

Article

A Simple Wide Range Approximation of Symmetric Binomial Distribution

Tamás Szabados 

Department of Stochastics, Institute of Mathematics, Budapest University of Technology and Economics,
Muegyetem rkp. 3, H ep, 5 em, 1521 Budapest, Hungary; szabados@math.bme.hu; Tel.: +36-30-390-0660

Abstract: The paper gives a wide range, uniform, local approximation of symmetric binomial distribution. The result clearly shows how one has to modify the classical de Moivre–Laplace normal approximation in order to give an estimate at the tail as well as to minimize the relative error.

Keywords: wide range; approximation; binomial distribution

MSC: 60E15; 60F05

The topic of this paper is a wide range, uniform, local approximation of symmetric binomial distribution, an extension of the classical de Moivre–Laplace theorem in the symmetric case [1] (Ch. VII) [2]. In other words, I would like to approximate individual binomial probabilities not only in a classical neighborhood of the center, but at the tail as well. The result clearly shows how one has to modify the normal approximation in order to give a wide range estimate to minimize the relative error. The method will be somewhat similar to the ones applied by [3–5] or in the proof of Tusnády’s lemma, see, e.g., [6]; however, my task is much simpler than those. This simplification makes the proof short, transparent and natural. Moreover, the result is non-asymptotic, that is, it gives explicit, nearly optimal, upper and lower bounds for the relative error with a finite n . Thus, I hope that it can be used in both applications and teaching.

Let $(X_r)_{r \geq 1}$ be a sequence of independent, identically distributed random steps with $\mathbb{P}(X_r = \pm 1) = \frac{1}{2}$ and $S_\ell = \sum_{r=1}^\ell X_r$ ($\ell \geq 1$), $S_0 = 0$, be the corresponding simple, symmetric random walk. Then

$$\mathbb{P}(S_\ell = j) = \binom{\ell}{\frac{\ell+j}{2}} 2^{-\ell} \quad (|j| \leq \ell). \quad (1)$$

Here, we use the convention that the above binomial coefficient is zero whenever $\ell + j$ is not divisible by 2. Since $\mathbb{P}(S_\ell = j) = \mathbb{P}(S_\ell = -j)$ for any j , it is enough to consider only the case $j \geq 0$ whenever it is convenient.

First we consider the case when $\ell = 2n$, even. Let us introduce the notation

$$a_{k,n} := \mathbb{P}(S_{2n} = 2k) = \binom{2n}{n+k} 2^{-2n} \quad (|k| \leq n).$$

Also, introduce the notation

$$b_{k,n} := n \left\{ \left(1 + \frac{k + \frac{1}{2}}{n} \right) \log \left(1 + \frac{k}{n} \right) + \left(1 - \frac{k - \frac{1}{2}}{n} \right) \log \left(1 - \frac{k}{n} \right) \right\} \quad (2)$$



Academic Editors: Antónia Földes
and Endre Csáki

Received: 1 December 2024

Revised: 13 December 2024

Accepted: 15 December 2024

Published: 30 December 2024

Citation: Szabados, T. A Simple Wide Range Approximation of Symmetric Binomial Distribution. *Entropy* **2025**, *27*, 21. <https://doi.org/10.3390/e27010021>

Copyright: © 2024 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

when $n \geq 1$ and $|k| < n$, and

$$b_{\pm n,n} := \left(2n + \frac{1}{2}\right) \log 2 - \frac{1}{2} \log(2\pi n) \quad (n \geq 1). \quad (3)$$

Theorem 1. (a) For any $n \geq 1$ and $|k| \leq n$, we have

$$a_{k,n} \leq \frac{1}{\sqrt{\pi n}} e^{-b_{k,n}}. \quad (4)$$

(b) For any $n \geq 1$ and $|k| \leq rn$, $r \in (0, 1)$, we have

$$a_{k,n} > \frac{1}{\sqrt{\pi n}} e^{-b_{k,n}} \exp\left(-\frac{1}{7n} - \frac{r^4}{3(1-r^2)^2 n}\right). \quad (5)$$

(c) In accordance with the classical de Moivre–Laplace normal approximation, for $n \geq 3$ and $|k| \leq n^{\frac{2}{3}}$ one has

$$1 - 2n^{-\frac{1}{3}} < \frac{a_{k,n}}{\frac{1}{\sqrt{\pi n}} e^{-\frac{k^2}{n}}} < 1 + 2n^{-\frac{1}{3}}. \quad (6)$$

Proof. As usual, the first step is to estimate the central term

$$a_{0,n} = \binom{2n}{n} 2^{-2n} = \frac{(2n)!}{(n!)^2} 2^{-2n}.$$

By Stirling’s formula, see, e.g., [1] (p. 54), we have:

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n} \left(\frac{n}{e}\right)^n e^{\frac{1}{12n}} \quad (n \geq 1). \quad (7)$$

Thus, after simplification we obtain

$$\frac{1}{\sqrt{\pi n}} e^{-\frac{1}{7n}} < a_{0,n} < \frac{1}{\sqrt{\pi n}} e^{-\frac{1}{9n}} \quad (n \geq 1). \quad (8)$$

Second, also by the standard way, for $1 \leq k \leq n$,

$$a_{k,n} = a_{0,n} \frac{n(n-1) \cdots (n-k+1)}{(n+1)(n+2) \cdots (n+k)} = a_{0,n} \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{\left(1 + \frac{1}{n}\right) \left(1 + \frac{2}{n}\right) \cdots \left(1 + \frac{k}{n}\right)}.$$

So it follows that

$$\begin{aligned} \log a_{k,n} &= \log a_{0,n} - \log \left(1 + \frac{k}{n}\right) - 2 \sum_{j=1}^{k-1} \frac{1}{2} \log \frac{1 + \frac{j}{n}}{1 - \frac{j}{n}} \\ &= \log a_{0,n} - \log \left(1 + \frac{k}{n}\right) - 2 \sum_{j=1}^{k-1} \tanh^{-1} \left(\frac{j}{n}\right). \end{aligned} \quad (9)$$

To approximate the sum here, let us introduce the integral

$$I(x) := \int_0^x \tanh^{-1}(t) dt = \frac{1}{2} (1+x) \log(1+x) + \frac{1}{2} (1-x) \log(1-x) \quad (10)$$

for $|x| < 1$, and its approximation by a trapezoidal sum

$$T_{k,n} := \frac{1}{n} \left\{ \frac{1}{2} \tanh^{-1}(0) + \sum_{j=1}^{k-1} \tanh^{-1} \left(\frac{j}{n}\right) + \frac{1}{2} \tanh^{-1} \left(\frac{k}{n}\right) \right\}, \quad (11)$$

where $0 \leq k < n$. It is well-known that the error of the trapezoidal formula for a function $f \in C^2([a, b])$ is

$$T_n(f) - \int_a^b f(t) dt = \frac{(b-a)^3}{12n^2} f''(x), \quad x \in [a, b].$$

Since $(\tanh^{-1})''(x) = 2x(1-x^2)^{-2}$, we obtain that

$$0 \leq T_{k,n} - I\left(\frac{k}{n}\right) \leq \frac{r^4}{6(1-r^2)^2 n^2}, \quad \text{when } 0 \leq k \leq rn, \quad r \in (0, 1). \quad (12)$$

Let us combine Formulas (8)–(12):

$$\log a_{k,n} = \log a_{0,n} - 2nT_{k,n} - \frac{1}{2} \log\left(1 + \frac{k}{n}\right) - \frac{1}{2} \log\left(1 - \frac{k}{n}\right),$$

thus

$$\log \frac{1}{\sqrt{\pi n}} - b_{k,n} - \frac{1}{7n} - \frac{r^4}{3(1-r^2)^2 n} < \log a_{k,n} < \log \frac{1}{\sqrt{\pi n}} - b_{k,n} - \frac{1}{9n}, \quad (13)$$

where $0 \leq k \leq rn$ and

$$b_{k,n} = 2nI\left(\frac{k}{n}\right) + \frac{1}{2} \log\left(1 - \frac{k^2}{n^2}\right). \quad (14)$$

Clearly, (14) is the same as (2). Thus, (13) proves (a) and (b) of the theorem.

Let us see now, using Taylor expansions, a series expansion of $b_{k,n}$ when $|k| < n$. First, for $|x| < 1$,

$$I(x) = \frac{x^2}{1 \cdot 2} + \frac{x^4}{3 \cdot 4} + \frac{x^6}{5 \cdot 6} + \frac{x^8}{7 \cdot 8} + \dots$$

Second, also for $|x| < 1$,

$$\log(1-x^2) = -x^2 - \frac{x^4}{2} - \frac{x^6}{3} - \frac{x^8}{4} - \dots$$

So by (14), for $|k| < n$ we have a convergent series for $b_{k,n}$:

$$\begin{aligned} b_{k,n} &= \frac{k^2}{n} \left(1 - \frac{1}{2n}\right) + \frac{k^4}{2n^3} \left(\frac{1}{3} - \frac{1}{2n}\right) + \frac{k^6}{3n^5} \left(\frac{1}{5} - \frac{1}{2n}\right) + \dots \\ &= \sum_{j=1}^{\infty} \left(\frac{k}{n}\right)^{2j} \frac{1}{j} \left(\frac{n}{2j-1} - \frac{1}{2}\right). \end{aligned} \quad (15)$$

By (15), for $n \geq 3$ and $|k| \leq n^{\frac{2}{3}}$ we obtain that

$$\begin{aligned} \left|b_{k,n} - \frac{k^2}{n}\right| &\leq \frac{n^{-\frac{2}{3}}}{2} + \sum_{j=2}^{\infty} n^{-\frac{2}{3}j} \frac{1}{j} \left(\frac{n}{2j-1} + \frac{1}{2}\right) \\ &\leq \frac{n^{-\frac{2}{3}}}{2} + \frac{n}{2} \left(\frac{1}{3} + \frac{1}{6}\right) \sum_{j=2}^{\infty} n^{-\frac{2}{3}j} \leq n^{-\frac{1}{3}}. \end{aligned}$$

Thus, by (4), for $n \geq 3$ and $|k| \leq n^{\frac{2}{3}}$,

$$a_{k,n} \leq \frac{1}{\sqrt{\pi n}} e^{-\frac{k^2}{n}} e^{n^{-1/3}} < \frac{1}{\sqrt{\pi n}} e^{-\frac{k^2}{n}} \left(1 + 2n^{-\frac{1}{3}}\right). \quad (16)$$

Similarly, by (5), for $n \geq 3$ and $|k| \leq n^{\frac{2}{3}}$,

$$\begin{aligned} a_{k,n} &> \frac{1}{\sqrt{\pi n}} e^{-\frac{k^2}{n}} \exp\left(-\frac{1}{7n} - \frac{n^{-\frac{7}{3}}}{3(1-n^{-\frac{2}{3}})^2} - n^{-\frac{1}{3}}\right) \\ &\geq \frac{1}{\sqrt{\pi n}} e^{-\frac{k^2}{n}} e^{-1.21n^{-1/3}} > \frac{1}{\sqrt{\pi n}} e^{-\frac{k^2}{n}} \left(1 - 2n^{-1/3}\right). \end{aligned} \quad (17)$$

(16) and (17) together prove (c) of the theorem. $\square$

The main moral of Theorem 1 is that the exponent k^2/n in the exponent of the normal approximation is only a first approximation of the series (15). In (c), the bound $|k| \leq n^{\frac{2}{3}}$ was chosen somewhat arbitrarily. It is clear from the series (15) that the bound should be $o(n^{\frac{3}{4}})$. The above given bound was picked because it seemed to be satisfactory for usual applications with large deviation and gave a nice relative error bound $2n^{-\frac{1}{3}}$.

It is not difficult to extend the previous results to the odd-valued case of the symmetric binomial probabilities

$$a_{k,n}^* := \mathbb{P}(S_{2n-1} = 2k-1) = \binom{2n-1}{n+k-1} 2^{-2n+1} \quad (-n+1 \leq k \leq n).$$

Define

$$b_{k,n}^* := n \left\{ \left(1 + \frac{k - \frac{1}{2}}{n}\right) \log\left(1 + \frac{k}{n}\right) + \left(1 - \frac{k - \frac{1}{2}}{n}\right) \log\left(1 - \frac{k}{n}\right) \right\} \quad (18)$$

for $n \geq 1$ and $|k| < n$, and

$$b_{n,n}^* := \left(2n - \frac{1}{2}\right) \log 2 - \frac{1}{2} \log(2\pi n) \quad (n \geq 1). \quad (19)$$

Theorem 2. (a) For any $n \geq 1$ and $-n+1 \leq k \leq n$, we have

$$a_{k,n}^* < \frac{1}{\sqrt{\pi n}} e^{-b_{k,n}^*} \exp\left(\frac{2}{3n}\right). \quad (20)$$

(b) For any $n \geq 1$ and $|k| \leq rn$, $r \in (0, 1)$, we have

$$a_{k,n}^* > \frac{1}{\sqrt{\pi n}} e^{-b_{k,n}^*} \exp\left(-\frac{1}{n} - \frac{r^4}{3(1-r^2)^2 n}\right). \quad (21)$$

(c) In accordance with the classical de Moivre–Laplace normal approximation, for $n \geq 4$ and $|k| \leq n^{\frac{2}{3}}$ one has

$$1 - 3n^{-\frac{1}{3}} < \frac{a_{k,n}^*}{\frac{1}{\sqrt{\pi n}} e^{-\frac{k^2}{n}}} < 1 + 6n^{-\frac{1}{3}}. \quad (22)$$

Proof. Since this proof is very similar to the previous one, several details are omitted. First, by Stirling's Formula (7), after simplifications we obtain that

$$\frac{1}{\sqrt{\pi n}} e^{-\frac{1}{n}} < a_{0,n}^* < \frac{1}{\sqrt{\pi n}} e^{\frac{2}{3n}} \quad (n \geq 1). \quad (23)$$

Second, similarly to (13), for $n \geq 1$ and $|k| \leq rn$, $r \in (0, 1)$, we obtain that

$$\log \frac{1}{\sqrt{\pi n}} - b_{k,n}^* - \frac{1}{n} - \frac{r^4}{3(1-r^2)^2 n} < \log a_{k,n}^* < \log \frac{1}{\sqrt{\pi n}} - b_{k,n}^* + \frac{2}{3n}, \quad (24)$$

where

$$b_{k,n}^* := 2nI\left(\frac{k}{n}\right) - \tanh^{-1}\left(\frac{k}{n}\right) = \sum_{j=1}^{\infty} \left(\frac{k}{n}\right)^{2j-1} \frac{1}{2j-1} \left(\frac{k}{n} - 1\right). \quad (25)$$

(25) clearly agrees with (18). (24) proves (a) and (b).

By the series in (25), for $n \geq 4$ and $|k| \leq n^{\frac{2}{3}}$ we obtain that

$$\left| b_{k,n}^* - \frac{k^2}{n} \right| < 2n^{-\frac{1}{3}}.$$

This and (24) imply (c). $\square$

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Feller, W. *An Introduction to Probability Theory and Its Applications*, 3rd ed.; Wiley: New York, NY, USA, 1968; Volume I.
2. Révész, P. *Random Walk in Random and Non-Random Environments*, 2nd ed.; World Scientific: Singapore, 2005.
3. Feller, W. On the normal approximation to the binomial distribution. *Ann. Math. Statist.* **1945**, *16*, 319–329. [[CrossRef](#)]
4. Littlewood, J.E. On the probability in the tail of a binomial distribution. *Adv. Appl. Prob.* **1969**, *1*, 43–72. [[CrossRef](#)]
5. McKay, B.D. On Littlewood's estimate for the binomial distribution. *Adv. Appl. Prob.* **1989**, *21*, 475–478. [[CrossRef](#)]
6. Massart, P. Tusnady's lemma, 24 years later. *Ann. I. H. Poincaré—PR* **2002**, *38*, 991–1007. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.