Statistics and Probability Letters 86 (2014) 91–98

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

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

Spencer Greenberg [a,*], Mehryar Mohri [a,b]

[a] Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, NY 10012, United States
[b] Google Research, 76 Ninth Avenue, New York, NY 10011, United States

ARTICLE INFO

Article history:
Received 2 November 2013
Received in revised form 11 December 2013
Accepted 12 December 2013
Available online 22 December 2013

Keywords:
Binomial distribution
Lower bound
Expected value
Relative deviation
Machine learning

ABSTRACT

We give the proof of a tight lower bound on the probability that a binomial random variable exceeds its expected value. The inequality plays an important role in a variety of contexts, including the analysis of relative deviation bounds in learning theory and generalization bounds for unbounded loss functions.

1. Motivation

This paper presents a tight lower bound on the probability that a binomial random variable exceeds its expected value. If the binomial distribution were symmetric around its mean, such a bound would be trivially equal to 1/2. And indeed, when the number of trials m for a binomial distribution is large, and the probability p of success on each trial is not too close to 0 or to 1, the binomial distribution is approximately symmetric. With p is fixed, and m sufficiently large, the de Moivre–Laplace theorem tells us that we can approximate the binomial distribution with a normal distribution. But, when p is close to 0 or 1, or the number of trials m is small, substantial asymmetry around the mean can arise, as illustrated by Fig. 1, which shows the binomial distribution for different values of m and p.

The lower bound we prove has been invoked several times in the machine learning literature, starting with work on relative deviation bounds by Vapnik (1998), where it is stated without proof. Relative deviation bounds are useful bounds in learning theory that provide more insight than the standard generalization bounds because the approximation error is scaled by the square root of the true error. In particular, they lead to sharper bounds for empirical risk minimization, and play a critical role in the analysis of the generalization bounds for unbounded loss functions (Cortes et al., 2010).

This binomial inequality is mentioned and used again without proof or reference in Anthony and Shawe-Taylor (1993), where the authors improve the original work of Vapnik (1998) on relative deviation bounds by a constant factor. The same claim later appears in Vapnik (2006) and implicitly in other publications referring to the relative deviations bounds of Vapnik (1998).

To the best of our knowledge, there is no publication giving an actual proof of this inequality in the machine learning literature. Our search efforts for a proof in the statistics literature were also unsuccessful. Instead, some references suggest in fact that such a proof is indeed not available. In particular, we found one attempt to prove this result in the context

* Corresponding author.
E-mail address: sgg247@nyu.edu (S. Greenberg).

0167-7152/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spl.2013.12.009

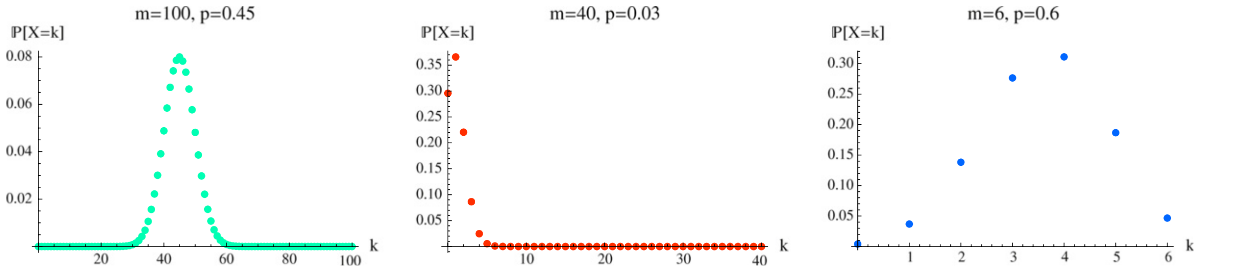

Fig. 1. Plots of the probability of getting different numbers of successes k, for the binomial distribution $B(m, p)$, shown for three different values of m, the number of trials, and p, the probability of a success on each trial. Note that in the second and third image, the distribution is clearly not symmetrical around its mean.

of the analysis of some generalization bounds (Jaeger, 2005), but the proof is not sufficient to show the general case needed for the proof of Vapnik (1998), and only pertains to cases where the number of Bernoulli trials is 'large enough'. A concise proof of this inequality for the special case where $p \leq \frac{1}{2}$ (Rigollet and Tong, 2011) was also recently brought to our attention. However that proof technique does not seem to readily extend to the general case. We also derived an alternative straightforward proof for the same special case using Slud's inequality following a suggestion of Luc Devroye (private communication). However, the proof of the general case turned out to be more challenging. Our proof therefore seems to be the first rigorous justification of this inequality, which is needed, among other things, for the analysis of relative deviation bounds in machine learning.

In Section 2, we start with some preliminaries and then give the presentation of our main result. In Section 3, we give a detailed proof of the inequality.

2. Main result

The following is the standard definition of a binomial distribution.

Definition 1. A random variable X is said to be distributed according to the binomial distribution with parameters m (the number of trials) and p (the probability of success on each trial), if for $k = 0, 1, \ldots, m$ we have

$$\mathbb{P}[X = k] = \binom{m}{k} p^k (1 - p)^{m-k}. \tag{1}$$

The binomial distribution with parameters m and p is denoted by $B(m, p)$. It has mean mp and variance $mp(1 - p)$.

The following theorem is the main result of this paper.

Theorem 1. *For any positive integer m and any probability p such that $p > \frac{1}{m}$, let X be a random variable distributed according to $B(m, p)$. Then, the following inequality holds:*

$$\mathbb{P}\big[X \geq \mathbb{E}[X]\big] > \frac{1}{4}, \tag{2}$$

where $\mathbb{E}[X] = mp$ is the expected value of X.

The lower bound is never reached but is approached asymptotically when $m = 2$ as $p \to \frac{1}{2}$ from the right. Note that when $m = 2$, the case $p = \frac{1}{2}$ is excluded from consideration, due to our assumption $p > \frac{1}{m}$. In words, the theorem says that a coin that is flipped a fixed number of times always has a probability of more than $1/4$ of getting at least as many heads as the expected value of the number of heads, as long as the coin's chance of getting a head on each flip is not so low that the expected value is less than or equal to 1. The inequality is tight, as illustrated by Fig. 2.

3. Proof

Our proof is based on the following series of lemmas and corollaries and makes use of Camp–Paulson's normal approximation to the binomial cumulative distribution function (Johnson et al., 1995, 2005; Lesch and Jeske, 2009). We start with a lower bound that reduces the problem to a simpler one.

Lemma 1. *For all $k = 1, 2, \ldots, m - 1$ and $p \in (\frac{k}{m}, \frac{k+1}{m}]$, the following inequality holds:*

$$\mathbb{P}_{X\sim B(m,p)}[X \geq \mathbb{E}[X]] \geq \mathbb{P}_{X\sim B(m,\frac{k}{m})}[X \geq k + 1].$$