Best lower bound on the probability of a binomial exceeding its expectation

Iosif Pinelis

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA
E-mail: ipinelis@mtu.edu

Abstract
Let $X$ be a random variable distributed according to the binomial distribution with parameters $n$ and $p$. It is shown that $P(X > E(X)) \geq 1/4$ if $1 > p \geq c/n$, where $c := \ln(4/3)$, the best possible constant factor.

Keywords: binomial distribution, probability inequalities, exact bounds
2010 MSC: 60E15, 62E15

1. Summary and discussion

Theorem 1. Let $X = X_{n,p}$ be a random variable (r.v.) with the binomial distribution with parameters $n$ and $p$. Then

$$P(X > E(X)) \geq 1/4$$

if

$$1 > p \geq c/n,$$

where

$$c := \ln(4/3) = 0.28768 \ldots.$$  

Under condition $[2]$, the equality in $[1]$ is attained only if $n = 2$ and $p = 1/2$. The constant factor $c$ in $[2]$ is the best possible.

Complementing Theorem 1 is the following simple proposition.

Proposition 2. If $c/n \geq p \geq 0$, then $P(X > E(X)) = 1 - (1-p)^n \geq \max(1, bn)p$, where $b := (1 - e^{-c})/c = 0.86901 \ldots$.

A very short proof of Theorem 1 will be given in Section 2. This proof is based on a monotonicity result due to Anderson and Samuels [2], which in turn follows from a more general result due to Hoeffding [7].
A bit longer proof of Theorem 1, which may still be of interest, is relegated to the appendix. This second proof is based on a version of the Berry–Esseen bound, which takes care of the main case when \( np \geq 2 \) and \( n(1 - p) \geq 2 \), that is, when \( 2 \leq \mathbb{E}X \leq n - 2 \). The remaining cases are rather easy to deal with, since all the values of \( X \) are in the set \( \{0, \ldots, n\} \).

Previously it was shown [5] that, for \( X \) as in Theorem 1, one has

\[ P(X \geq \mathbb{E}X) > \frac{1}{4} \quad (4) \]

if

\[ p > \frac{1}{n}. \quad (5) \]

Theorem 1 improves the result of [5] in two ways at once:

(i) The (optimal) constant factor \( c = 0.28768 \ldots \) in (2) is better than the corresponding constant factor 1 in (5). (Concerning the strictness of the inequality \( P(X \geq \mathbb{E}X) > \frac{1}{4} \) in (4), here one may recall that the inequality \( P(X > \mathbb{E}X) \geq 1/4 \) in (1) is strict unless \( n = 2 \) and \( p = 1/2 \) – in which latter case condition (5) fails to hold.)

(ii) Instead of the probability \( P(X \geq \mathbb{E}X) \) in (4), we have the (possibly) smaller probability \( P(X > \mathbb{E}X) \) in (1).

Improvement (i) and the optimality of the constant factor \( c \) are illustrated in Figure 1, showing the graphs

- \( \{(p, P(X_{n,p} > np)) : 1/n \leq p < 1\} \) (solid)
- \( \{(p, P(X_{n,p} > np)) : c/n < p \leq 1/n\} \) (dashed, black)
- \( \{(p, P(X_{n,p} > np)) : 0 < p \leq c/n\} \) (dashed, gray)

for \( n = 5 \). This figure is similar to [5, Figure 2], where the graphs over the interval \( (c/n, 1/n) \] were dashed, too.

\[
P(X > \mathbb{E}X)
\]

\[
\begin{array}{cccccc}
 & 0 & \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & 1 \\
\hline
\mathbb{P} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 1 \\
\end{array}
\]

Figure 1: Graphs of \( P(X > \mathbb{E}X) \).

However, concerning improvement (i), one should note that the case when \( c \leq np < 1 \) – considered in the beginning of the proof of Theorem 1 – is comparatively simple. As for improvement (ii), inequality (4) follows from its
non-strict counterpart \( P(X \geq E X) \geq 1/4 \) upon noting that \( P(X > E X) = P(X_n > np) \) is right-continuous in \( p \) and \( P(X > E X) = P(X \geq E X) \geq 1/4 \) if \( np \) is not an integer.

So, the main distinction of the present note from [5] is perhaps that each of the two proofs of Theorem 1 given here appears to be significantly simpler than the proof in [5].

As noted in [5], inequality (4) was used several times in the machine learning literature, including [4, 14, 13], to bound the probability of the so-called relative deviation of frequencies from the corresponding probabilities for certain classes of events. Such results have applications to the so-called probably-approximately-correct (PAC) models of machine learning; concerning PAC models, see e.g. [12, 6, 8, 1].

In [3], the non-strict version, \( P(X \geq E X) \geq 1/4 \), of inequality (1) was obtained, but only for large enough \( n \) and \( p \geq 2/n \).

In [11, Lemma 13], it was shown that
\[
P(X \geq E X) \geq \min(p, 1/4)
\] (6) for
\[
p \in (0, 1/2].
\] (7)

This was used to prove a part of [11, Proposition 8]. To state that result, we need to reproduce several definitions from [11]. Let \((X, Y)\) be a random vector in \( X \times \{-1, 1\} \), where \( X \) is a Borel subset of \( \mathbb{R}^d \). A classifier is a Borel-measurable map from \( X \) to \( \{-1, 1\} \). For any classifier \( h \), consider the two types of error probabilities,
\[
R^-(h) := P(h(X) \geq 0 | Y = -1) \quad \text{and} \quad R^+(h) := P(h(X) < 0 | Y = 1),
\]
and also the empirical counterpart
\[
\hat{R}^-(h) := \frac{1}{n'} \sum_{i=1}^{n'} 1(h(X_i^-) \geq 0)
\]
of \( R^-(h) \), where \( X_1, \ldots, X_n^- \) is a (training) iid sample from the conditional distribution of \( X \) given \( Y = -1 \), and \( 1(A) \) denotes the indicator of an assertion \( A \) (so that \( 1(A) = 1 \) if \( A \) is true and \( 1(A) = 0 \) if \( A \) is false).

The mentioned result in [11] is as follows: there exist classifiers \( h_1 \) and \( h_2 \) and a probability distribution for \((X, Y)\) such that, for any \( \alpha \in (0, 1/2] \) and any r.v. \( A \) with values in \([0, 1]\) such that for the random “pseudo-classifier” \( h_A := \Lambda h_1 + (1 - \Lambda) h_2 \) we have \( \hat{R}^- (h_A) < \alpha \), the event that the “excess type II risk”
\[
R^+(h_A) - \min_{x \in [0,1]: R^+(h_x)} R^+(h_x)
\]
is \( \geq \alpha \) occurs with a probability \( P \geq \min(\alpha, 1/4) \).

Using inequality (1) with condition (2) – instead of inequality (6) with condition (7), we can replace the conditions \( \alpha \in (0, 1/2] \) and \( P \geq \min(\alpha, 1/4) \) in
the cited result in [11] by the respective conditions $\alpha \in [c/n, 1]$ and $P \geq 1/4$, which will constitute a substantial improvement, in the case when $\alpha \geq c/n$. For the simpler case of $\alpha \in (0, c/n]$, an improvement over the result in [11] can be similarly obtained using Proposition 2.

2. Proofs

Here and in what follows, \( q := 1 - p. \) \hspace{1cm} (8)

**Proof of Theorem 1** If $n = 1$, then

\[
P(X > \mathbb{E}X) = P(X > p) = P(X = 1) = p = np \geq c > 1/4,
\]

so that (1) holds, with the strict inequality.

Fix now any natural $n \geq 2$. Consider first the case when $c \leq np < 1$. Then

\[
P(X > np) = 1 - q^n \geq 1 - q^{c/p} = 1 - \left(\frac{4}{3}\right)^{\ln(1 - p)}/p > 1 - \left(\frac{4}{3}\right)^{-1} = \frac{1}{4}, \tag{9}
\]

so that $P(X > np) > 1/4$. Moreover, if $c = \ln \frac{4}{3}$ is replaced here by any $c_1 \in (0, c)$, and if $p = c_1/n$ with $n \to \infty$, then $P(X > np) = 1 - q^n = 1 - \left(1 - c_1/n\right)^n \to 1 - e^{-c_1} < 1 - e^{-c} = 1/4$.

Therefore, the constant factor $c$ in (2) cannot be improved and, moreover, without loss of generality (wlog)

\[
np \geq 1. \tag{10}
\]

So,

\[
m := m_n := \lfloor np \rfloor + 1 \in [2, n]. \tag{11}
\]

Introduce also

\[
p_j := p_{n,j} := (m_n - 1)/j = (m - 1)/j \tag{12}
\]

for $j \in \{m, \ldots, n\}$. Then

\[
P(X > \mathbb{E}X) = P(X_{n,p} > np) = P(X_{n,p} \geq m) \geq P(X_{n,p} \geq m). \tag{13}
\]

The latter inequality, which follows from the (strict) stochastic monotonicity of $X_{n,p}$ in $p$ and the inequality $p \geq p_n$, is strict unless $p = p_n$ (that is, unless $np$ is an integer). Next, by part (i) of [10 Theorem 3] (which immediately follows from the second inequality in [2 Theorem 2.1], again by the stochastic monotonicity of $X_{n,p}$ in $p$), we have $P(X_{j+1,p_{j+1}} \geq m) > P(X_{j,p_j} \geq m)$ for all $j \in \{m, \ldots, n - 1\}$. So, $P(X_{n,p_n} \geq m) \geq P(X_{m,p_m} \geq m)$, and this inequality is strict unless $m = n$. Also, $P(X_{m,p_m} \geq m) = (1 - 1/m)^m \geq (1 - 1/2)^2 = 1/4$, and $P(X_{m,p_m} \geq m) > 1/4$ unless $m = 2$. It follows that $P(X > \mathbb{E}X) > 1/4$ unless $n = m = 2$ and $np$ is an integer. Thus, in view of [10], $P(X > \mathbb{E}X) > 1/4$ unless $n = 2$ and $p = 1/2$. That $P(X > \mathbb{E}X) = 1/4$ if $n = 2$ and $p = 1/2$ is trivial. This completes the proof of Theorem 1. \( \square \)
Proof of Proposition 2. If $c/n \geq p \geq 0$, then $P(X > EX) = 1 - (1-p)^n$. Next, $(1 - (1-p)^n)/(np)$ is decreasing in $p \in (0,1]$, so that for $p \in (0,c/n]$ we have $(1 - (1-p)^n)/(np) \geq (1 - (1-c/n)^n)/c \geq (1-e^{-c})/c = b$, so that $1 - (1-p)^n \geq bnp$. The inequality $1 - (1-p)^n \geq p$ is obvious. This completes the proof of Proposition 2.

References

[1] Alon, N., Hanneke, S., Holzman, R., Moran, S., 2021. A theory of PAC learnability of partial concept classes. arXiv:2107.08444.

[2] Anderson, T.W., Samuels, S.M., 1967. Some inequalities among binomial and Poisson probabilities, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, University of California Press, Berkeley, Calif., pp. 1–12. URL: https://projecteuclid.org/euclid.bsmsp/1200512976

[3] Andonova Jaeger, S., 2005. Generalization bounds and complexities based on sparsity and clustering for convex combinations of functions from random classes. J. Mach. Learn. Res. 6, 307–340.

[4] Anthony, M., Shawe-Taylor, J., 1993. A result of Vapnik with applications. Discrete Appl. Math. 47, 207–217. doi:10.1016/0166-218X(93)90126-9

[5] Greenberg, S., Mohri, M., 2014. Tight lower bound on the probability of a binomial exceeding its expectation. Statist. Probab. Lett. 86, 91–98. URL: https://doi.org/10.1016/j.spl.2013.12.009 doi:10.1016/j.spl.2013.12.009

[6] Haussler, D., 1992. Decision theoretic generalizations of the PAC model for neural net and other learning applications. Inf. Comput. 100, 78–150. URL: http://dx.doi.org/10.1016/0890-5401(92)90010-D doi:10.1016/0890-5401(92)90010-D

[7] Hoeffding, W., 1956. On the distribution of the number of successes in independent trials. Ann. Math. Statist. 27, 713–721. URL: https://doi.org/10.1214/aoms/1177728178 doi:10.1214/aoms/1177728178

[8] Kontorovich, A., Pinelis, I., 2019. Exact lower bounds for the agnostic probably-approximately-correct (PAC) machine learning model. Ann. Statist. 47, 2822–2854. doi:10.1214/18-AOS1766

[9] Korolev, V., Shevtsova, I., 2012. An improvement of the Berry-Esseen inequality with applications to Poisson and mixed Poisson random sums. Scand. Actuar. J. , 81–105 doi:10.1080/03461238.2010.485370

[10] Pinelis, I., 2020. Monotonicity properties of the Poisson approximation to the binomial distribution. Statist. Probab. Lett. 167, 108901, 7. doi:10.1016/j.spl.2020.108901
Appendix

Second proof of Theorem 1. At least one of the following five cases must occur:

Case 1: $np \geq 2$ and $nq \geq 2$ (recall the convention $q := 1 - p$ in (8)).

Case 2: $c \leq np < 1$ and $n \geq 1$, where $c$ is as in (3).

Case 3: $1 \leq np < 2$ and $n \geq 3$.

Case 4: $1 < nq \leq 2$ and $n \geq 3$.

Case 5: $0 < nq \leq 1$ and $n \geq 2$.

In particular, note that the cases when either (i) $n = 1$ or (ii) $n = 2$ and $p < 1/2$ are covered by Case 2 whereas the case when $n = 2$ and $p \geq 1/2$ is covered by Case 5.

Consider now each of the five listed cases.

Case 1 The version of the Berry–Esseen bound given in [9, Theorem 1] implies

$$P(X > \mathbb{E}X) = P(X > np) \geq \frac{1}{2} - \varepsilon(n, p),$$

where $\varepsilon(n, p) := \frac{c_3}{\sqrt{n}} \left( \frac{\rho}{\sigma^3} + c_2 \right)$,

$$\rho = p^3q + q^3p, \quad \sigma = \sqrt{pq}, \quad c_3 := \frac{34177}{100000}, \quad c_2 = \frac{429}{1000}.$$

Note that $p^3q/\sigma^3 = p^{3/2}(1 - p)^{-1/2}$ is convex in $p$ and, similarly, $q^3p/\sigma^3$ is convex in $p$, so that $\rho/\sigma^3$ and $\varepsilon(n, p)$ are convex in $p$. Therefore and in view of the Case 1 conditions $np \geq 2$ and $nq \geq 2$, we have $\varepsilon(n, p) \leq \varepsilon(n, 2/n) = \varepsilon(n, 1 - 2/n) =: \varepsilon_*(n)$, which is a simple algebraic function of $n$. For the derivative $\varepsilon'_*(n)$ of $\varepsilon_*(n)$ in $n$, we see that $\varepsilon'_*(n)n^{5/2}(n - 2)^{3/2}$ is a polynomial in $(n - 2)^{1/2}$, of degree 5. Therefore, it is easy to see that $\varepsilon_*(n)$ is decreasing in $n \in [4, 6]$, increasing in $n \in [7, 89]$, and decreasing in $n \in [90, \infty)$. Also, the conditions $np \geq 2$ and $nq \geq 2$ imply $n = np + nq \geq 4$. So, in Case 1 $P(X > np) \geq \frac{1}{2} - \max(\varepsilon_*(4), \varepsilon_*(89), \varepsilon_*(90)) > 0.25587 > 1/4$.

Case 2 Then, by (9), $P(X > np) > \frac{1}{3}$. Moreover, it was shown in the paragraph containing (2) that the constant factor $c$ in (2) cannot be improved.

Case 3 Then

$$P(X > np) = P(X > 1) = 1 - q^n - nq^{n-1}p,$$
which is increasing in \( p \), by the stochastic monotonicity of \( X_{n,p} \) in \( p \). So, wlog \( p = 1/n \), in which case \( \mathbb{P}(X > np) = f_3(n) := 1 - (2 - 1/n)(1 - 1/n)^{n-1} \).

The second derivative of \( \ln(1 - f_3(n)) \) is \( 1/((2n - 1)^2(n - 1)n) > 0 \), so that \( \ln(1 - f_3) \) is convex. Also, \( \ln(1 - f_3(n)) \to \ln(2/e) \). Therefore, \( \ln(1 - f_3(n)) \) is decreasing (in \( n \geq 3 \)) and \( f_3(n) \) is increasing, from \( f_3(3) = 7/27 > 1/4 \). Thus, \( \mathbb{P}(X > np) > 1/4 \) in Case 3.

Case 4: Then \( n - 2 \leq np < n - 1 \), \( p \geq 1 - 2/n \), and

\[
\mathbb{P}(X > np) = \mathbb{P}(X \geq n - 1) = f_1(p) := f_1(p, n) := p^n + np^{n-1}q,
\]

and \( f_1(p) \) is increasing in \( p \), by the stochastic monotonicity of \( X_{n,p} \) in \( p \). Therefore, here wlog \( p = 1 - 2/n \), and

\[
f_1(n) := f_1(1 - 2/n, n) = \frac{3n - 2}{n - 2} \left(1 - \frac{2}{n}\right)^n.
\]

Letting

\[
Df_1(n) := f_1'(n) = \frac{(1 - 2/n)^n (3n - 2)}{n - 2} = \ln(1 - 2/n) + \frac{6n - 8}{(n - 2)(3n - 2)},
\]

we have

\[
(Df_1)'(n) = \frac{-4(3n^2 - 4n + 4)}{(3n - 2)^2(n - 2)^2 n} < 0.
\]

So, \( Df_1 \) is decreasing. Also, \( Df_1(\infty) = 0 \). It follows that \( Df_1 > 0 \) and hence \( f_1 \) is increasing, from \( f_1(3) = \frac{5}{27} > \frac{1}{4} \). Thus, \( \mathbb{P}(X > np) > \frac{1}{4} \) in Case 4.

Case 5: Then \( n > np \geq n - 1 \), \( p \geq 1 - 1/n \), and hence

\[
\mathbb{P}(X > np) = \mathbb{P}(X = n) = p^n \geq \left(1 - \frac{1}{n}\right)^n,
\]

and \( \left(1 - \frac{1}{n}\right)^n \) is increasing in \( n \geq 2 \), from \( (1 - 1/2)^2 = 1/4 \). So, \( \mathbb{P}(X > np) > \frac{1}{4} \) in Case 5 – except when \( n = 2 \) and \( p = 1/2 \), in which case \( \mathbb{P}(X > np) = \frac{1}{4} \).

This completes the second proof of Theorem 1. \( \square \)