The Density Ratio of Poisson Binomial versus Poisson Distributions

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

June 12, 2020

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 maximum of the density ratio $b(x)/\pi_\lambda(x)$, where π_λ is the probability mass function of the Poisson distribution with parameter λ.

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

1 Introduction and main results

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 sum $X = \sum_{i\ge1} Z_i$. By the first and second Borel–Cantelli lemmas, X is almost surely finite if and only if the sequence $\boldsymbol{p} = (p_i)_{i\ge1}$ satisfies

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

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

$$b(x) = b_{\boldsymbol{p}}(x) := \mathbb{P}(X = x) \ = \ \sum_{J\in\mathcal{J}(x)} \prod_{i\in J} p_i \prod_{k\in J^c} (1 - p_k) \tag{2}$$

for integers $x \ge 0$, where $\mathcal{J}(x) := \{J \subset \mathbb{N} : \#J = x\}$ and $J^c := \mathbb{N} \setminus J$.

It is well-known that the distribution Q may be approximated by the Poisson distribution Poiss_λ with probability mass function $\pi = \pi_\lambda$ given by $\pi(x) = e^{-\lambda}\lambda^x/x!$, provided

*Research supported by Swiss National Science Foundation
†Research supported in part by: (a) NSF Grant DMS-1566514; and (b) NI-AID Grant 2R01 AI291968-04

that the quantity

$$\Delta \ := \ \lambda^{-1} \sum_{i \ge 1} p_i^2$$

is small. Indeed, Barbour and Hall (1984) obtained the remarkable bound

$$d_{\mathrm{TV}}(Q, \mathrm{Poiss}_\lambda) \ \le \ (1 - e^{-\lambda})\Delta$$

via a suitable version of Stein's method developed by Chen (1975). Here $d_{\mathrm{TV}}(\cdot,\cdot)$ stands for total variation distance. Note also that $\mathrm{Var}(X) = \sum_{i \ge 1} p_i(1-p_i) = \lambda(1-\Delta)$, and

$$\Delta \ \le \ p_* \ := \ \max_{i \ge 1} p_i.$$

Main results. Motivated by Dümbgen et al. (2020), we are aiming at upper bounds for the maximal density ratio

$$\rho(Q, \mathrm{Poiss}_\lambda) \ := \ \sup_{x \ge 0} r(x)$$

with $r(x) = r_{\boldsymbol{p}}(x) := b(x)/\pi(x)$. Note that the probability mass functions b and π are densities (in the sense of the Radon-Nikodym theorem) of Q and Poiss_λ with respect to counting measure on the set $\mathbb{N}_0$ of nonnegative integers. Thus $r = b/\pi_\lambda$ is the "density ratio" in the title. 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 $\rho\bigl(Q, \mathrm{Poiss}_\lambda\bigr)$, no matter how small the Poisson probability is. Hence, $\rho(Q, \mathrm{Poiss}_\lambda)$ is a strong measure of error when Q is approximated by Poiss_λ, see also Remark 3 below. While Dümbgen et al. (2020) obtained explicit and essentially sharp bounds for $\rho(Q, P)$ for various pairs of distributions P and Q, the present setting with the particular Poisson binomial distribution Q and $P = \mathrm{Poiss}_\lambda$ seems to be substantially more difficult. In this note we prove the following result:

Theorem 1. *For any sequence* $\boldsymbol{p}$ *of probabilities* $p_i \in [0,1)$ *with* $\lambda = \sum_{i \ge 1} p_i < \infty$,

$$\rho(Q, \mathrm{Poiss}_\lambda) \ \le \ (1 - p_*)^{-1}.$$

We conjecture that Theorem 1 is true with Δ in place of p_*. In the case of $\lambda \le 1$ we can prove the following result:

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

$$\Delta\Bigl(1 - \frac{\Delta}{2} - \frac{\lambda}{2(1-p_*)}\Bigr) \ \le \ \log \rho(Q, \mathrm{Poiss}_\lambda) \ \le \ \Delta.$$

In particular, $\lambda \le 1$ implies that $\rho(Q, \mathrm{Poiss}_\lambda) \le e^\Delta < 1/(1-\Delta)$. And since $\Delta \le p_* \le \lambda$, Theorem 2 implies that

$$\frac{\log \rho(Q, \mathrm{Poiss}_\lambda)}{\Delta} \ \to \ 1 \quad \text{as } \lambda \to 0.$$

Remark 3 (Total variation distance). Proposition 1 (a) of Dümbgen et al. (2020) implies that $d_{\mathrm{TV}}(Q, \mathrm{Poiss}_\lambda) \le Q(\{b > \pi\})\bigl(1 - \rho(Q, \mathrm{Poiss}_\lambda)^{-1}\bigr)$. Since $b(0) = \prod_{i\ge 1}(1 - p_i)$ satisfies the two inequalities $1 - \lambda \le b(0) < e^{-\lambda} = \pi(0)$, we obtain the inequality $Q(\{b > \pi\}) \le 1 - b(0) \le \min(1, \lambda)$ and the bounds

$$\begin{aligned} d_{\mathrm{TV}}(Q, \mathrm{Poiss}_\lambda) \ &\le \ \min(1,\lambda)\bigl(1 - \rho(Q, \mathrm{Poiss}_\lambda)^{-1}\bigr) \\ &\le \ \begin{cases} \min(1,\lambda) p_* \\ \lambda(1 - e^{-\Delta}) \ \le \ \lambda\Delta = \sum_{i\ge 1} p_i^2 \quad \text{if } \lambda \le 1. \end{cases} \end{aligned}$$

The remainder of this note is structured as follows: In Section 2 we provide some basic formulae for the probability masses $b(x)$ and the ratios $r(x)$. Then we present the proofs of Theorems 1 and 2 in Section 3.

2 Auxiliary results

2.1 The probability mass function of Q

Since $b(0) < 1$ (see Remark 3), we know that $\rho(Q, \mathrm{Poiss}_\lambda) = \sup_{x\ge 1} r(x)$. Writing

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

equation (2) may be reformulated as

$$b(x) \ = \ b(0) \sum_{J\in\mathcal{J}(x)} W(J)$$

with

$$W(J) \ := \ \prod_{i\in J} q_i \quad \text{and} \quad q_i \ := \ \frac{p_i}{1 - p_i} \ \in \ [0,\infty),$$

i.e. $p_i = q_i/(1 + q_i)$. Note also that the support of Q is equal to an integer interval containing 0. Precisely,

$$b(x) \ > \ 0 \quad \text{if and only if} \quad x \ \le \ \#\{i \ge 1 : p_i > 0\} \in \mathbb{N} \cup \{\infty\}.$$

2.2 Discrete scores

For any $x \ge 0$,

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

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_o is a maximizer of $r(\cdot)$, then

$$\frac{(x_o+1)b(x_o+1)}{b(x_o)} \ \leq \ \lambda \ \leq \ \frac{x_o b(x_o)}{b(x_o-1)} \tag{3}$$

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

There are various ways to represent the ratios $b(x+1)/b(x)$. The following notation will be useful for that task: 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 $x := \#J < \infty$ we set

$$\bar{s}(J) \ := \ s(J)/x, \quad \bar{S}(J) \ := \ S(J)/x \quad \text{and} \quad \bar{W}(J) \ := \ W(J) \Big/ \sum_{L \in \mathcal{J}(x)} W(L)$$

with the convention $0/0 := 0$. The numbers $\bar{W}(J)$ are probability weights in the sense that $\sum_{J \in \mathcal{J}(x)} \bar{W}(J) = 1$ whenever $b(x) > 0$. In that case,

$$\begin{aligned}
\frac{b(x+1)}{b(0)} = \sum_{L \in \mathcal{J}(x+1)} W(L) \ &= \ \sum_{L \in \mathcal{J}(x+1)} \frac{1}{x+1} \sum_{k \in L} W(L \setminus \{k\}) q_k \\
&= \ \frac{1}{x+1} \sum_{J \in \mathcal{J}(x)} W(J) \sum_{k \in J^c} q_k \\
&= \ \frac{1}{x+1} \sum_{J \in \mathcal{J}(x)} W(J) S(J^c).
\end{aligned}$$

Consequently,

$$\frac{(x+1)b(x+1)}{b(x)} \ = \ \sum_{J \in \mathcal{J}(x)} \bar{W}(J) S(J^c). \tag{4}$$

Alternatively, if $b(x+1) > 0$, then

$$\begin{aligned}
\frac{b(x)}{b(0)} = \sum_{J \in \mathcal{J}(x)} W(J) \ &= \ \sum_{J \in \mathcal{J}(x)} W(J) \sum_{k \in J^c} \frac{q_k}{S(J^c)} \\
&= \ \sum_{J \in \mathcal{J}(x)} \sum_{k \in J^c} \frac{W(J \cup \{k\})}{q_k + S((J \cup \{k\})^c)} \\
&= \ \sum_{L \in \mathcal{J}(x+1)} W(L) \sum_{k \in L} \frac{1}{q_k + S(L^c)}.
\end{aligned}$$

Consequently,

$$\frac{b(x)}{(x+1)b(x+1)} \ = \ \sum_{L \in \mathcal{J}(x+1)} \bar{W}(L) \frac{1}{x+1} \sum_{k \in L} \frac{1}{q_k + S(L^c)}. \tag{5}$$

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

$$\frac{b(x)}{b(0)} = \sum_{J\in\mathcal{J}(x)} \sum_{k\in J^c} \frac{W(J)p_k}{p_k + s((J\cup\{k\})^c)} = \sum_{L\in\mathcal{J}(x+1)} W(L) \sum_{k\in L} \frac{1-p_k}{p_k + s(L^c)},$$

because $W(J)p_k = W(J\cup\{k\})(1-p_k)$ for $k \in J^c$. Consequently,

$$\frac{b(x)}{(x+1)b(x+1)} = \sum_{L\in\mathcal{J}(x+1)} \bar{W}(L) \frac{1}{x+1} \sum_{k\in L} \frac{1-p_k}{p_k + s(L^c)}. \tag{6}$$

Analyzing equation (6) leads to a first result about the location of maximizers of $r(\cdot)$:

Proposition 1. *Any maximizer $x_o \in \mathbb{N}_0$ of $r(\cdot)$ satisfies the inequalities $1 \le x_o \le \lceil\lambda\rceil$.*

Proof of Proposition 1. The inequality $x_o \ge 1$ follows from $r(0) < 1$, see Remark 3. To verify the inequality $x_o \le \lceil\lambda\rceil$, it suffices to show that $r(x+1)/r(x) < 1$ for any integer $x \ge \lambda$ with $b(x) > 0$. This is equivalent to

$$\frac{b(x)}{(x+1)b(x+1)} > \lambda^{-1}. \tag{7}$$

If $b(x+1) = 0$, this inequality is trivial. Otherwise, the left hand side of (7) is given by (6). Since $(1-y)/(y+s(L^c))$ is a strictly convex function of $y \ge 0$, Jensen's inequality implies that

$$\frac{1}{x+1} \sum_{k\in L} \frac{1-p_k}{p_k + s(L^c)} > \frac{1-\bar{s}(L)}{\bar{s}(L) + s(L^c)} = \frac{1-\bar{s}(L)}{\bar{s}(L) + \lambda - s(L)} = \frac{1-\bar{s}(L)}{\lambda - x\bar{s}(L)}.$$

But in case of $x \ge \lambda$,

$$\frac{1-\bar{s}(L)}{\lambda - x\bar{s}(L)} \ge \frac{1-\bar{s}(L)}{\lambda - \lambda\bar{s}(L)} = \lambda^{-1},$$

whence (7) holds true. □

Finally, let us mention that the probability mass function b is ultra-log-concave in the sense that $\log r = \log(b/\pi)$ is concave, i.e. $r(x+1)/r(x)$ is monotone decreasing in $x \in \{y \ge 0 : b(y) > 0\}$, see Section 4 of Saumard and Wellner (2014) and the references therein. Equivalently, $(x+1)b(x+1)/b(x)$ is monotone decreasing in $x \in \{y \ge 0 : b(y) > 0\}$. With a direct argument one can even show a stronger result.

Proposition 2. *The ratio $(x+1)b(x+1)/b(x)$ is strictly decreasing in $x \in \{y \ge 0 : b(y) > 0\}$.*

Proof of Proposition 2. We have to show that for any integer $x \ge 0$ with $b(x+1) > 0$,

$$\frac{(x+2)b(x+2)}{b(x+1)} < \frac{(x+1)b(x+1)}{b(x)}.$$

It follows from (4) that the left hand side equals $S(\mathbb{N}) - \sum_{L\in\mathcal{J}(x+1)} \bar{W}(L)S(L)$ while the right hand side equals $S(\mathbb{N}) - \sum_{J\in\mathcal{J}(x)} \bar{W}(J)S(J)$. Thus the assertion is equivalent to

$$\sum_{J\in\mathcal{J}(x),L\in\mathcal{J}(x+1)} W(J)W(L)\big(S(L) - S(J)\big) \ > \ 0. \tag{8}$$

But each pair $(J, L) \in \mathcal{J}(x)\times\mathcal{J}(x+1)$ is uniquely determined by the three sets $M := J\cap L$, $K := (J \setminus M) \cup (L \setminus M)$ and $L' := L \setminus M$, and

$$W(J)W(L) \ = \ W(M)^2 W(K) \quad \text{and} \quad S(L) - S(J) \ = \ 2S(L') - S(K).$$

Moreover, $\#K = 2x + 1 - 2\#M$ and $\#L' = x + 1 - \#M$. Hence, the left hand side of (8) equals

$$\sum_{s=0}^{x} \sum_{M\in\mathcal{J}(s)} \sum_{K\in\mathcal{J}(2x+1-2s)} 1_{[M\cap K=\emptyset]} W(M)^2 W(K)H(K) \tag{9}$$

with

$$\begin{aligned} H(K) \ &:= \sum_{L'\subset K\,:\,\#L'=x+1-s} \big(2S(L') - S(K)\big) \\ &= \sum_{i\in K} q_i \sum_{L'\subset K\,:\,\#L'=x+1-s} (2\cdot 1_{L'}(i) - 1) \\ &= S(K)\binom{2x-2s}{x-s}\Big/(x + 1 - s). \end{aligned}$$

Hence, all summands in (9) are non-negative, and $W(M)^2 W(K)S(K) > 0$ for suitable sets $M \in \mathcal{J}(x)$ and $K \in \mathcal{J}(1)$ with $M \cap K = \emptyset$. $\square$

2.3 Log-density ratios along a ray

In what follows we consider the sequence $t\boldsymbol{p}$ for arbitrary $t \in (0, 1]$, leading to the distributions $Q_{t\boldsymbol{p}}$ with probability mass functions $b_{t\boldsymbol{p}}$, weights $W_{t\boldsymbol{p}}(J)$ and sums $S_{t\boldsymbol{p}}(J)$. The corresponding Poisson probability mass functions are $\pi_{t\lambda}$, and this leads to the ratios $r_{t\boldsymbol{p}}$. According to Proposition 1,

$$f(t) \ := \ \log\rho(Q_{t\boldsymbol{p}}, \mathrm{Poiss}_{t\lambda}) \ = \ \max_{1\le x\le\lceil t\lambda\rceil} \log r_{t\boldsymbol{p}}(x) \ = \ \max_{1\le x\le\lceil\lambda\rceil} \log r_{t\boldsymbol{p}}(x).$$

Now we analyze the functions $L_x : (0, 1] \to \mathbb{R}$,

$$\begin{aligned} L_x(t) \ &:= \ \log r_{t\boldsymbol{p}}(x) \\ &= t\lambda + \log\Big((t\lambda)^{-x} x! \sum_{J\in\mathcal{J}(x)} \prod_{i\in J} \frac{tp_i}{1 - tp_i} \prod_{k\ge1} (1 - tp_k)\Big) \\ &= t\lambda + \sum_{k\ge1} \log(1 - tp_k) + \log\Big(\lambda^{-x} x! \sum_{J\in\mathcal{J}(x)} \prod_{i\in J} \frac{p_i}{1 - tp_i}\Big), \end{aligned}$$

for integers $x \geq 0$ with $b(x) > 0$. Note first that $L_x(t)$ can be extended to a real-analytic function of $t \in (-\infty, 1/p_*) \supset [0,1]$, and

$$\begin{aligned} L_x(0) &= \log\Bigl(\lambda^{-x} x! \sum_{J \in \mathcal{J}(x)} \prod_{i \in J} p_i\Bigr) \\ &\leq \log\Bigl(\lambda^{-x} \sum_{i(1),\ldots,i(x) \geq 1} \prod_{s=1}^{x} p_{i(s)}\Bigr) \;=\; \log(\lambda^{-x}\lambda^x) \;=\; 0 \end{aligned}$$

with equality for $x = 0, 1$ and strict inequality for $x > 1$. This shows already that f is a Lipschitz-continuous function on $(0,1]$ with limit $f(0+) = 0$.

Concerning the first derivative of L_x, for $t \in (0,1]$,

$$\frac{d}{dt} \prod_{i \in J} \frac{p_i}{1 - tp_i} \;=\; \sum_{k \in J} \frac{p_k^2}{(1 - tp_k)^2} \prod_{i \in J \setminus \{k\}} \frac{p_i}{1 - tp_i} \;=\; \prod_{i \in J} \frac{p_i}{1 - tp_i} \sum_{k \in J} \frac{p_k}{1 - tp_k},$$

whence

$$\begin{aligned} L_x'(t) &= \lambda - \sum_{k \geq 0} \frac{p_k}{1 - tp_k} + \sum_{J \in \mathcal{J}(x)} \prod_{i \in J} \frac{p_i}{1 - tp_i} \sum_{k \in J} \frac{p_k}{1 - tp_k} \Big/ \sum_{J \in \mathcal{J}(x)} \prod_{i \in J} \frac{p_i}{1 - tp_i} \\ &= \lambda - \frac{1}{t}\Bigl(S_{t\boldsymbol{p}}(\mathbb{N}) - \sum_{J \in \mathcal{J}(x)} \bar{W}_{t\boldsymbol{p}}(J) S_{t\boldsymbol{p}}(J)\Bigr) \\ &= \lambda - \frac{1}{t} \sum_{J \in \mathcal{J}(x)} \bar{W}_{t\boldsymbol{p}}(J) S_{t\boldsymbol{p}}(J^c). \end{aligned}$$

Combining this formula with (4) yields

$$\begin{aligned} L_x'(t) &= \lambda - \frac{1}{t} \frac{(x+1) b_{t\boldsymbol{p}}(x+1)}{b_{t\boldsymbol{p}}(x)} \\ &= \lambda - \lambda \frac{r_{t\boldsymbol{p}}(x+1)}{r_{t\boldsymbol{p}}(x)} \\ &= \lambda\bigl(1 - \exp\bigl(L_{x+1}(t) - L_x(t)\bigr)\bigr). \end{aligned} \tag{10}$$

In particular,

$$L_x'(t) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0 \quad \text{if and only if} \quad L_x(t) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} L_{x+1}(t). \tag{11}$$

There is also an explicit expression for the second derivative of L_x: If $b(x+1) = 0$, then $x = n = \#\{i \geq 1 : p_i > 0\}$ and $L_x(t) = \lambda t + \log(\lambda^{-n} n!\, b(n))$, whence $L_x'' \equiv 0$. Otherwise, for $0 < t \leq 1$,

$$L_x''(t) \;=\; \lambda \exp\bigl(L_{x+1}(t) - L_x(t)\bigr)\bigl(L_x'(t) - L_{x+1}'(t)\bigr),$$

and

$$L_x'(t) - L_{x+1}'(t) \;=\; \frac{1}{t}\Bigl(\frac{(x+2) b_{t\boldsymbol{p}}(x+2)}{b_{t\boldsymbol{p}}(x+1)} - \frac{(x+1) b_{t\boldsymbol{p}}(x+1)}{b_{t\boldsymbol{p}}(x)}\Bigr) \;<\; 0$$

by Proposition 2. Hence L_x defines a smooth concave function on $[0,1]$.

3 Proofs of the main results

Proof of Theorem 1. We know that $f(t) = \log \rho(Q_{t\boldsymbol{p}}, \mathrm{Poiss}_{t\lambda})$ is equal to the maximum of $L_x(t)$ over $x \in \{1, \ldots, \lceil \lambda \rceil\}$, and that $f(0+) = 0$. Note also that

$$f'(t+) \ = \ \max_{x \in N(t)} L'_x(t)$$

where

$$N(t) \ := \ \operatorname*{arg\,max}_{x \in \{1,\ldots,\lceil \lambda \rceil\}} \ r_{t\boldsymbol{p}}(x).$$

Since $g(t) := -\log(1 - tp_*)$ satisfies $g(0) = 0$ and $g'(t) = p_*/(1 - tp_*)$, it suffices to show that

$$L'_x(t) \ \le \ \frac{p_*}{1 - tp_*} \quad \text{for any } x \in N(t).$$

According to (10), the latter requirement is equivalent to

$$\frac{(x+1) b_{t\boldsymbol{p}}(x+1)}{b_{t\boldsymbol{p}}(x)} \ \ge \ t\lambda - \frac{tp_*}{1 - tp_*} \quad \text{for any } x \in N(t).$$

Note that $x \in N(t)$ implies that $L_{x-1}(t) \le L_x(t)$. But the latter inequality is equivalent to $L'_{x-1}(t) \le 0$, see (11), and by (10), this is equivalent to

$$\frac{x b_{t\boldsymbol{p}}(x)}{b_{t\boldsymbol{p}}(x-1)} \ \ge \ t\lambda.$$

Consequently, it suffices to show that

$$\frac{(x+1) b_{t\boldsymbol{p}}(x+1)}{b_{t\boldsymbol{p}}(x)} \ \ge \ t\lambda - \frac{tp_*}{1 - tp_*} \quad \text{whenever} \quad \frac{x b_{t\boldsymbol{p}}(x)}{b_{t\boldsymbol{p}}(x-1)} \ \ge \ t\lambda.$$

We may simplify notation by replacing $t\boldsymbol{p}$ with $\boldsymbol{p}$ and prove that

$$\frac{(x+1) b(x+1)}{b(x)} \ \ge \ \lambda - \frac{p_*}{1 - p_*} \quad \text{whenever} \quad \frac{x b(x)}{b(x-1)} \ \ge \ \lambda. \tag{12}$$

Note that for $1 \le x \le \lceil \lambda \rceil$, the representation (5) with $x - 1$ in place of x reads

$$\frac{b(x-1)}{x b(x)} \ = \ \sum_{J \in \mathcal{J}(x)} \bar{W}(J) \frac{1}{x} \sum_{i \in J} \frac{1}{q_i + S(J^c)}.$$

By Jensen's inequality,

$$\frac{1}{x} \sum_{i \in J} \frac{1}{q_i + S(J^c)} \ \ge \ \Bigl(\frac{1}{x} \sum_{i \in J} (q_i + S(J^c)) \Bigr)^{-1} \ = \ \bigl(\bar{S}(J) + S(J^c) \bigr)^{-1},$$

so

$$\frac{b(x-1)}{x b(x)} \ \ge \ \sum_{J \in \mathcal{J}(x)} \bar{W}(J) \bigl(\bar{S}(J) + S(J^c) \bigr)^{-1}.$$

A second application of Jensen's inequality yields that

$$\frac{b(x-1)}{xb(x)} \;\ge\; \Big(\sum_{J \in \mathcal{J}(x)} \bar{W}(J) \big(\bar{S}(J) + S(J^c) \big) \Big)^{-1}.$$

Consequently, if $xb(x)/b(x-1) \ge \lambda$, then

$$\sum_{J \in \mathcal{J}(x)} \bar{W}(J) \big(\bar{S}(J) + S(J^c) \big) \;\ge\; \lambda.$$

On the other hand, (4) yields

$$\begin{aligned} \frac{(x+1)b(x+1)}{b(x)} \;&=\; \sum_{J \in \mathcal{J}(x)} \bar{W}(J) \big(\bar{S}(J) + S(J^c) \big) - \sum_{J \in \mathcal{J}(x)} \bar{W}(J) \bar{S}(J) \\ &\ge\; \lambda - \frac{p_*}{1-p_*}, \end{aligned}$$

because $\bar{S}(J) = x^{-1} \sum_{i \in J} p_i/(1-p_i) \le p_*/(1-p_*)$ for any set $J \in \mathcal{J}(x)$. This proves (12). □

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

$$\log \rho(Q, \mathrm{Poiss}_\lambda) \;=\; \log r(1) \;=\; L_1(1)$$

with

$$L_1(t) \;=\; t\lambda + \sum_{i \ge 1} \log(1 - tp_i) + \log\Big(\lambda^{-1} \sum_{i \ge 1} \frac{p_i}{1 - tp_i} \Big).$$

First of all, $L_1(0) = 0$, and

$$\begin{aligned} L_1'(t) \;&=\; \lambda - \sum_{i \ge 1} \frac{p_i}{1 - tp_i} + \sum_{i \ge 1} \frac{p_i^2}{(1 - tp_i)^2} \Big/ \sum_{i \ge 1} \frac{p_i}{1 - tp_i} \\ &=\; -t \sum_{i \ge 1} \frac{p_i^2}{1 - tp_i} + \sum_{i \ge 1} \frac{p_i^2}{(1 - tp_i)^2} \Big/ \sum_{i \ge 1} \frac{p_i}{1 - tp_i}, \end{aligned}$$

whence $L_1'(0) = \Delta$. Moreover, we have seen before that $L_1'' \le 0$ by ultra-log-concavity of the probability mass functions $b_{t\boldsymbol{p}}$. Consequently, for some $\xi \in (0,1)$,

$$L_1(1) \;=\; L_1(0) + L_1'(0) + 2^{-1} L_1''(\xi) \;=\; 0 + \Delta + 2^{-1} L_1''(\xi) \;\le\; \Delta.$$

As to the lower bound, recall that

$$L_1(1) \;=\; \sum_{i \ge 1} (p_i + \log(1 - p_i)) + \log\Big(\lambda^{-1} \sum_{i \ge 1} \frac{p_i}{1 - p_i} \Big).$$

On the one hand,

$$p_i + \log(1 - p_i) \;=\; -\sum_{k \ge 2} \frac{p_i^k}{k} \;\ge\; -\frac{p_i^2}{2} \sum_{\ell \ge 0} p_*^\ell \;=\; -\frac{p_i^2}{2(1 - p_*)},$$

so

$$\sum_{i\geq 1}(p_i + \log(1-p_i)) \ \geq \ -\frac{1}{2(1-p_*)}\sum_{i\geq 1} p_i^2 \ = \ -\frac{\lambda}{2(1-p_*)}\Delta.$$

Moreover,

$$\log\Bigl(\lambda^{-1}\sum_{i\geq 1}\frac{p_i}{1-p_i}\Bigr) \ \geq \ \log\Bigl(\lambda^{-1}\sum_{i\geq 1}(p_i+p_i^2)\Bigr) \ = \ \log(1+\Delta) \ \geq \ \Delta - \Delta^2/2,$$

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

Acknowledgement. Part of this research was conducted at the Mathematical Research Institute Oberwolfach (MFO), Germany, in June and July 2019. We are grateful to the MFO for its generous hospitality and support. We also thank a referee and an associate editor for constructive comments.

References

Barbour, A. D. and Hall, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95** 473–480.

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

Dümbgen, L., Samworth, R. J. and Wellner, J. A. (2020). Bounding distributional errors via density ratios. Preprint (arXiv:1905.03009).

Saumard, A. and Wellner, J. A. (2014). Log-concavity and strong log-concavity: a review. *Stat. Surv.* **8** 45–114.