

Alma Mater Studiorum Università di Bologna  
Archivio istituzionale della ricerca

On the Chvatal-Janson conjecture

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Barabesi, L., Pratelli, L., Rigo, P. (2023). On the Chvatal-Janson conjecture. STATISTICS & PROBABILITY LETTERS, 194(March), 1-6 [10.1016/j.spl.2022.109744].

*Availability:*

This version is available at: <https://hdl.handle.net/11585/903684> since: 2023-03-05

*Published:*

DOI: <http://doi.org/10.1016/j.spl.2022.109744>

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

**Lucio Barabesi, Luca Pratelli, Pietro Rigo. (2023). “On the Chvátal–Janson conjecture”. *Statistics & Probability Letters*, Vol. 194, March 2023, Art. No. 109744.**

The final published version is available online at:

<https://doi.org/10.1016/j.spl.2022.109744>

#### Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

*This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)*

***When citing, please refer to the published version.***

# ON THE CHVÁTAL-JANSON CONJECTURE

LUCIO BARABESI, LUCA PRATELLI, AND PIETRO RIGO

ABSTRACT. Let  $q_m = P(X \leq m)$ , where  $m$  is a positive integer and  $X$  a binomial random variable with parameters  $n$  and  $m/n$ . Vašek Chvátal conjectured that, for fixed  $n \geq 2$ ,  $q_m$  attains its minimum when  $m$  is the integer closest to  $2n/3$ . As shown by Svante Janson, this conjecture is true for large  $n$ . Here, we prove that the conjecture is actually true for every  $n \geq 2$ .

## 1. INTRODUCTION

Denoting by  $B(n, p)$  a binomial random variable with parameters  $n$  and  $p$ , Janson [4] investigates the following conjecture suggested by Chvátal in a personal communication.

**Conjecture 1 (Chvátal).** *For any fixed  $n \geq 2$ , as  $m$  ranges over  $\{0, \dots, n\}$ ,*

$$q_m := P(B(n, m/n) \leq m)$$

*is smallest when  $m = \llbracket 2n/3 \rrbracket$  where  $\llbracket \cdot \rrbracket$  represents the nearest integer function.*

In addition to be intriguing, Conjecture 1 may have useful applications, since the probability that a binomial random variable exceeds its expected value plays a role in the machine learning framework; see e.g. [1], [3], [9] and references therein. Such a probability is also connected to an equation given by Ramanujan, as emphasized by [5]. See also [6] and [8] for further results on this topic.

For large  $n$ , Conjecture 1 is actually true and the  $q_m$  have a unique minimum.

**Theorem 1 (Janson, [4]).** *There exists an integer  $n_0$  such that, for each  $n \geq n_0$ : i)  $q_m$  is minimum for  $m = \llbracket 2n/3 \rrbracket$  and ii)  $q_m > q_{m+1}$  or  $q_m < q_{m+1}$  according to whether  $m + \frac{1}{2} < 2n/3$  or  $m + \frac{1}{2} > 2n/3$ .*

As noted in [4, Remark 1.5], in principle, the value of  $n_0$  could be computed and then Conjecture 1 could be proved (or disproved) by considering all  $n < n_0$ . Even if potentially possible, however, this strategy looks not practically feasible and Janson wishes for a general proof of Conjecture 1.

The only purpose of this note is to prove that Conjecture 1 is actually true.

**Theorem 2.** *For each  $n \geq 2$ , one obtains  $q_m > q_{m+1}$  or  $q_m < q_{m+1}$  according to whether  $m + \frac{1}{2} < 2n/3$  or  $m + \frac{1}{2} > 2n/3$ . Hence, if  $m_0 = \llbracket 2n/3 \rrbracket$ , then  $q_{m_0} < q_m$  for each  $m \neq m_0$ .*

---

*Key words and phrases.* Binomial distribution, Binomial tail probability, Bernoulli inequality.

Our proof of Theorem 2 is quite plain and relies on completely different arguments with respect to [4]. In fact, [4] exploits the version for integer-valued random variables of the asymptotic Edgeworth expansion for probabilities in the central limit theorem - as proposed by Esseen [2]. Instead, our proof is closer to the approach introduced by [7, Appendix B] for showing that  $q_m \geq q_{m+1}$  for  $0 \leq m < n/2$  and  $n \geq 2$ .

## 2. TWO PRELIMINARY LEMMAS

Let  $U_1, \dots, U_n$  be  $n$  independent copies of a uniform random variable on  $[0, 1]$  and  $U_{(1)} \leq \dots \leq U_{(n)}$  the corresponding order statistics. For  $m < n$ , since  $U_{(m+1)}$  has a beta distribution with parameters  $m+1$  and  $n-m$ , one obtains

$$\begin{aligned} q_m &= P\left(\sum_{i=1}^n I_{\{U_i \leq m/n\}} \leq m\right) = P(U_{(m+1)} > m/n) \\ &= (m+1) \binom{n}{m+1} \int_{m/n}^1 x^m (1-x)^{n-m-1} dx. \end{aligned}$$

**Lemma 1.** *Let  $n \geq 2$  and  $m \leq n-2$ . Then,  $q_m \geq q_{m+1}$  if and only if*

$$\int_0^1 \left(1 - \frac{v}{m+1}\right)^m \left(1 + \frac{v}{n-m-1}\right)^{n-m-1} dv \geq 1. \quad (2)$$

*Proof.* First note that  $q_m \geq q_{m+1}$  is equivalent to

$$\frac{m+1}{n-m-1} \int_{m/n}^1 x^m (1-x)^{n-m-1} dx \geq \int_{(m+1)/n}^1 x^{m+1} (1-x)^{n-m-2} dx.$$

Integrating the left-hand side by parts, this inequality becomes

$$\int_{m/n}^{(m+1)/n} x^{m+1} (1-x)^{n-m-2} dx \geq \frac{(m/n)^{m+1} (1-m/n)^{n-m-1}}{n-m-1}.$$

Letting  $x = (m+t)/n$  in the integral, one obtains

$$\int_0^1 \left(1 + \frac{t}{m}\right)^{m+1} \left(1 - \frac{t}{n-m}\right)^{n-m-2} dt \geq \frac{n-m}{n-m-1}.$$

Integrating again the left-hand side by parts, such inequality turns into

$$\int_0^1 \left(1 + \frac{t}{m}\right)^m \left(1 - \frac{t}{n-m}\right)^{n-m-1} dt \geq \left(1 - \frac{1}{n-m}\right)^{n-m-1} \left(1 + \frac{1}{m}\right)^m$$

or equivalently

$$\int_0^1 \left(\frac{t+m}{1+m}\right)^m \left(\frac{n-m-t}{n-m-1}\right)^{n-m-1} dt \geq 1.$$

Now, inequality (2) follows from the transformation  $t = 1 - v$ .  $\square$

**Lemma 2.** Fix  $n \geq 3$  and define

$$g_v(x) = \left(1 - \frac{v}{x+1}\right)^x \left(1 + \frac{v}{n-x-1}\right)^{n-x-1}$$

for all  $v \in (0, 1]$  and  $x \in [1, n-2]$ . Then,  $x \mapsto g_v(x)$  is strictly decreasing for each fixed  $v$ . In particular, if

$$h(m) = \int_0^1 g_v(m) dv = \int_0^1 \left(1 - \frac{v}{m+1}\right)^m \left(1 + \frac{v}{n-m-1}\right)^{n-m-1} dv$$

for  $m \in \{1, \dots, n-2\}$ , the function  $h$  is strictly decreasing.

*Proof.* Fix  $(v, x) \in (0, 1] \times [1, n-2]$ , and note that

$$g'_v(x) = g_v(x) \left[ \log\left(1 - \frac{v}{x+1}\right) + \frac{\frac{vx}{(x+1)^2}}{1 - \frac{v}{x+1}} - \log\left(1 + \frac{v}{n-x-1}\right) + \frac{\frac{v}{n-x-1}}{1 + \frac{v}{n-x-1}} \right].$$

Therefore,

$$g'_v(x) < 0 \iff \frac{\frac{vx}{(x+1)^2}}{1 - \frac{v}{x+1}} + \frac{\frac{v}{n-x-1}}{1 + \frac{v}{n-x-1}} < \log\left(\frac{1 + \frac{v}{n-x-1}}{1 - \frac{v}{x+1}}\right).$$

In addition,

$$\begin{aligned} \log\left(\frac{1 + \frac{v}{n-x-1}}{1 - \frac{v}{x+1}}\right) &= \log\left[\left(1 + \frac{v}{n-x-1}\right)\left(1 + \frac{\frac{v}{x+1}}{1 - \frac{v}{x+1}}\right)\right] \\ &= \log\left(1 + \frac{v}{n-x-1}\right) + \log\left(1 + \frac{\frac{v}{x+1}}{1 - \frac{v}{x+1}}\right). \end{aligned}$$

Hence, in order to prove  $g'_v(x) < 0$ , it suffices to show that

$$\frac{\frac{v}{n-x-1}}{1 + \frac{v}{n-x-1}} < \log\left(1 + \frac{v}{n-x-1}\right) \quad (3)$$

and

$$\frac{\frac{vx}{(x+1)^2}}{1 - \frac{v}{x+1}} < \log\left(1 + \frac{\frac{v}{x+1}}{1 - \frac{v}{x+1}}\right). \quad (4)$$

To prove (3)-(4), first note that  $\log(1+c) > c/(1+c)$  for each  $c > 0$ . Therefore, (3) holds with  $c = v/(n-x-1)$ . Similarly, letting  $c = v/(x+1-v)$ , inequality (4) reduces to

$$\log(1+c) - c + \frac{c^2}{v(c+1)} > 0.$$

Finally, the above inequality is true, since

$$\log(1+c) - c + \frac{c^2}{v(c+1)} \geq \log(1+c) - c + \frac{c^2}{c+1} = \log(1+c) - \frac{c}{c+1} > 0.$$

□

## 3. A PROOF OF THE CHVÁTAL-JANSON CONJECTURE

We are now ready to attack Theorem 2. By a direct computation, Theorem 2 holds true for  $n \leq 5$ . Hence, it can be assumed  $n = 3s + r$  where  $s \geq 2$  and  $r \in \{0, 1, 2\}$ . In this case, because of Lemmas 1-2, it suffices to prove that

$$\int_0^1 \left(1 - \frac{v}{2s}\right)^{2s-1} \left(1 + \frac{v}{s}\right)^s dv > 1 \quad (5)$$

and

$$\int_0^1 \left(1 - \frac{v}{2s+1}\right)^{2s} \left(1 + \frac{v}{s-1}\right)^{s-1} dv < 1 \quad (6)$$

if  $r = 0$ , while

$$\int_0^1 \left(1 - \frac{v}{2s+1}\right)^{2s} \left(1 + \frac{v}{s+r-1}\right)^{s+r-1} dv > 1 \quad (7)$$

and

$$\int_0^1 \left(1 - \frac{v}{2s+2}\right)^{2s+1} \left(1 + \frac{v}{s+r-2}\right)^{s+r-2} dv < 1 \quad (8)$$

if  $r \in \{1, 2\}$ .

We point out that, since all the previous inequalities are strict, one obtains  $q_m \neq q_{m+1}$  for all  $m$  provided such inequalities are true.

**Inequalities (5) and (6).** Let  $r = 0$ . Recalling the Bernoulli inequality

$$(1 + c)^s \geq 1 + sc \quad \text{for all } c > -1,$$

one obtains

$$\begin{aligned} \int_0^1 \left(1 - \frac{v}{2s}\right)^{2s-1} \left(1 + \frac{v}{s}\right)^s dv &= \int_0^1 \left[\left(1 - \frac{v}{2s}\right)^2 \left(1 + \frac{v}{s}\right)\right]^s \left(1 - \frac{v}{2s}\right)^{-1} dv \\ &= \int_0^1 \left[1 - \frac{3v^2}{4s^2} + \frac{v^3}{4s^3}\right]^s \left(1 - \frac{v}{2s}\right)^{-1} dv \\ &\geq \int_0^1 \left(1 - \frac{3v^2}{4s} + \frac{v^3}{4s^2}\right) \left(1 - \frac{v}{2s}\right)^{-1} dv \\ &> \int_0^1 \left(1 - \frac{3v^2}{4s} + \frac{v^3}{4s^2}\right) \left(1 + \frac{v}{2s} + \frac{v^2}{4s^2}\right) dv \\ &= 1 + \frac{5}{96s^2} - \frac{1}{80s^3} + \frac{1}{96s^4} > 1. \end{aligned}$$

Here, the first inequality is because of the Bernoulli's one while the second depends on  $(1 - c)^{-1} > 1 + c + c^2$  for all  $c \in (0, 1)$ . Hence, inequality (5) is actually true.

Let us turn to inequality (6). We have to show that  $I_s < 1$ , where

$$I_s = \int_0^1 \left(1 - \frac{v}{2s+1}\right)^{2s} \left(1 + \frac{v}{s-1}\right)^{s-1} dv.$$

First note that

$$\begin{aligned} I_s &= \int_0^1 \frac{2s+1}{2s+1-v} \exp \left( (2s+1) \log(1 - \frac{v}{2s+1}) + (s-1) \log(1 + \frac{v}{s-1}) \right) dv \\ &= \int_0^1 \frac{2s+1}{2s+1-v} \exp \left( \sum_{k=2}^{\infty} \frac{v^k}{k} \left( \frac{(-1)^{k-1}}{(s-1)^{k-1}} - \frac{1}{(2s+1)^{k-1}} \right) \right) dv \\ &< \int_0^1 \frac{2s+1}{2s+1-v} \exp \left( \sum_{k=2}^3 \frac{v^k}{k} \left( \frac{(-1)^{k-1}}{(s-1)^{k-1}} - \frac{1}{(2s+1)^{k-1}} \right) \right) dv \end{aligned}$$

where the last inequality depends on

$$\sum_{k=4}^{\infty} \frac{v^k}{k} \left( \frac{(-1)^{k-1}}{(s-1)^{k-1}} - \frac{1}{(2s+1)^{k-1}} \right) < 0.$$

Since  $\exp(c) < 1 + c + \frac{c^2}{2}$  for  $c < 0$  and

$$\gamma(s, v) := \sum_{k=2}^3 \frac{v^k}{k} \left( \frac{(-1)^{k-1}}{(s-1)^{k-1}} - \frac{1}{(2s+1)^{k-1}} \right) = \frac{-3v^2s}{2(s-1)(2s+1)} + \frac{v^3s(s+2)}{(s-1)^2(2s+1)^2} < 0,$$

one also obtains

$$I_s < \int_0^1 \frac{2s+1}{2s+1-v} \exp(\gamma(s, v)) dv < \int_0^1 \frac{2s+1}{2s+1-v} \left( 1 + \gamma(s, v) + \frac{\gamma(s, v)^2}{2} \right) dv.$$

Moreover, since

$$\begin{aligned} \frac{2s+1}{2s+1-v} &= \frac{1}{1 - \frac{v}{2s+1}} = 1 + \frac{v}{2s+1} + \frac{v^2}{(2s+1)^2} \frac{1}{1 - \frac{v}{2s+1}} \\ &\leq 1 + \frac{v}{2s+1} + \frac{5v^2}{4(2s+1)^2}, \end{aligned}$$

it follows that

$$I_s < \int_0^1 \left( 1 + \frac{v}{2s+1} + \frac{5v^2}{4(2s+1)^2} \right) \left( 1 + \gamma(s, v) + \frac{\gamma(s, v)^2}{2} \right) dv.$$

After some (tedious) algebra, the above integral can be evaluated and the previous inequality can be written as

$$I_s < 1 + \frac{1 + 3s(-11s^3 + (s+4)^2)}{(s-1)^4(2s+1)^6} + \frac{s^6(29 + 7s - 15s^2/2)}{(s-1)^4(2s+1)^6}.$$

Both fractions in the previous expression are negative for  $s \geq 3$ . Hence,  $I_s < 1$  for each  $s \geq 3$ . Finally,  $I_2 < 1$  follows from a direct calculation.

This concludes the proof of inequality (6).

**Inequalities (7) and (8).** Let  $r \in \{1, 2\}$  and

$$J_s^{(r)} = \int_0^1 \left( 1 - \frac{v}{2s+1} \right)^{2s} \left( 1 + \frac{v}{s+r-1} \right)^{s+r-1} dv.$$

Since  $(1 + \frac{v}{s})^s \leq (1 + \frac{v}{s+1})^{s+1}$ , one obtains  $J_s^{(1)} \leq J_s^{(2)}$ . Hence, to prove (7), it suffices to show  $J_s^{(1)} > 1$ . To this end, we first write

$$\begin{aligned} J_s^{(1)} &= \int_0^1 (1 - \frac{v}{2s+1})^{2s} (1 + \frac{v}{s})^s dv \\ &= \int_0^1 [(1 - \frac{v}{2s+1})^2 (1 + \frac{v}{s})]^s dv \\ &= \int_0^1 [1 + \frac{v(1-2v)}{s(2s+1)} + \frac{v^2}{(2s+1)^2} (1 + \frac{v}{s})]^s dv. \end{aligned}$$

Hence, the Bernoulli inequality yields

$$\begin{aligned} J_s^{(1)} &\geq \int_0^1 [1 + \frac{v(1-2v)}{2s+1} + \frac{sv^2}{(2s+1)^2} (1 + \frac{v}{s})] dv \\ &= 1 - \frac{1}{6(2s+1)} + \frac{s}{3(2s+1)^2} + \frac{1}{4(2s+1)^2} = 1 + \frac{1}{12(2s+1)^2}. \end{aligned}$$

This proves inequality (7).

Finally, we turn to (8). Let

$$H_s^{(r)} = \int_0^1 (1 - \frac{v}{2s+2})^{2s+1} (1 + \frac{v}{s+r-2})^{s+r-2} dv.$$

Once again,  $H_s^{(1)} \leq H_s^{(2)}$ . Thus, to prove (8), it suffices to show that  $H_s^{(2)} < 1$ . To this end, we argue as in the proof of  $I_s < 1$ . Precisely, we first note that

$$\begin{aligned} H_s^{(2)} &= \int_0^1 (1 - \frac{v}{2s+2})^{2s+1} (1 + \frac{v}{s})^s dv \\ &= \int_0^1 \frac{2s+2}{2s+2-v} \exp \left( (2s+2) \log(1 - \frac{v}{2s+2}) + s \log(1 + \frac{v}{s}) \right) dv \\ &= \int_0^1 \frac{2s+2}{2s+2-v} \exp \left( \sum_{k=2}^{\infty} \frac{v^k}{k} \left( \frac{(-1)^{k-1}}{s^{k-1}} - \frac{1}{(2s+2)^{k-1}} \right) \right) dv \\ &< \int_0^1 \frac{2s+2}{2s+2-v} \exp \left( \sum_{k=2}^3 \frac{v^k}{k} \left( \frac{(-1)^{k-1}}{s^{k-1}} - \frac{1}{(2s+2)^{k-1}} \right) \right) dv \end{aligned}$$

where the last inequality is because

$$\sum_{k=4}^{\infty} \frac{v^k}{k} \left( \frac{(-1)^{k-1}}{s^{k-1}} - \frac{1}{(2s+2)^{k-1}} \right) < 0.$$

Moreover,

$$\lambda(s, v) := \sum_{k=2}^3 \frac{v^k}{k} \left( \frac{(-1)^{k-1}}{s^{k-1}} - \frac{1}{(2s+2)^{k-1}} \right) < 0,$$

$$\text{and} \quad \frac{2s+2}{2s+2-v} \leq 1 + \frac{v}{2s+2} + \frac{6v^2}{5(2s+2)^2}.$$



Hence, recalling that  $\exp(c) < 1 + c + \frac{c^2}{2}$  for  $c < 0$ , one obtains

$$\begin{aligned} H_s^{(2)} &< \int_0^1 \frac{2s+2}{2s+2-v} \exp(\lambda(s, v)) dv \\ &< \int_0^1 \left(1 + \frac{v}{2s+2} + \frac{6v^2}{5(2s+2)^2}\right) \left(1 + \lambda(s, v) + \frac{\lambda(s, v)^2}{2}\right) dv. \end{aligned}$$

Finally, evaluating the integral, the previous inequality turns into

$$H_s^{(2)} < 1 + \frac{-2s^5 - 9(s^4 + s^3) + 8s^2 + 22s + 18}{64s(s+1)^6} < 1.$$

This proves (8) and concludes the proof of Theorem 2.

**Added in proof:** After writing this paper, we learned (from an anonymous referee) of the existence of another paper very similar to ours, that is: Ping Sun (2021) Strictly unimodality of the probability that the binomial distribution is more than its expectation, *Discrete Applied Mathematics* 301, 1–5. However, we point out that a preliminary draft of our paper appeared on arXiv previous to Sun's paper; see: arXiv:2104.11971v1 [math.PR]

#### REFERENCES

- [1] Doerr B., (2018) An elementary analysis of the probability that a binomial random variable exceeds its expectation, *Statist. Probab. Lett.*, 139, 67-74.
- [2] Esseen C.G., (1945) Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law, *Acta Mathematica*, 77, 1-125.
- [3] Greenberg S. and Mohri M., (2014) Tight lower bound on the probability of a binomial exceeding its expectation, *Statist. Probab. Lett.*, 86, 91-98,
- [4] Janson S., (2021) On the probability that a binomial variable is at most its expectation, *Statist. Probab. Lett.*, 171, 109020.
- [5] Jogdeo K. and Samuels S.M., (1968) Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation, *Annals of Mathematical Statistics*, 39, 1191-1195.
- [6] Pelekis C. and Ramon J., (2016) A lower bound on the probability that a binomial random variable is exceeding its mean, *Statistics and Probability Letters*, 119, 305-309.
- [7] Rigollet P. and Tong Xin, (2011) Neyman-pearson classification, convexity and stochastic constraints, *J. Mach. Learn. Res.*, 12, 2831-2855.
- [8] Slud E.V., (1977) Distribution inequalities for the binomial law, *Annals of Probability*, 5, 404-412.
- [9] Vapnik V.N., (1998) Statistical Learning Theory, Wiley, New York.

LUCIO BARABESI, DIPARTIMENTO DI ECONOMIA POLITICA E STATISTICA, UNIVERSITÀ DI SIENA,  
VIA SAN FRANCESCO 7, 53100 SIENA, ITALY  
Email address: [lucio.barabesi@unisi.it](mailto:lucio.barabesi@unisi.it)

LUCA PRATELLI, ACCADEMIA NAVALE, VIALE ITALIA 72, 57100 LIVORNO, ITALY  
Email address: [pratel@mail.dm.unipi.it](mailto:pratel@mail.dm.unipi.it)

PIETRO RIGO (CORRESPONDING AUTHOR), DIPARTIMENTO DI SCIENZE STATISTICHE "P. FORTU-  
NATI", UNIVERSITÀ DI BOLOGNA, VIA DELLE BELLE ARTI 41, 40126 BOLOGNA, ITALY  
Email address: [pietro.rigo@unibo.it](mailto:pietro.rigo@unibo.it)