i\theta}|}{q_l + pr} \ d\theta \geq \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \exp \left\{ - \frac{6}{5} \log 2 (\lambda - \lambda_2) \theta^2 \right\} \ d\theta
$$

$$
= \frac{1}{2\pi} \frac{\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)}{\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)} \int_{-\frac{\pi}{2\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)}}^{\frac{\pi}{2\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)}} \exp \left\{ - \frac{1}{2} x^2 \right\} \ dx
$$

$$
= 0.3093 \frac{1}{\sqrt{\lambda - \lambda_2}}.
$$

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

$$
\frac{1}{2\pi} \frac{\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)}{\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)} \int_{-\frac{\pi}{2\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)}}^{\frac{\pi}{2\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)}} e^{-\frac{1}{2} x^2} \ dx \geq \frac{1}{2\pi} \frac{\sqrt{\frac{6}{5} \log 4}}{\sqrt{\frac{6}{5} \log 4}} \int_{-\frac{5\pi}{2\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)}}^{\frac{5\pi}{2\sqrt{\frac{6}{5} \log 4} (\lambda - \lambda_2)}} e^{-\frac{1}{2} x^2} \ dx
$$

$$
= \frac{1}{2\pi} \frac{\sqrt{\frac{6}{5} \log 4}}{\sqrt{\frac{6}{5} \log 4}} \ P\left\{ |\xi| \leq 5\pi \sqrt{\frac{6}{5} \log 4} \right\} > 0.3093,
$$
where $\xi \sim N(0, 1)$. In addition (recalling one of the upper bounds when bounding
the integral $I_{k2}$ from above), and using $\sin(\theta/2) \geq \sqrt{2}/\pi \theta$ for $0 \leq \theta \leq \pi/2$, we get that
\[
\frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \prod_{l=1}^{n} \left| q_l + pr^2 e^{i\theta} \right| \theta^6 d\theta \leq \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \exp \left\{ -\frac{5}{3} (\lambda - \lambda_2) \sin^2 \frac{\theta}{2} \right\} \theta^6 d\theta
\]
\[
\leq \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \exp \left\{ -\frac{10}{3\pi^2} (\lambda - \lambda_2) \theta^2 \right\} \theta^6 d\theta
\]
\[
\leq \frac{1}{\pi} \left( \frac{20}{3\pi^2} (\lambda - \lambda_2) \right)^{-7/2} \int_{-\infty}^{\infty} e^{-x^2/2} dx
\]
\[
= \pi^{13/2} \left( \frac{3}{20} \right)^{7/2} \frac{1}{(\lambda - \lambda_2)^{7/2}} < \frac{48}{(\lambda - \lambda_2)^{7/2}}.
\]

The assumption (10.1) may be rewritten as
\[
\text{Im} \left( \sum_{l=1}^{n} \log(q_l + pr e^{i\theta}) \right) \bigg|_{\theta=0} = \left( \sum_{l=1}^{n} \text{Im} \left( \log(q_l + pr e^{i\theta}) \right) \right) \bigg|_{\theta=0} = k.
\]

Note that the functions $\text{Im} \left( \log(q_l + pr e^{i\theta}) \right)$ are odd, so their 2nd derivatives are
vanishing at zero. We now apply the Taylor formula up to the cubic term to the
function
\[A_k(r, \theta) = -k\theta + \text{Im} \sum_{l=1}^{n} \log(q_l + pr e^{i\theta})\]
on the interval $\theta \in [-\pi/2, \pi/2]$ to get that
\[A_k(r, \theta) = \frac{1}{6} \left( \text{Im} \sum_{l=1}^{n} \log(q_l + pr e^{i\theta}) \right)^{'''} \bigg|_{\theta=0} \theta^3\]
with some $\theta_0 \in [-\pi/2, \pi/2]$. To perform the differentiation, consider a function of
the form
\[h(v) = \log(q + pr e^{iv}), \quad p, q, r > 0.\]

We have
\[h'(v) = \frac{pr e^{iv}}{q + pr e^{iv}} = i \left( 1 - \frac{q}{q + pr e^{iv}} \right) = i - iq (q + pr e^{iv})^{-1},\]
\[h''(v) = -pq r e^{iv} (q + pr e^{iv})^{-2},\]
\[h'''(v) = -pq r \left( ie^{iv} (q + pr e^{iv})^{-2} - 2i pr e^{2iv} (q + pr e^{iv})^{-3} \right).\]

Therefore,
\[-\left( \text{Im} \sum_{l=1}^{n} \log(q_l + pr e^{i\theta}) \right)''' = \text{Im} \left( i \sum_{l=1}^{n} \frac{pq r e^{2i\theta}}{(q_l + pr e^{i\theta})^2} \right) - 2 \text{Im} \left( i \sum_{l=1}^{n} \frac{pq r^2 e^{2i\theta}}{(q_l + pr e^{i\theta})^3} \right),\]

implying that
\[\left| \text{Im} \sum_{l=1}^{n} \log(q_l + pr e^{i\theta}) \right|''' \leq \sum_{l=1}^{n} \frac{pq r}{(q_l + pr e^{i\theta})^2} + 2 \sum_{l=1}^{n} \frac{pq r^2}{(q_l + pr e^{i\theta})^3}.\]
But, for \( \frac{5}{6} \leq r \leq 1 \) and \(|\theta| \leq \frac{\pi}{2}\),

\[
|q_l + p_l r e^{i\theta}|^2 = (q_l + p_l r)^2 (1 - \frac{4q_l p_l r}{(q_l + p_l r)^2} \sin^2 \frac{\theta}{2}) \\
\geq (q_l + p_l r)^2 - 2q_l p_l r = q_l^2 + p_l^2 r^2.
\]

Hence

\[
\frac{r}{|q_l + p_l r e^{i\theta}|^2} \leq \frac{r}{q_l^2 + p_l^2 r^2} = u_l(r) \leq \frac{121}{60}.
\]

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

\[
|A_l(1)| = \frac{1}{2p_l q_l} \leq \frac{1}{2 \cdot \frac{6}{11} \cdot \frac{5}{6}} = \frac{121}{60}.
\]

From this,

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

so that

\[
\left| \left( \text{Im} \sum_{l=1}^{n} \log(q_l + p_l r e^{i\theta}) \right) \right|^2 \leq \frac{121}{60} \sum_{l=1}^{n} q_l^2 p_l^2 + 2 \left( \frac{121}{60} \right)^{3/2} \sum_{l=1}^{n} q_l^2 p_l^2 \leq c_0 (\lambda - \lambda_2)
\]

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

\[
|A_k(r, \theta)| \leq \frac{c_0}{6} (\lambda - \lambda_2) |\theta|^3, \quad \frac{5}{6} \leq r \leq 1, \quad |\theta| \leq \frac{\pi}{2}.
\]

Now, as we mentioned before, the function \( A_k \) is odd in \( \theta \), so that

\[
I_{k1}(r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{l=1}^{n} \frac{|q_l + p_l r e^{i\theta}|}{q_l + p_l r} \cos(A_k(r, \theta)) \, d\theta
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} \prod_{l=1}^{n} \frac{|q_l + p_l r e^{i\theta}|}{q_l + p_l r} \, d\theta - \frac{1}{\pi} \int_{-\pi}^{\pi} \prod_{l=1}^{n} \frac{|q_l + p_l r e^{i\theta}|}{q_l + p_l r} \sin^2(A(r, \theta)/2) \, d\theta.
\]

Hence, using

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

from the previous estimates we may deduce the lower bound

\[
I_{k1}(r) \geq 0.3093 \frac{1}{\sqrt{\lambda - \lambda_2}} - \frac{c_0^2}{144} (\lambda - \lambda_2)^2 \frac{48}{(\lambda - \lambda_2)^{7/2}}
\]

\[
= 0.3093 \frac{1}{\sqrt{\lambda - \lambda_2}} - \frac{c_0^2}{3} \frac{1}{(\lambda - \lambda_2)^{3/2}}
\]

\[
\geq \frac{1}{\sqrt{\lambda - \lambda_2}} (0.3093 - \frac{20}{\lambda - \lambda_2}) \geq 0.1093 \frac{1}{\sqrt{\lambda - \lambda_2}},
\]
where on the last step we assume that $\lambda - \lambda_2 \geq 100$. Together with the upper bound on $I_k$, we arrive at the lower bound

$$I_k(r) \geq 0.1093 \frac{1}{\sqrt{\lambda - \lambda_2}} - \frac{1}{2} e^{-\frac{5}{6} (\lambda - \lambda_2)}$$

$$\geq (0.1093 - 5 e^{-\frac{5}{6}}) \frac{1}{\sqrt{\lambda - \lambda_2}} > \frac{0.1}{\sqrt{\lambda - \lambda_2}}.$$  

Thus, Lemma 10.4 is proved.

**Proof of Proposition 10.1.** We conclude from Lemmas 10.3 and 10.4 that

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

(10.2)

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

On the other hand, according to the bound (2.6) of Lemma 2.3, $\mathbb{P}\{Z = k\} \leq \sqrt{\frac{2\pi k}{2\pi}}$. Since $k \geq \lambda - \frac{1}{6} (\lambda - \lambda_2) \geq \frac{5}{6} \lambda$, we have

$$\mathbb{P}\{Z = k\} \leq \sqrt{\frac{6/5}{2\pi \lambda}} < \frac{1}{2\sqrt{\lambda}}.$$

As a consequence,

$$1 + \chi^2(W, Z) \geq \sum_{0 \leq \lambda - k \leq \frac{1}{6} \sqrt{\lambda - \lambda_2}} \frac{\mathbb{P}\{W = k\}^2}{\mathbb{P}\{Z = k\}}$$

$$\geq \frac{\sqrt{\lambda}}{8 (\lambda - \lambda_2)} \sum_{0 \leq \lambda - k \leq \frac{1}{6} \sqrt{\lambda - \lambda_2}} e^{-8 \frac{(\lambda - k)^2}{\lambda - \lambda_2}} \geq c \frac{\sqrt{\lambda}}{\lambda - \lambda_2}$$

with an absolute constant $c > 0$. In order to clarify the last inequality, note that the condition $\lambda - \lambda_2 \geq 100$ implies that $\lambda \geq 100$. 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 of order $\frac{1}{6} \sqrt{\lambda - \lambda_2}$. Therefore,

$$\sum_{0 \leq \lambda - k \leq \frac{1}{6} \sqrt{\lambda - \lambda_2}} e^{-8 \frac{(\lambda - k)^2}{\lambda - \lambda_2}} \sim \int_0^{\frac{1}{6} \sqrt{\lambda - \lambda_2}} e^{-8 x^2 / (\lambda - \lambda_2)} dx$$

$$= \frac{1}{4} \sqrt{\lambda - \lambda_2} \int_0^{\frac{1}{2}} e^{-y^2 / 2} dy \sim \sqrt{\lambda - \lambda_2},$$

where $\sim$ means the equivalence within positive absolute factors.

In order to treat the region $\lambda - \lambda_2 \leq 100$, we apply Proposition 2.2. Let $W_1 = W$ and $W_2 = Y_1 + \cdots + Y_m$, where $Y_1, \ldots, Y_m$ 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 $\lambda = \lambda + m/2$ and $\lambda_2 = \lambda_2 + m/4$ satisfy the condition $\lambda - \lambda_2 \geq 100$. 


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

$$c \sqrt{\frac{\lambda}{\lambda - \lambda_2}} \leq \chi^2(W_1 + W_2, Z_1 + Z_2) + 1 \leq (\chi^2(W_1, Z_1) + 1)(\chi^2(W_2, Z_2) + 1)$$

with some absolute constant $c > 0$. Here $\chi^2(W_2, Z_2)$ is just a numerical value, while $\frac{\lambda}{\lambda - \lambda_2}$ is equivalent to $\lambda / \max\{1, \lambda - \lambda_2\}$ as long as $\lambda$ is bounded away from zero, for example, if $\lambda \geq 1/8$. Hence, Proposition 10.1 holds in the case $\lambda - \lambda_2 \leq 100$ as well.

11. Lower Bound for $D$ in the Degenerate Case

An analogue of Proposition 10.1 is the following statement about the relative entropy.

**Proposition 11.1.** There exist absolute constants $0 < \kappa < 1$ and $\lambda_0 > 0$, such that, for $\lambda_2 \geq \kappa \lambda$ and $\lambda \geq \lambda_0$,

$$D(W||Z) \geq c \log \frac{e^\lambda}{\max\{1, \lambda - \lambda_2\}}. \quad (11.1)$$

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

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

that have been obtained for $0 \leq \lambda - k \leq \frac{1}{6} (\lambda - \lambda_2)$ under the assumption $\lambda - \lambda_2 \geq 100$. This is fulfilled if $0 \leq \lambda - k \leq \frac{5}{3} \sqrt{\lambda - \lambda_2}$ and $\lambda - \lambda_2 \geq 100$, and if additionally $\lambda_2 \geq \kappa \lambda$ with $0 < \kappa < 1$, then

$$w_k \geq \frac{1}{10 \sqrt{\lambda - \lambda_2}} e^{-100/3} \geq \frac{1}{10 \sqrt{(1 - \kappa) \lambda}} e^{-100/3}.$$  

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

$$\frac{1}{10 \sqrt{1 - \kappa}} e^{-100/3} \geq \frac{1}{\sqrt{5\pi/3}},$$

i.e., $1 - \kappa \leq \frac{60}{\pi} e^{-200/3}$. Let $\kappa = 1 - \frac{60}{\pi} e^{-200/3}$ for definiteness.

We have, moreover,

$$\log \frac{w_k}{v_k} \geq \frac{1}{2} \log \frac{\lambda}{\lambda - \lambda_2} + \log \left(\frac{\sqrt{5\pi/3}}{10} e^{-100/3}\right) \geq \frac{1}{2} \log \frac{\lambda}{\lambda - \lambda_2} - 32.$$
Now, applying the general inequality (2.1) of Proposition 2.1, we get
\[
D(W||Z) \geq \sum_{w_k \geq v} w_k \log \frac{w_k}{v_k} - 1
\]
\[
\geq \sum_{0 \leq \lambda - k \leq \frac{2}{3} \sqrt{\lambda - \lambda_2}} w_k \log \frac{w_k}{v_k} - 1
\]
\[
\geq \sum_{0 \leq \lambda - k \leq \frac{2}{3} \sqrt{\lambda - \lambda_2}} w_k \left( \frac{1}{2} \log \frac{\lambda}{\lambda - \lambda_2} - 32 \right) - 1
\]
\[
\geq \frac{1}{2} \log \frac{\lambda}{\lambda - \lambda_2} \sum_{0 \leq \lambda - k \leq \frac{2}{3} \sqrt{\lambda - \lambda_2}} \frac{1}{10 \sqrt{\lambda - \lambda_2}} e^{-4 \frac{(\lambda - k)^2}{\lambda - \lambda_2}} - 33.
\]
Note that, if \( \lambda - \lambda_2 \geq 100 \), the \( x \)-interval \( \Delta : 0 \leq \lambda - x \leq \frac{2}{3} \sqrt{\lambda - \lambda_2} \) has length \( |\Delta| \) at least \( 50/3 \), so, the total number of such integer points in this interval is at least
\[
|\Delta| \geq \frac{3}{5} |\Delta| \geq \sqrt{\lambda - \lambda_2}.
\]
Hence, the last sum can be bounded from below by
\[
\frac{1}{10 \sqrt{\lambda - \lambda_2}} e^{-100/3} \sum_{0 \leq \lambda - k \leq \frac{2}{3} \sqrt{\lambda - \lambda_2}} 1 \geq \frac{1}{10} e^{-100/3}.
\]
Thus,
\[
D(W||Z) \geq \frac{1}{10} e^{-100/3} \log \frac{\lambda}{\lambda - \lambda_2} - 33 \geq c \log \frac{\lambda}{\lambda - \lambda_2}
\]
with some constant \( c > 0 \).

The proposition is thus proved under the conditions \( \lambda - \lambda_2 \geq 100 \) and \( \lambda_2 \geq \kappa \lambda \).

It remains to eliminate the first condition, assuming that \( \lambda - \lambda_2 < 100 \) and again that \( \lambda_2 \geq \kappa \lambda \). To this aim, we appeal to Proposition 2.2 again like on the last step of the proof of the analogous Proposition 10.1. Namely, using the same notations and assumptions, from the inequality (2.3) and using the previous step, we obtain that
\[
c \log \frac{\lambda}{\max\{1, \lambda - \lambda_2\}} \leq D(W_1 + W_2||Z_1 + Z_2) \leq D(W_1||Z_1) + D(W_2||Z_2),
\]
where \( W_1 = W \) and \( Z_1 = Z \). It holds with some absolute constant \( c > 0 \), as long as the second condition is fulfilled: \( \lambda_2 \geq \kappa \lambda \), i.e.,
\[
\lambda_2 + m/4 \geq \kappa (\lambda + m/2).
\]
Since \( \lambda - \lambda_2 < 100 \), the latter would follow from
\[
\lambda - 100 + m/4 \geq \kappa (\lambda + m/2)
\]
which is solved as \( \lambda \geq 50 \frac{\kappa}{1 - \kappa} \).

Since \( C = D(W_2||Z_2) \) is just a numerical value, we conclude that
\[
D(W||Z) \geq c \log \frac{\lambda}{\max\{1, \lambda - \lambda_2\}} - C \geq c' \log \frac{\lambda}{\max\{1, \lambda - \lambda_2\}}.
\]
Here the last inequality holds true with a suitable constant \( c' \in (0, c) \), since \( \lambda - \lambda_2 < 100 \) and since one may choose a proper value of \( \lambda_0 \). \( \square \)
12. Summarizing Remarks. Proof of Theorem 1.2

Let us summarize. The obtained bounds for different regions are united in Theorem 1.2 as the two-sided bounds

\[ c_1 \left( \frac{\lambda_2}{\lambda} \right)^2 \frac{2 + \lambda}{\max\{1, \lambda - \lambda_2\}} \leq D(W \| Z) \leq c_2 \left( \frac{\lambda_2}{\lambda} \right)^2 \frac{2 + \lambda}{\max\{1, \lambda - \lambda_2\}}, \]  
\[ (12.1) \]

\[ c_1 \left( \frac{\lambda_2}{\lambda} \right)^2 \sqrt{\frac{2 + \lambda}{\max\{1, \lambda - \lambda_2\}}} \leq \chi^2(W, Z) \leq c_2 \left( \frac{\lambda_2}{\lambda} \right)^2 \sqrt{\frac{2 + \lambda}{\max\{1, \lambda - \lambda_2\}}}, \]  
\[ (12.2) \]

valid up to some absolute positive constants \( c_1 \) and \( c_2 \).

If the parameter \( \lambda \) is bounded from above by an absolute constant, for example, if \( \lambda \leq 1/8 \), these bounds simplify to

\[ c_1 \left( \frac{\lambda_2}{\lambda} \right)^2 \leq D(W \| Z) \leq \chi^2(W, Z) \leq c_2 \left( \frac{\lambda_2}{\lambda} \right)^2. \]  
\[ (12.3) \]

For these values, this was proved in Section 4, cf. Proposition 4.3.

In the case where \( \lambda \geq 1/8 \) and \( \lambda_2 \leq \kappa \lambda \) with an absolute constant \( \kappa \in (0, 1) \), the inequalities (12.1)-(12.2) are simplified to (12.3) again. The lower bound in this case follows Theorem 1.1 giving a lower bound for the total variation distance.

In the case where \( \lambda \geq 1/8 \) and \( \lambda_2 \geq \kappa \lambda \), (12.2) is simplified to

\[ c_1 \sqrt{\frac{\lambda}{\max\{1, \lambda - \lambda_2\}}} \leq \chi^2(W, Z) \leq c_2 \sqrt{\frac{\lambda}{\max\{1, \lambda - \lambda_2\}}}. \]

It was proved in Sections 9-10, cf. Propositions 9.1 and Proposition 10.1. In this region, \( \chi^2(W, Z) \) is necessarily bounded away from zero.

In the same region, (12.1) is simplified to

\[ c_1 \log \frac{2 + \lambda}{\max\{1, \lambda - \lambda_2\}} \leq D(W \| Z) \leq c_2 \log \frac{2 + \lambda}{\max\{1, \lambda - \lambda_2\}}. \]  
\[ (12.4) \]

In this case, the upper bound was derived in Proposition 9.1, while the lower bound is given in Proposition 11.1, cf. (11.1), – but only for larger values \( \lambda \geq \lambda_0 \). For the remaining region \( \frac{1}{8} < \lambda < \lambda_0 \), the inequality (12.4) is further simplified to

\[ c_1 \leq D(W \| Z) \leq c_2, \]

where we only need to recover the lower bound. But again, since the relative entropy is stronger than the total variation distance, the result follows from Theorem 1.1 via the Pinsker inequality. Theorem 1.2 is thus fully proved.

13. Difference of Entropies

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

\[ H_2(Z) = \left( \mathbb{E} (\log v(Z))^2 \right)^{1/2} = \left( \sum_k v_k (\log v_k)^2 \right)^{1/2}, \]

where \( Z \) is an integer-valued random variable with probability function \( v(k) = v_k = P\{Z = k\}, k \in \mathbb{Z} \). 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 13.1.** For all integer-valued random variables $W$ and $Z$ with finite entropies, we have

$$H(W||Z) \leq \chi^2(W, Z) + H_2(Z) \sqrt{\chi^2(W, Z)}.$$  \hfill (13.1)

**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 $k \in \mathbb{Z}$, $v_k = 0$ implies $w_k = 0$, where $w_k = \mathbb{P}\{W = k\}$. Define $t_k = w_k/v_k$ in case that $v_k > 0$. Recalling the definition (1.10), 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

$$H(W||Z) \leq \sum_{v_k > 0} (t_k - 1) v_k + \sum_{v_k > 0} (t_k - 1)^2 v_k + \sum_{v_k > 0} (t_k - 1) v_k \log v_k.$$

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

$$\sum_{k} \frac{(w_k - v_k)^2}{v_k} \sum_{k} v_k (\log v_k)^2 \leq \chi^2(W, Z) H_2^2(Z).$$

□

In view of (13.1), we also need:

**Proposition 13.2.** If $Z \sim P_\lambda$, then with some absolute constant $c > 0$

$$H_2(Z) \leq \begin{cases} c \log(4\lambda), & \text{if } \lambda \geq \frac{1}{2}, \\ c\sqrt{\lambda} \log(1/\lambda), & \text{if } \lambda \leq \frac{1}{2}. \end{cases}$$

**Proof.** First note that the first probabilities of $v_k = \mathbb{P}\{Z = k\}$ satisfy $v_0 (\log v_0)^2 = \lambda^2 e^{-\lambda}$ which is bounded for $\lambda \geq 1$ and is equivalent to $\lambda^2$ as $\lambda \to 0$. Also, $v_1 (\log v_1)^2 = \lambda e^{-\lambda} (\lambda + \log(1/\lambda))^2$. This shows that the above upper bound for $\lambda \leq \frac{1}{2}$ can be reversed. For such values, given $k \geq 1$, from

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

we get

$$\sum_{k \geq 1} v_k (\log v_k)^2 \leq \mathbb{E} \sum_{k \geq 1} \frac{1}{v_k} \log \frac{1}{v_k} \leq \mathbb{E} \sum_{k \geq 1} \frac{1}{v_k} \log \frac{1}{\lambda} \leq 14 \lambda \log^2 \frac{e}{\lambda},$$

thus proving the upper bound.
Now, assuming that $\lambda \geq \frac{1}{2}$, let us apply the lower bound (2.7) from Lemma 2.3, which for $1 \leq k \leq 2\lambda$ gives

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

and

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

Using the concavity of the function $\log^2 x$ in $x \geq e$ and applying Jensen’s inequality, we therefore obtain that

$$\sum_{1 \leq k \leq 2\lambda} v_k (\log v_k)^2 \leq 2 \log^2(e(Z + 1)) + \frac{6(\lambda + 2)}{\lambda} \leq c \log^2(4\lambda).$$

In addition, using $\log \frac{1}{v_k} < k^2$ for $k > 2\lambda \geq 1$, we also have

$$\sum_{k > 2\lambda} v_k (\log v_k)^2 \leq \mathbb{E} Z^4 1_{\{Z > 2\lambda\}} \leq c$$

with some absolute constant $c > 0$. These bounds give $H_2(Z) \leq c \log(4\lambda)$. Applying the upper bound (2.7) from Lemma 2.3, we also see that this upper bound on $H_2$ can also be reversed.

\[\square\]

**Remark 13.3.** With the same arguments based on Lemma 2.4, it follows that

$$H(Z) \leq \begin{cases} c \log(4\lambda), & \text{if } \lambda \geq \frac{1}{2}, \\ c \log(1/\lambda), & \text{if } \lambda \leq \frac{1}{2}, \end{cases}$$

which can also be reversed modulo an absolute factor $c > 0$. Hence, $H_2(Z) \sim H(Z)$ as long as $\lambda$ 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

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

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 $\lambda_2 \leq \lambda$).

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

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

By Proposition 13.1, this gives

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

and it remains to note that $1 + H_2(Z) \leq C \log(2 + \lambda)$, according to Proposition 13.2. \[\square\]
Acknowledgement. The authors would like to thank Igal Sason for drawing our attention to additional references related to the Poisson approximation in informational distances.

References


[V] 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.