On the monotonicity of tail probabilities

Fokkink, Robbert; Papavassiliou, Symeon; Pelekis, Christos

DOI
10.37190/0208-4147.00050

Publication date
2022

Document Version
Final published version

Published in
Probability and Mathematical Statistics

Citation (APA)
Fokkink, R., Papavassiliou, S., & Pelekis, C. (2022). On the monotonicity of tail probabilities. *Probability and Mathematical Statistics*, *42*(1), 133-141. https://doi.org/10.37190/0208-4147.00050

PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 42, Fasc. 1 (2022), pp. 133–141
Published online 6.6.2022
doi:10.37190/0208-4147.00050

ON THE MONOTONICITY OF TAIL PROBABILITIES*

BY

ROBBERT FOKKINK (DELFT), SYMEON PAPAVASSILIOU (ATHENS), AND
CHRISTOS PELEKIS (ATHENS)

Abstract. Let S and X be independent random variables, assuming values in the set of non-negative integers, and suppose further that both $\mathbb{E}(S)$ and $\mathbb{E}(X)$ are integers satisfying $\mathbb{E}(S) \geqslant \mathbb{E}(X)$. We establish a sufficient condition for the tail probability $\mathbb{P}(S \geqslant \mathbb{E}(S))$ to be larger than the tail $\mathbb{P}(S + X \geqslant \mathbb{E}(S + X))$, when the mean of S is equal to the mode.

2020 Mathematics Subject Classification: Primary 60G50; Secondary 60E15.

Key words and phrases: tail comparisons, sums of independent random variables, (negative) binomial distribution, Poisson distribution, Simmons' inequality.

1. MAIN RESULT

We are interested in the comparison between the tails $\mathbb{P}(S \geqslant \mathbb{E}(S))$ and $\mathbb{P}(S+X \geqslant \mathbb{E}(S + X))$, where S and X are independent random variables. In everyday language, suppose an enterprise S is successful if the result exceeds the mean; would it be beneficial to include one more enterprise X? In many applications, S is a sum of independent random variables and X adds one more to the sum. By the central limit theorem, $\mathbb{P}(S \geqslant \mathbb{E}(S))$ converges to $1/2$. Therefore, if $\mathbb{P}(S \geqslant \mathbb{E}(S)) > 1/2$ (the enterprise is favorably skewed), one would expect that adding one more term to the sum would lower this probability.

All random variables under consideration take values in $\mathbb{N} \cup \{0\}$. We establish an inequality that applies to random variables that satisfy certain "skewness" conditions. Throughout the text, given a positive integer n, we denote the set $\{1, \ldots, n\}$ by $[n]$.

* Research was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: HFRI-FM17-2436).

DEFINITION 1.1 (Right-skewness). Assume that S is unimodal with mode s. Then we say that S is *right-skewed* if

$$\mathbb{P}(S = s - i) \leqslant \mathbb{P}(S = s + i - 1) \quad \text{for all } i \in [s].$$

In our definition, we allow that the mode is not unique. It is possible that $\mathbb{P}(S = s - 1) = \mathbb{P}(S = s)$ and that is why we put the $\leqslant$ sign. If the inequality is strict, then the inequality in our main result is also strict.

DEFINITION 1.2 (Left-loadedness). Let X be a random variable such that $m := \mathbb{E}(X)$ is an *integer*. For $i \in [m]$, set $\alpha_i := \mathbb{P}(X \leqslant m - i) - \mathbb{P}(X \geqslant m + i)$. Then we say that X is *left-loaded* if either of the following two conditions holds true:

(L_1): The sequence $\{\alpha_i\}_{i=1}^{m}$ changes sign once from positive to negative, i.e., there exists $\ell \in [m]$ such that $\alpha_i \geqslant 0$ for $i \leqslant \ell$, and $\alpha_i \leqslant 0$ for $i > \ell$.

(L_2): $\sum_{i=1}^{k} \alpha_i \geqslant 0$ for all $k \in [m]$.

A random variable can be both right-skewed and left-loaded. For instance, if $\mathbb{E}(S) = 1$ then it is not hard to prove that S is left-loaded. If such an S is unimodal, such as the binomial distribution $Bin(n, 1/n)$, then it is also right-skewed. Another example is a geometric random variable with parameter $1/n$. Our main result reads as follows.

THEOREM 1.1. *Let $s \geqslant m$ be two positive integers. Suppose that S and X are* independent *random variables, assuming values in the set of non-negative integers, that satisfy the following conditions:*

- *S is right-skewed with mode s.*
- *X is left-loaded with mean m.*

Then $\mathbb{P}(S \geqslant s) \geqslant \mathbb{P}(S + X \geqslant s + m)$.

Note that we have replaced the mean of S by its mode. If S is binomial or Poisson with integer mean, then the mean is equal to the mode. We will show that Poisson random variables with integer mean are both right-skewed and left-loaded, and that binomial random variables are right-skewed if $p \leqslant 1/2$. We conjecture that a binomial random variable is left-loaded if it has integer mean and $p \leqslant 1/2$. This seems to be hard to prove and is related to an old inequality of Simmons [6].

Our inequality is well-established for standard random variables. Let $Poi(\lambda)$ denote a Poisson random variable of mean λ. Teicher [7] showed that

$$\mathbb{P}(Poi(k) \geqslant k) \geqslant \mathbb{P}(Poi(k+1) \geqslant k+1) \quad \text{for all } k \geqslant 1, \tag{1.1}$$

which follows from our result if we take $S \sim Poi(k)$ and $X \sim Poi(1)$. Let $Bin(m, p)$ denote a binomial random variable of parameters m and $p \in (0, 1)$.

Chaundy and Bullard [1] showed that for every fixed positive integer $n \geqslant 1$ and probability $p = 1/n$,

$$\mathbb{P}(Bin(nk,p) \geqslant k) \geqslant \mathbb{P}(Bin(n(k+1),p) \geqslant k+1) \quad \text{for all } k \geqslant 1. \tag{1.2}$$

This follows from our result if we take $S \sim Bin(nk,p)$ and $X \sim Bin(n,p)$ for $p = 1/n$. We remark that both inequalities (1.1) and (1.2) concern the monotonicity of tail probabilities of the form $\mathbb{P}(S_k \geqslant \mathbb{E}(S_k))$, where S_k is a sum of k independent random variables of mean 1. These results have been extended to the case of integer means (see [3, Theorem 2.1] and [4, Theorem 2.3]), and several of those extensions can be deduced from our main result. However, Theorem 1.1 provides a bit more, since it allows one to convolute different distributions. For example, it follows from the results in Section 3 that Theorem 1.1 implies that $\mathbb{P}(S \geqslant s) \geqslant \mathbb{P}(S + X \geqslant \mathbb{E}(S + X))$ for $S \sim Bin(n, s/n)$ with $n \geqslant 2s$, and $X \sim Poi(m)$ with $s \geqslant m$, a result which may be seen as a "mixture" of (1.1) and (1.2).

The tail probability $\mathbb{P}(S \geqslant \mathbb{E}(S))$ has been extensively studied for Poisson random variables, motivated by a conjecture by Ramanujan that was eventually settled by Flajolet. This research is ongoing and results continue to be sharpened and extended; see [2] for recent progress and further references. It is not possible to deduce such refined results for parametrized families from our inequality, which puts relatively weak constraints on the distributions of S and X.

2. PROOF OF MAIN RESULT

We begin with an observation.

LEMMA 2.1. *Let X be a random variable, assuming non-negative integer values, such that $m := \mathbb{E}(X)$ is an* integer. *Then*

$$\sum_{i=1}^{m} \big(\mathbb{P}(X \leqslant m-i) - \mathbb{P}(X \geqslant m+i)\big) = \sum_{i \geqslant m+1} \mathbb{P}(X \geqslant m+i).$$

In particular, $\sum_{i=1}^{m} \big(\mathbb{P}(X \leqslant m-i) - \mathbb{P}(X \geqslant m+i)\big) \geqslant 0$.

Proof. Notice that

$$m = \sum_{i=1}^{m} \mathbb{P}(X \geqslant i) + \sum_{i=m+1}^{2m} \mathbb{P}(X \geqslant i) + \sum_{i \geqslant 2m+1} \mathbb{P}(X \geqslant i),$$

which, upon transferring the first two sums on the right to the other side, is equivalent to

$$\sum_{i=1}^{m} (\mathbb{P}(X \leqslant m-i) - \mathbb{P}(X \geqslant m+i)) = \sum_{i \geqslant m+1} \mathbb{P}(X \geqslant m+i). \quad \blacksquare$$

We now prove our main result, which applies to random variables that are skewed to the right. One would expect that there exists a corresponding result for variables that are skewed to the left. However, our proof does not easily transfer to this case. One problem is that the inequality $\sum_{i=1}^{m}\big(\mathbb{P}(X \leqslant m-i) - \mathbb{P}(X \geqslant m+i)\big) \geqslant 0$ holds for all random variables. It does not change sign if we skew the random variable to the left.

Proof of Theorem 1.1. If we condition on S we have

$$\begin{aligned}\mathbb{P}(S+X \geqslant s+m) &= \sum_{i\geqslant 0} \mathbb{P}(X \geqslant s+m-i)\cdot \mathbb{P}(S=i)\\ &= \mathbb{P}(S \geqslant s+m) + \sum_{i=0}^{s+m-1} \mathbb{P}(X \geqslant s+m-i)\cdot \mathbb{P}(S=i).\end{aligned}$$

Hence $\mathbb{P}(S+X \geqslant s+m) \leqslant \mathbb{P}(S \geqslant s)$ is equivalent to

$$\sum_{i=0}^{s+m-1} \mathbb{P}(X \geqslant s+m-i)\cdot \mathbb{P}(S=i) \leqslant \sum_{i=s}^{s+m-1} \mathbb{P}(S=i),$$

which can be rearranged as

$$\sum_{i=0}^{s-1} \mathbb{P}(S=i)\cdot \mathbb{P}(X \geqslant s+m-i) \leqslant \sum_{i=s}^{s+m-1} \mathbb{P}(S=i)\cdot \mathbb{P}(X \leqslant s+m-i-1).$$

This is equivalent to

$$\sum_{i=1}^{s} \mathbb{P}(S=s-i)\cdot \mathbb{P}(X \geqslant m+i) \leqslant \sum_{i=1}^{m} \mathbb{P}(S=s+i-1)\cdot \mathbb{P}(X \leqslant m-i). \tag{2.1}$$

Let L and R denote the left-hand side and the right-hand side of (2.1). Since S is unimodal with mode $s \geqslant m$, we can estimate L as follows:

$$\begin{aligned}L &\leqslant \sum_{i=1}^{m} \mathbb{P}(S=s-i)\cdot \mathbb{P}(X \geqslant m+i)\\ &\quad + \mathbb{P}(S=s-m-1)\cdot \sum_{i=m+1}^{s} \mathbb{P}(X \geqslant m+i)\\ &=: \ell_1 + \ell_2,\end{aligned}$$

with the convention that ℓ_2 is equal to 0 when $s=m$. Now, since S is right-skewed, we have

$$\ell_1 \leqslant \sum_{i=1}^{m} \mathbb{P}(S=s+i-1)\cdot \mathbb{P}(X \geqslant m+i) =: R_1. \tag{2.2}$$

Using again the right-skewness of S and Lemma 2.1, we have

$$\ell_2 \leqslant \mathbb{P}(S = s + m) \cdot \Big(\sum_{i=1}^{m} (\mathbb{P}(X \leqslant m - i) - \mathbb{P}(X \geqslant m + i))\Big) =: R_2. \tag{2.3}$$

It follows from (2.1)–(2.3) that it is enough to show that $R_1 + R_2 \leqslant R$, or equivalently

$$\sum_{i=1}^{m} \big(\mathbb{P}(S = s+i-1) - \mathbb{P}(S = s+m)\big) \cdot \big(\mathbb{P}(X \leqslant m-i) - \mathbb{P}(X \geqslant m+i)\big) \geqslant 0. \tag{2.4}$$

For each $i \in [m]$, let $\Delta_i := \mathbb{P}(S = s + i - 1) - \mathbb{P}(S = s + m)$ as well as $\alpha_i := \mathbb{P}(X \leqslant m - i) - \mathbb{P}(X \geqslant m + i)$, and note that (2.4) is equivalent to

$$\sum_{i=1}^{m} \Delta_i \cdot \alpha_i \geqslant 0. \tag{2.5}$$

The unimodality of S implies that $\Delta_1 \geqslant \cdots \geqslant \Delta_m \geqslant 0$. We distinguish two cases.

Suppose first that X satisfies condition (L_1). Let $\ell \in [m]$ be such that $\alpha_i \geqslant 0$ for $i \leqslant \ell$, and $\alpha_i \leqslant 0$ for $i > \ell$. Then, since $\{\Delta_i\}_{i\in[m]}$ is non-increasing, it follows that

$$\sum_{i=1}^{m} \Delta_i \cdot \alpha_i \geqslant \Delta_\ell \sum_{i=1}^{\ell} \alpha_i + \Delta_\ell \sum_{i=\ell+1}^{m} \alpha_i = \Delta_\ell \sum_{i\in[m]} \alpha_i \geqslant 0,$$

where the last estimate follows from the second statement in Lemma 2.1. Hence we obtain (2.5) and the result follows.

Now assume that X satisfies condition (L_2). Set $\Sigma_i := \sum_{j=1}^{i} \alpha_j$ for $i \in [m]$, and notice that $\Sigma_i \geqslant 0$ by assumption. Using summation by parts, we have

$$\sum_{i=1}^{m} \Delta_i \cdot \alpha_i = \Delta_m \cdot \Sigma_m + \sum_{i=1}^{m-1} (\Delta_i - \Delta_{i+1}) \cdot \Sigma_i \geqslant 0.$$

Hence, we obtain (2.5) and the result follows. ■

3. SKEWNESS OF RANDOM VARIABLES

The standard examples of non-negative random variables that take values in $\mathbb{N} \cup \{0\}$ are Poisson, binomial, or negative binomial. We examine their "skewness" properties.

LEMMA 3.1. *Fix a positive integer s, and let $S \sim Poi(s)$. Then S is right-skewed.*

Proof. Since s is a positive integer it follows that the mode of S is equal to s. For $i \in [s]$, let $\beta_i = \frac{\mathbb{P}(S=s-i)}{\mathbb{P}(S=s+i-1)}$. Since the mode of S is equal to s, it follows that

$\beta_1 \leqslant 1$. Next, note that $\beta_i \geqslant \beta_{i+1}$ is equivalent to $s^2 \geqslant s^2 - i^2$, which is clearly correct for each $i \in [s]$. Hence, the sequence $\{\beta_i\}_{i=1}^s$ is non-increasing, and the fact that $\beta_1 \leqslant 1$ finishes the proof. ∎

LEMMA 3.2. *Fix a positive integer s, and let $S \sim Bin(n,p)$ for some $n \geqslant 2s$ with $p = s/n$. Then S is right-skewed.*

Proof. The proof is similar to the proof of Lemma 3.1. Let $\beta_i = \frac{\mathbb{P}(S=s-i)}{\mathbb{P}(S=s+i-1)}$ for $i \in [s]$. Since S is unimodal with mode s, we have $\beta_1 \leqslant 1$. Furthermore, $\beta_i \geqslant \beta_{i+1}$ is equivalent to

$$s^2 \cdot ((n-s+1)^2 - i^2) \geqslant (n-s)^2 \cdot (s^2 - i^2). \tag{3.1}$$

Now observe that (3.1) holds true when $s^2 \cdot ((n-s)^2 - i^2) \geqslant (n-s)^2 \cdot (s^2 - i^2)$ and the latter is equivalent to $n - s \geqslant s$, which is true by assumption. Hence (3.1) holds true and we conclude that the sequence $\{\beta_i\}_{i\in[s]}$ is non-decreasing. The result follows. ∎

We denote the negative binomial distribution by $NB(r,p)$ where $r \in \mathbb{N}$ is the number of failures and $p \in (0,1)$ is the probability of success. If $S \sim NB(r,p)$ then $\mathbb{P}(S = k) = \binom{k+r-1}{r-1} p^k q^r$ with $q = 1 - p$ the probability of failure. If $q = 1/n$, the negative binomial has mean $r(n-1)$ and mode $(r-1)(n-1)$.

LEMMA 3.3. *Let $S \sim NB(r,p)$ with $p = 1 - 1/n$ for some integer $n > 1$. Then S is right-skewed.*

Proof. Let $a_k = \mathbb{P}(S = k)$. Then

$$\frac{a_{k+1}}{a_k} = \frac{(k+r)p}{k+1}$$

is $\leqslant 1$ if and only if $k + 1 \geqslant p(r-1)/q$. In particular, S is unimodal with mode $\lfloor p(r-1)/q \rfloor$, which is equal to $s = (n-1)(r-1)$ for our choice of p. To prove that S is right-skewed, it suffices to show that $\frac{a_{s-i-1}}{a_{s-i}} \leqslant \frac{a_{s+i}}{a_{s+i-1}}$, in other words,

$$\frac{s-i}{(s+r-1-i)p} \leqslant \frac{(s+r-1+i)p}{s+i}.$$

For our choice of p, this is equivalent to

$$\frac{s-i}{s-ip} \leqslant \frac{s+ip}{s+i},$$

which obviously holds true. ∎

We have thus established the right-skewness of standard non-negative discrete random variables for certain parameters. Left-loadedness is more difficult to verify. We will prove that a Poisson random variable with integer mean is left-loaded.

Simmons [6] proved that a binomial random variable X with integer mean m satisfies $\mathbb{P}(X \leqslant m-1) > \mathbb{P}(X \geqslant m+1)$ if $n > 2m$. This has been generalized to other distributions by Perrin and Redside [5, Proposition 3.3].

LEMMA 3.4. *Let X be a random variable with integer mean m. Then*

$$\mathbb{P}(X \leqslant m-1) > \mathbb{P}(X \geqslant m+1)$$

if X is Poisson.

LEMMA 3.5. *Fix a positive integer $m \geqslant 3$, and let $X \sim Poi(m)$. Then*

$$\mathbb{P}(X \geqslant 2m) > \mathbb{P}(X = 0).$$

Proof. It is enough to show that $\mathbb{P}(X = 2m) > \mathbb{P}(X = 0)$, or equivalently that $m^{2m} > (2m)!$. This holds if $m = 3$ and we proceed by induction:

$$\begin{aligned}(m+1)^{2(m+1)} &= \left(\frac{m+1}{m}\right)^{2m} \cdot (m+1)^2 \cdot m^{2m} \\ &> 4(m+1)^2 \cdot (2m)! > (2(m+1))!. \quad \blacksquare\end{aligned}$$

A sequence $\{a_i\}_{i=1}^m$ of real numbers is said to be *U-shaped* if there exists $\ell \in [m]$ such that $a_1 \geqslant \cdots \geqslant a_\ell$ and $a_\ell \leqslant \cdots \leqslant a_m$.

LEMMA 3.6. *Let $m \geqslant 3$ be an integer, and let $X \sim Poi(m)$. Then X is left-loaded.*

Proof. We show that X satisfies condition (L_1). Recall that $\alpha_i = \mathbb{P}(X \leqslant m-i) - \mathbb{P}(X \geqslant m+i)$. We have to show that $\{\alpha_i\}_{i=1}^m$ changes sign once. Lemma 3.4 implies that $\alpha_1 > 0$ and Lemma 3.5 implies that $\alpha_m \leqslant 0$, and it suffices to show that the sequence $\{\alpha_i\}_{i=1}^m$ is U-shaped. Since for every $i \in [m-1]$ we have

$$\alpha_{i+1} = \alpha_i - \mathbb{P}(X = m-i) + \mathbb{P}(X = m+i),$$

it is enough to show that the sequence $\{b_i\}_{i=1}^m$, where $b_i := \mathbb{P}(X = m-i) - \mathbb{P}(X = m+i)$, changes sign once. To this end, for $i \in [m]$, let

$$\beta_i = \frac{\mathbb{P}(X = m+i)}{\mathbb{P}(X = m-i)}.$$

Then $\beta_i \geqslant \beta_{i+1}$ is equivalent to $i^2 + i \leqslant m$. Since the sequence $\{i^2 + i\}_{i=1}^m$ is increasing, it follows that the sequence $\{\beta_i\}_{i=1}^m$ is U-shaped. Now note that $\beta_1 < 1$, and the proof of Lemma 3.5 implies that $\beta_m \geqslant 1$. Since $\{\beta_i\}_{i=1}^m$ is U-shaped, there exists a unique $k \in [m]$ such that $\beta_i < 1$ for $i \leqslant k$, and $\beta_i \geqslant 1$ for $i \geqslant k+1$, which in turn yields $b_i > 0$ for $i \leqslant k$, and $b_i \leqslant 0$ for $i \geqslant k+1$. In other words, the sequence $\{b_i\}_{i=1}^m$ changes sign once, as desired. $\blacksquare$

LEMMA 3.7. *Let $X \sim Poi(m)$ for a natural number m. Then X is left-loaded.*

Proof. We need to verify the remaining two cases of $m = 1$ and $m = 2$. If $m = 1$, then the second statement in Lemma 2.1 implies that X satisfies condition (L_2). If $m = 2$, then Lemma 3.4 and the second statement in Lemma 2.1 imply that X satisfies condition (L_2). If $m \geqslant 3$ then Lemma 3.6 implies that X satisfies condition (L_1). The result follows. ∎

In a similar way, one can show that a $Bin(n, m/n)$ random variable is left-loaded for a certain range of parameters. More precisely, it satisfies condition (L_2) when $m \in \{1, 2\}$, and condition (L_1) when $4 \leqslant m \leqslant n/3$, but numerical experiments suggest that it is left-loaded for $m \leqslant n/2$ (see the conjecture below). The same appears to be true for a negative binomial distribution with parameter $p = 1 - 1/n$.

4. CONCLUDING REMARKS

We expect that a binomial random variable is left-loaded if $p \leqslant 1/2$. More specifically, we conjecture the following.

CONJECTURE 4.1. *Fix positive integers n, m such that $n \geqslant 2m$, and let $X \sim Bin(n, m/n)$. Then X is left-loaded.*

Condition (L_2) says that $\sum_{i=1}^{k} \alpha_i \geqslant 0$ for all $1 \leqslant k \leqslant m$. Note that our conjecture extends Simmons' inequality (see [6] and [5]).

We have established the right-skewness of random variables for a limited set of parameter values. It is likely that this parameter range can be considerably extended.

The main restriction on our result is that $\mathbb{E}(X)$ is an integer. This is used in Lemma 2.1, which is just a rearrangement of terms. To extend our result to X with non-integer mean, one needs to find a way around this lemma.

REFERENCES

[1] T. W. Chaundy and J. E. Bullard, *John Smith's problem*, Math. Gazette 44 (1960), 253–260.
[2] D. Dmitriev and M. Zhukovskii, *On monotonicity of Ramanujan function for binomial random variables,* Statist. Probab. Lett. 177 (2021), 109–147.
[3] K. Jogdeo and S. M. Samuels, *Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation*, Ann. Math. Statist. 39 (1968), 1191–1195.
[4] J. M. Kane, *Monotonic approach to central limits*, Proc. Amer. Math. Soc. 129 (2000), 2127–2133.
[5] O. Perrin and E. Redside, *Generalization of Simmons' theorem*, Statist. Probab. Lett. 77 (2007), 604–606.
[6] T. C. Simmons, *A new theorem in probability*, Proc. London Math. Soc. 1 (1894), 290–325.
[7] H. Teicher, *An inequality on Poisson probabilities*, Ann. Math. Statist. 26 (1955), 147–149.

Robbert Fokkink
Institute of Applied Mathematics
Delft University of Technology
Mourikbroekmanweg 6
2628 XE Delft, The Netherlands
E-mail: r.j.fokkink@tudelft.nl

Symeon Papavassiliou
School of Electrical and Computer Engineering
National Technical University of Athens
Zografou, Athens, Greece, 15780
E-mail: papavass@mail.ntua.gr

Christos Pelekis
School of Electrical and Computer Engineering
National Technical University of Athens
Zografou, Athens, Greece, 15780
E-mail: pelekis.chr@gmail.com

Received 20.11.2021;
accepted 8.4.2022