The Density Ratio of Generalized Binomial versus Poisson Distributions

Lutz Dümbgen (University of Bern)∗
and
Jon A. Wellner (University of Washington, Seattle)†

October 9, 2019

Abstract

Let $b(x)$ be the probability that a sum of independent Bernoulli random variables with parameters $p_1, p_2, p_3, \ldots \in [0, 1]$ equals $x$, where $\lambda := p_1 + p_2 + p_3 + \cdots$ is finite. We prove two inequalities for the maximal ratio $b(x)/\pi_\lambda(x)$, where $\pi_\lambda$ is the weight function of the Poisson distribution with parameter $\lambda$.

Key words: Poisson approximation, relative errors, total variation distance.

1 Introduction

We consider independent Bernoulli random variables $Z_1, Z_2, Z_3, \ldots \in \{0, 1\}$ with parameters $\mathbb{P}(Z_i = 1) = \mathbb{E}(Z_i) = p_i \in [0, 1)$ and their random sum $X = \sum_{i \geq 1} Z_i$. By the first and second Borel–Cantelli lemmas, $X$ is almost surely finite if and only if the sequence $p = (p_i)_{i \geq 1}$ satisfies

$$\lambda := \sum_{k=1}^{\infty} p_k < \infty,$$

and we exclude the trivial case $\lambda = 0$. Under this assumption, the distribution $Q$ of $X$ is given by

$$\mathbb{P}(X = x) =: b(x) = \sum_{J : \# J = x} \prod_{i \in J} p_i \prod_{k \in J^c} (1 - p_k)$$

for $x \in \mathbb{N}_0$, where $J$ denotes a generic subset of $\mathbb{N}$ and $J^c := \mathbb{N} \setminus J$. It is well-known that the distribution $Q$ may be approximated by the Poisson distribution $\text{Pois}(\lambda)$ with weights

$$\pi(x) = e^{-\lambda} \lambda^x / x!,$$
provided that the quantity
\[ \Delta := \lambda^{-1} \sum_{i \geq 1} p_i^2 \]
is small. Indeed, a suitable version of Stein’s method, developed by [Chen (1975)], leads to the remarkable bound
\[ d_{TV}(Q, \text{Poiss}(\lambda)) \leq (1 - e^{-\lambda}) \Delta \leq \sum_{i \geq 1} p_i^2 / \max(1, \lambda), \]
where \( d_{TV}(\cdot, \cdot) \) stands for total variation distance; see Theorem 2.3 in [Barbour et al. (1992)]. Note also that
\[ \text{Var}(X) = \sum_{i \geq 1} p_i (1 - p_i) = \lambda (1 - \Delta). \]

**Conjecture and main results.** Motivated by [Dümbgen et al. (2019)], we are aiming at upper bounds for the maximal density ratio
\[ \rho(Q, \text{Poiss}(\lambda)) := \sup_{x \geq 0} r(x) \]
with
\[ r(x) := \frac{b(x)}{\pi(x)}. \]
Note that for arbitrary sets \( A \subset \mathbb{N}_0 \), the probability \( Q(A) = \mathbb{P}(X \in A) \) is never larger than the corresponding Poisson probability times \( \rho(Q, \text{Poiss}(\lambda)) \), no matter how small the Poisson probability is. Moreover, \( d_{TV}(Q, \text{Poiss}(\lambda)) \leq 1 - \rho(Q, \text{Poiss}(\lambda))^{-1} \). Hence, \( \rho(Q, \text{Poiss}(\lambda)) \) is a strong measure of error when \( Q \) is approximated by \( \text{Poiss}(\lambda) \). We conjecture that
\[ \rho(Q, \text{Poiss}(\lambda)) \leq (1 - \Delta)^{-1}. \] (2)
In this note we prove that
\[ \rho(Q, \text{Poiss}(\lambda)) \leq (1 - p_\ast)^{-1} \] (3)
for arbitrary values of \( \lambda \), where
\[ p_\ast := \max_{i \geq 1} p_i \geq \Delta. \]
In addition, we prove that in case of \( \lambda \leq 1 \), a stronger version of (2) is true:
\[ \rho(Q, \text{Poiss}(\lambda)) \leq e^{\Delta} \quad \text{if} \quad \lambda \leq 1. \] (4)
Note that \( e^{-\Delta} > 1 - \Delta \), whence \( e^{\Delta} < (1 - \Delta)^{-1} \).

In Section 2 we provide some basic formulae for the weights \( b(x) \) and the ratios \( r(x) \). These lead to a preliminary bound for the maximizer(s) of \( r = b/\pi \) and a first bound for \( \rho(Q, \text{Poiss}(\lambda)) \). Then in Section 3 we derive the upper bound (3). In Section 4 we discuss the case \( 0 < \lambda \leq 1 \) and provide lower and upper bounds for \( \rho(Q, \text{Poiss}(\lambda)) \).
2 Preparations

Discrete scores. With $n := \# \{ i \geq 1 : p_i > 0 \} \in \mathbb{N} \cup \{ \infty \}$, note that $b(x) > 0$ if and only if $x \leq n$. For any $x \geq 0$,

$$\frac{\pi(x+1)}{\pi(x)} = \frac{\lambda}{x+1},$$

so the “scores” $r(x+1)/r(x)$ are given by

$$\frac{r(x+1)}{r(x)} = \frac{(x+1)b(x+1)}{\lambda b(x)}$$

for $x \geq 0$ with $b(x) > 0$. If $x$ is a maximizer of $r(\cdot)$, then

$$\frac{(x+1)b(x+1)}{b(x)} \leq \lambda \leq \frac{xb(x)}{b(x-1)} \quad (5)$$

with $b(-1) := 0$.

Representing the weight function of $Q$. The weight function $b$ may be written as

$$b(x) = \sum_{J: \#J = x} w(J) \quad \text{with} \quad w(J) := \prod_{i \in J} p_i \prod_{k \in J^c} (1 - p_k).$$

In particular,

$$b(0) = \prod_{k \geq 1} (1 - p_k) = \exp\left(\sum_{k \geq 1} \log(1 - p_k)\right) < \exp(-\lambda) = \pi(0),$$

because $\log(1 + y) < y$ for $-1 < y \neq 0$. Since

$$w(J) = \prod_{i \in J} \frac{p_i}{1 - p_i} \prod_{k \geq 1} (1 - p_k) = b(0) \prod_{i \in J} \frac{p_i}{1 - p_i},$$

we can also write $w(J) = b(0) W(J)$ and

$$b(x) = b(0) \sum_{J: \#J = x} W(J) \quad \text{with} \quad W(J) := \prod_{i \in J} q_i,$$

where

$$q_i := \frac{p_i}{1 - p_i} \in [0, \infty), \quad p_i = \frac{q_i}{1 + q_i}.$$  

Ratios of consecutive binomial weights. There are various ways to represent the ratios $b(x+1)/b(x)$. In the subsequent versions, the following notation will be useful: For any set $J \subset \mathbb{N}$, we define

$$s(J) := \sum_{i \in J} p_i \quad \text{and} \quad S(J) := \sum_{i \in J} q_i.$$
In case of \( \#J < \infty \) we set
\[
\bar{s}(J) := s(J)/\#J, \\
\bar{S}(J) := S(J)/\#J, \\
\bar{W}(J) := W(J)/\sum_{L : \#L = \#J} W(L) = w(J)/\sum_{L : \#L = \#J} w(L)
\]
with the convention \( 0/0 := 0 \). Then for any integer \( x \geq 0 \) with \( b(x) > 0 \),
\[
\frac{b(x + 1)}{b(0)} = \sum_{L : \#L = x + 1} W(L) = \sum_{L : \#L = x + 1} \frac{1}{x + 1} \sum_{k \in L} W(L \setminus \{k\}) q_k \\
= \frac{1}{x + 1} \sum_{J : \#J = x} W(J) \sum_{k \in J^c} q_k \\
= \frac{1}{x + 1} \sum_{J : \#J = x} W(J) S(J^c).
\]
Consequently,
\[
\frac{(x + 1)b(x + 1)}{b(x)} = \sum_{J : \#J = x} W(J) S(J^c). \tag{6}
\]
Alternatively, if \( b(x + 1) > 0 \), then
\[
\frac{b(x)}{b(0)} = \sum_{J : \#J = x} W(J) = \sum_{J : \#J = x} W(J) \sum_{k \in J^c} \frac{q_k}{s(J^c)} \\
= \sum_{J : \#J = x} \sum_{k \in J^c} \frac{W(J \cup \{k\})}{q_k + S((J \cup \{k\})^c)} \\
= \sum_{L : \#L = x + 1} \frac{W(L)}{\sum_{k \in L} q_k + S(L^c)}.
\]
Consequently,
\[
\frac{b(x)}{(x + 1)b(x + 1)} = \sum_{L : \#L = x + 1} \bar{W}(L) \frac{1}{x + 1} \sum_{k \in L} \frac{1}{q_k + S(L^c)}. \tag{7}
\]
One can repeat the previous arguments with the sums \( \sum_{k \in J^c} p_j/s(J^c) = 1 \) in place of \( \sum_{k \in J^c} q_k/s(J^c) = 1 \). This leads to
\[
\frac{b(x)}{b(0)} = \sum_{J : \#J = x} \sum_{k \in J^c} \frac{W(J)p_k}{p_k + s((J \cup \{k\})^c)} \\
= \sum_{L : \#L = x + 1} W(L) \sum_{k \in L} \frac{1 - p_k}{p_k + s(L^c)},
\]
because \( W(J)p_k = W(J \cup \{k\})(1 - p_k) \) for \( k \in J^c \). Consequently,
\[
\frac{b(x)}{(x + 1)b(x + 1)} = \sum_{L : \#L = x + 1} \bar{W}(L) \frac{1}{x + 1} \sum_{k \in L} \frac{1 - p_k}{p_k + s(L^c)}. \tag{8}
\]
Analyzing equation (8) will lead to a first result about the location of maximizers of \( r(\cdot) \) plus a preliminary bound for \( \rho(Q, \text{Poiss}(\lambda)) \).
Proposition 1. Any maximizer \( x \in \mathbb{N}_0 \) of \( r(x) \) satisfies the inequalities
\[
1 \leq x \leq \lceil \lambda \rceil.
\]

Moreover,
\[
\rho(Q, \text{Poiss}(\lambda)) \leq \left(1 + \Delta \frac{e^{p_*} - 1}{p_*}\right)^{\lceil \lambda \rceil} \leq e^{[\lambda]p_*}.
\]

Proof of Proposition 1. Since \( r(0) < 1 \), any maximizer \( x_o \) of \( r(\cdot) \) has to satisfy \( x_o \geq 1 \).

To verify the inequality \( x_o \leq \lceil \lambda \rceil \), it suffices to show that for any \( x \geq \lambda \) with \( b(x) > 0 \),
\[
\frac{r(x + 1)}{r(x)} \leq 1.
\]

This is equivalent to
\[
\frac{b(x)}{(x + 1)b(x + 1)} \geq \lambda^{-1}.
\] (9)

If \( b(x + 1) = 0 \), this inequality is trivial. Otherwise, according to (8), the left hand side of (9) equals
\[
\sum_{L: \#L = x + 1} \bar{W}(L) \frac{1}{x + 1} \sum_{k \in L} \frac{1 - p_k}{p_k + s(L^c)}.
\]

Since \( (1 - y)/(y + s(L^c)) \) is a convex function of \( y \geq 0 \), Jensen’s inequality implies that
\[
\frac{1}{x + 1} \sum_{k \in L} \frac{1 - p_k}{p_k + s(L^c)} \geq \frac{1 - \bar{s}(L)}{\bar{s}(L) + s(L^c)} = \frac{1 - \bar{s}(L)}{\bar{s}(L) + \lambda - s(L)} = \frac{1 - \bar{s}(L)}{\lambda - x\bar{s}(L)}.
\]

But in case of \( x \geq \lambda \),
\[
\frac{1 - \bar{s}(L)}{\lambda - x\bar{s}(L)} \geq \frac{1 - \bar{s}(L)}{\lambda - \lambda \bar{s}(L)} = \lambda^{-1},
\]
whence (9) holds true.

Now we only need an upper bound for \( r(x) \) and apply it with \( x \leq \lceil \lambda \rceil \). First of all,
\[
r(x) = \lambda^{-x} x! e^\lambda \sum_{J: \#J = x} \prod_{i \in J} p_i \prod_{k \in J^e} (1 - p_k)
\]
\[
= \lambda^{-x} x! \sum_{J: \#J = x} \prod_{i \in J} p_i e^{p_i} \prod_{k \in J^e} \exp(p_k + \log(1 - p_k))
\]
\[
\leq \lambda^{-x} x! \sum_{J: \#J = x} \prod_{i \in J} p_i e^{p_i}
\]
\[
\leq \lambda^{-x} \sum_{k(1), \ldots, k(x) \geq 1} \prod_{s=1}^{x} p_{k(s)} e^{p_{k(s)}} = \left( \sum_{k \geq 1} \frac{p_k}{\lambda} e^{p_k} \right)^x.
\]

Moreover,
\[
e^{p_k} \leq 1 + (p_k/p_*) (e^{p_*} - 1) \leq e^{p_*}
\]
by convexity and monotonicity of the exponential function, whence
\[
\sum_{k \geq 1} \frac{p_k}{\lambda} e^{p_k} \leq 1 + \Delta \frac{e^{p_*} - 1}{p_*} \leq e^{p_*}.
\]

\[\square\]
3 Bounds in terms of $p_*$

3.1 A general strategy to verify upper bounds

In what follows, the dependency of objects such as $Q, b, r, w(J), \ldots$ on the sequence $p$ is indicated by a subscript $p$ if necessary, leading to $Q_p, b_p, r_p, w_p(J), \ldots$, and we write $\pi = \pi_{\lambda}$. Let $A = A(p) \in [0, 1)$ stand for a positively homogeneous functional of $p$, i.e.

$$A(tp) = tA(p) \quad \text{for } t \in (0, 1].$$

Two examples for such a functional are $A = \Delta$ and $A = p_*

Suppose we want to prove that

$$\log \rho(Q, \text{Pois}(\lambda)) \leq g(A)$$

for a given differentiable function $g : [0, 1) \to [0, \infty)$ with $g(0) = 0$ and $g'(0) \geq 1$. An explicit example is given by $g(s) := -\log(1 - s)$. To verify this conjecture, we analyze the function $f : (0, 1] \to \mathbb{R}$ given by

$$f(t) := \log \rho(Q_{tp}, \text{Pois}(t\lambda)) - g(tA),$$

so the assertion is equivalent to $f(1) \leq 0$. Hence, it suffices to show that $f(0+) = 0$ and that $f$ is nonincreasing.

Note that replacing $p$ with $tp$ amounts to replacing $\lambda$ and $\Delta$ with $t\lambda$ and $t\Delta$, respectively. By Proposition 1 we know that

$$\rho(Q_{tp}, \text{Pois}(t\lambda)) = \max_{1 \leq x \leq \lceil t\lambda \rceil} r_{tp}(x)$$

and

$$f(t) \leq [t\lambda]tp_* - g(tA).$$

This implies already that $f(0+) = 0$. If we can show that for any fixed $x \in \{1, \ldots, \lceil \lambda \rceil\}$, the log-density ratio $L_x(t) := \log r_{tp}(x)$ is a continuously differentiable function of $t \in (0, 1]$, then $f$ is continuous on $(0, 1]$ with limit $f(0+) = 0$, and for $t < 1$,

$$f'(t+) = \max_{x \in N(t)} L'_x(t) - \tilde{g}(tA)/t,$$

where

$$N(t) := \arg \max_{1 \leq x \leq \lceil t\lambda \rceil} r_{tp}(x)$$

and

$$\tilde{g}(s) := sg'(s).$$

6
Then a sufficient condition for $f(1) \leq 0$ is that $f'(t+) \leq 0$ for all $t \in (0, 1)$, and this can be rewritten as follows: For $t \in (0, 1)$ and $1 \leq x \leq \lceil t \lambda \rceil$,

$$L'_x(t) \leq \tilde{g}(tA)/t$$

if $x \in N(t)$.

In view of (7), a sufficient condition for that is

$$L'_x(t) \leq \tilde{g}(tA)/t$$

if $rac{xb_{tp}(x)}{b_{tp}(x-1)} \geq t\lambda$.  \hspace{1cm} \text{(11)}

Now it is high time to analyze the functions $L_x(\cdot)$ for $1 \leq x \leq \lceil \lambda \rceil$. The inequality $x \leq \lceil \lambda \rceil$ implies that $b(x) > 0$, because otherwise, $\lambda$ would be a sum of $x - 1$ weights $p_i \in [0, 1)$, and this would lead to the contradiction $\lceil \lambda \rceil \leq x - 1$. For a set $J \subset \mathbb{N}$ with $\# J = x$,

$$\frac{\partial}{\partial t} w_{tp}(J) = \frac{\partial}{\partial t} \prod_{i \in J} \prod_{k \in J^c} (1 - tp_k)$$

$$= xt^{-1} \prod_{i \in J} \prod_{k \in J^c} (1 - tp_k) - \sum_{\ell \in J^c} t^x \prod_{i \in J} \prod_{k \in J \setminus \{\ell\}} (1 - tp_k)$$

$$= xt^{-1} \prod_{i \in J} \prod_{k \in J^c} (1 - tp_k) - \sum_{\ell \in J^c} t^x \prod_{i \in J \cup \{\ell\}} \prod_{k \in (J \cup \{\ell\})^c} (1 - tp_k)$$

$$= \frac{x}{t} w_{tp}(J) - \frac{1}{t} \sum_{\ell \in J^c} w_{tp}(J \cup \{\ell\}).$$

Consequently,

$$\frac{\partial}{\partial t} b_{tp}(x) = \sum_{J: \# J = x} \frac{\partial}{\partial t} w_{tp}(J)$$

$$= \frac{x}{t} \sum_{J: \# J = x} w_{tp}(J) - \frac{1}{t} \sum_{J: \# J = x} \sum_{\ell \in J^c} w_{tp}(J \cup \{\ell\})$$

$$= \frac{x}{t} \sum_{J: \# J = x} w_{tp}(J) - \frac{1}{t} \sum_{L: \# L = x+1} \sum_{\ell \in L} w_{tp}(L)$$

$$= \frac{x}{t} w_{tp}(J) - \frac{x + 1}{t} \sum_{L: \# L = x+1} w_{tp}(L)$$

$$= \frac{x}{t} b_{tp}(x) - \frac{x + 1}{t} b_{tp}(x + 1).$$

This gives us the identity

$$\frac{\partial}{\partial t} \log b_{tp}(x) = \frac{x}{t} - \frac{x + 1}{t} \frac{b_{tp}(x + 1)}{b_{tp}(x)}.$$

An elementary calculation yields

$$\frac{\partial}{\partial t} \log \pi_{t\lambda}(x) = \frac{x}{t} - \lambda.$$
so
\[ L'_x(t) = \frac{\partial}{\partial t} \log r_{tp}(x) = \lambda - \frac{x + 1}{t} \frac{b_{tp}(x + 1)}{b_{tp}(x)}. \]

Consequently, (11) may be rewritten as follows: For each \( t \in (0, 1) \) and \( 1 \leq x \leq \lfloor t\lambda \rfloor \),
\[
\frac{(x + 1)b_{tp}(x + 1)}{b_{tp}(x)} \geq t\lambda - \tilde{g}(tA) \quad \text{if} \quad \frac{xb_{tp}(x)}{b_{tp}(x - 1)} \geq t\lambda.
\]

Since we could replace \( p \) with \( t p \), it even suffices to show that for \( 1 \leq x \leq \lceil \lambda \rceil \),
\[
\frac{(x + 1)b(x + 1)}{b(x)} \geq \lambda - \tilde{g}(A) \quad \text{if} \quad \frac{xb(x)}{b(x - 1)} \geq \lambda. \tag{12}
\]

Note that \( b(1)/b(0) = \sum_{i \geq 1} q_i > \sum_{i \geq 1} p_i = \lambda \), so (12) implies that
\[
\frac{2b(2)}{b(1)} \geq \lambda - \tilde{g}(A).
\]

### 3.2 The main result

In case of \( A = p^* \) and \( g(s) = -\log(1 - s) \), the strategy just outlined works nicely, leading to our first main result. Note that \( \tilde{g}(s) = s/(1 - s) \).

**Theorem 1.** For any sequence \( p \) of probabilities \( p_i \in [0, 1) \) with \( \lambda = \sum_{i \geq 1} p_i < \infty \),
\[
\rho(Q, \text{Pois} (\lambda)) \leq (1 - p^*)^{-1}.
\]

**Proof of Theorem 1.** For \( 1 \leq x \leq \lceil \lambda \rceil \), the representation (7) with \( x - 1 \) in place of \( x \) reads
\[
\frac{b(x - 1)}{xb(x)} = \sum_{J : \# J = x} \tilde{W}(J) \frac{1}{x} \sum_{i \in J} \frac{1}{q_i + S(J^c)}.
\]

By Jensen’s inequality,
\[
\frac{1}{x} \sum_{i \in J} \frac{1}{q_i + S(J^c)} \geq \left( \frac{1}{x} \sum_{i \in J} (q_i + S(J^c)) \right)^{-1} = (\tilde{S}(J) + S(J^c))^{-1},
\]
so
\[
\frac{b(x - 1)}{xb(x)} \geq \sum_{J : \# J = x} \tilde{W}(J) (\tilde{S}(J) + S(J^c))^{-1}.
\]

A second application of Jensen’s inequality yields that
\[
\frac{b(x - 1)}{xb(x)} \geq \left( \sum_{J : \# J = x} \tilde{W}(J) (\tilde{S}(J) + S(J^c)) \right)^{-1}.
\]

Consequently, if \( xb(x)/b(x - 1) \geq \lambda \), then
\[
\sum_{J : \# J = x} \tilde{W}(J) (\tilde{S}(J) + S(J^c)) \geq \lambda.
\]
On the other hand, (6) yields
\[
\frac{(x + 1)b(x + 1)}{b(x)} = \sum_{J, \# J = x} W(J)S(J^c)
\]
\[
= \sum_{J, \# J = x} W(J)(\bar{S}(J) + S(J^c)) - \sum_{J, \# J = x} W(J)\bar{S}(J)
\]
\[
\geq \lambda - \sum_{J, \# J = x} W(J)\bar{S}(J)
\]
\[
\geq \lambda - \frac{p_*}{1 - p_*} = \lambda - \bar{g}(p_*),
\]
because for any set \(J\) with \(x\) elements,
\[
\bar{S}(J) = \frac{1}{x} \sum_{i \in J} q_i \leq \frac{p_*}{1 - p_*}.
\]
Consequently, (12) is satisfied with \(A = p_*\), and this yields the assertion.

\[\]

4 Bounds in terms of \(\Delta\)

At the moment we do not know whether our general strategy works for \(A = \Delta\). Instead we derive some bounds via direct arguments. We start with an elementary result about the log-density ratio \(L_1(t) = \log r_p(1)\).

**Proposition 2.** The function \(L_1 : [0, 1] \rightarrow \mathbb{R}\) is twice differentiable with \(L_1(0) = 0\), \(L'_1(0) = \Delta\) and \(L''_1 \leq 0\) with equality if and only if \(#\{i \geq 1 : p_i > 0\} = 1\).

**Proof of Proposition 2.** Note first that for \(t \in (0, 1]\),
\[
L_1(t) = t\lambda + \log \left( (t\lambda)^{-1} \sum_{i \geq 1} (tp_i) \prod_{k \neq i} (1 - tp_k) \right)
\]
\[
= t\lambda + \log \left( \lambda^{-1} \sum_{i \geq 1} p_i \prod_{k \neq i} (1 - tp_k) \right)
\]
\[
= \sum_{i \geq 1} (tp_i + \log(1 - tp_i)) + \log \left( \lambda^{-1} \sum_{i \geq 1} \frac{p_i}{1 - tp_i} \right).
\]
The right hand side is a smooth function of \(t \in [0, 1]\) with \(L_1(0) = 0\). Moreover,
\[
L'_1(t) = \sum_{i \geq 1} \left( p_i - \frac{p_i}{1 - tp_i} \right) + \sum_{i \geq 1} \frac{p_i^2}{(1 - tp_i)^2} / \sum_{i \geq 1} \frac{p_i}{1 - tp_i}
\]
\[
= -t \sum_{i \geq 1} p_i^2 / (1 - tp_i) + \sum_{i \geq 1} \frac{p_i^2}{(1 - tp_i)^2} / \sum_{i \geq 1} \frac{p_i}{1 - tp_i},
\]
\[
L'_1(0) = \sum_{i \geq 1} p_i^2 / \sum_{i \geq 1} p_i = \Delta.
\]
Finally, with $a_i(t) := p_i/(1 - tp_i)$ and $S(t) := \sum_{i \geq 1} a_i(t)$,

\[
L''_1(t) = - \sum_{i \geq 1} \frac{p_i^2}{(1 - tp_i)^2} + 2 \sum_{i \geq 1} \frac{p_i^3}{(1 - tp_i)^3} / \sum_{i \geq 1} \frac{p_i}{1 - tp_i} \\
- \left( \sum_{i \geq 1} \frac{p_i^2}{(1 - tp_i)^2} / \sum_{i \geq 1} \frac{p_i}{1 - tp_i} \right)^2 \\
= - \sum_{i \geq 1} a_i(t)^2 + 2 \sum_{i \geq 1} a_i(t)^3 / S(t) - \sum_{i, j \geq 1} a_i(t)^2 a_j(t)^2 / S(t)^2 \\
\leq - \sum_{i \geq 1} (a_i(t)^2 - 2a_i(t)^3 / S(t) + a_i(t)^4 / S(t)^2) \\
= - \sum_{i \geq 1} a_i(t)^2 (1 - a_i(t) / S(t))^2 \\
\leq 0.
\]

The second last inequality is strict, unless $\#\{i \geq 1 : p_i > 0\} = 1$, and in that case both preceding inequalities are equalities.

Propositions 1 and 2 are the main ingredients for the following upper bound for $\log r(Q, \text{Poiss}(\lambda))$.

**Theorem 2.** For any sequence $p$ of probabilities $p_i \in [0, 1)$ with $\lambda = \sum_{i \geq 1} p_i \leq 1$,

\[
\Delta \geq \log r(Q, \text{Poiss}(\lambda)) \geq \Delta \left( 1 - \frac{\Delta}{2} - \frac{\lambda}{2(1 - p_\star)} \right).
\]

Since $\Delta \leq p_\star \leq \lambda$, this theorem shows that

\[
\frac{\log r(Q, \text{Poiss}(\lambda))}{\Delta} \rightarrow 1 \quad \text{as} \quad \lambda \rightarrow 0.
\]

**Proof of Theorem 2.** We know from Proposition 1 that in case of $\lambda \leq 1$,

\[
\log r(Q, \text{Poiss}(\lambda)) = \log r(1) = L_1(1).
\]

But Proposition 2 implies that for some $\xi \in (0, 1)$,

\[
L_1(1) = L_1(0) + L'_1(0) + 2^{-1} L''_1(\xi) = 0 + \Delta + 2^{-1} L''_1(\xi) \leq \Delta.
\]

As to the lower bound, recall that

\[
L_1(1) = \sum_{i \geq 1} (p_i + \log(1 - p_i)) + \log \left( \lambda^{-1} \sum_{i \geq 1} \frac{p_i}{1 - p_i} \right).
\]

On the one hand,

\[
p_i + \log(1 - p_i) = - \sum_{k \geq 2} \frac{p_i^k}{k} \geq - \frac{p_i^2}{2} \sum_{\ell \geq 0} p_\star^\ell = - \frac{p_i^2}{2(1 - p_\star)}.
\]
\[
\sum_{i \geq 1} (p_i + \log(1 - p_i)) \geq -\frac{1}{2(1 - p^*)} \sum_{i \geq 1} p_i^2 = -\frac{\lambda}{2(1 - p^*)} \Delta.
\]

Moreover,
\[
\log\left(\lambda^{-1} \sum_{i \geq 1} \frac{p_i}{1 - p_i}\right) \geq \log\left(\lambda^{-1} \sum_{i \geq 1} (p_i + p_i^2)\right) = \log(1 + \Delta) \geq \Delta - \Delta^2/2,
\]

and this implies the asserted lower bound for \(L_1(1)\).

**Remark 3** (Total variation distance). Since \(b(0) \leq \pi(0)\), Theorem 2 implies that in case of \(\lambda \leq 1\),
\[
d_{TV}(Q, \text{Poiss}(\lambda)) = \sup_{A \subseteq \{1, 2, 3, \ldots\}} (Q(A) - \text{Poiss}(\lambda)(A))
\]
\[
\leq \sup_{A \subseteq \{1, 2, 3, \ldots\}} Q(A)(1 - \rho(Q, \text{Poiss}(\lambda))^{-1})
\]
\[
\leq (1 - b(0))(1 - e^{-\Delta})
\]
\[
\leq \lambda(1 - e^{-\Delta}) \leq \lambda \Delta = \sum_{i \geq 1} p_i^2.
\]

Here we used the elementary inequalities \(1 - b(0) = 1 - \prod_{i \geq 1} (1 - p_i) \leq \sum_{i \geq 1} p_i = \lambda\) and \(1 - e^{-\Delta} \leq \Delta\). Consequently, Theorem 2 implies a reasonable upper bound for \(d_{TV}(Q, \text{Poiss}(\lambda))\).

**Acknowledgement.** Part of this research was conducted at Mathematisches Forschungsinstitut Oberwolfach (MFO), Germany, in June and July 2019. We are grateful to the MFO for its generous hospitality and support.

**References**

Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson approximation*, vol. 2 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York. Oxford Science Publications.

Chen, L. H. Y. (1975). Poisson approximation for dependent trials. *Ann. Probability* 3 534–545.

Dümbgen, L., Samworth, R. J. and Wellner, J. A. (2019). Bounding distributional errors via density ratios. Tech. rep., University of Bern. (arXiv:1905.03009).