How tight is Chernoff bound?

Nima Mousavi
University of Waterloo

1 Abstract

The well known Chernoff bound says that sum of m independent binary random variables with parameter p deviates from its expectation $\mu = mp$ with the standard deviation of at most $\sigma = \sqrt{m}$ in general and $\sigma = \sqrt{\mu}$ for small deviations. It is shown here that the sum deviates from its mean μ with standard deviation of at least $\sqrt{\mu}$.

2 A Motivating Example

Consider a biased coin where one side comes up with probability $p = 1/2 - \epsilon$. Your task is to find out which side (head or tail) is the less probable one. One can prove that the best solution for this problem is to toss the coin sufficiently many times and declare the side that appears less. How many times do you need to toss the coin to be confident about your prediction?

To make it more formal, we say that one is δ-confident about his prediction if it is correct with probability at least $1 - \delta$. Using Chernoff bound you find that $m = 3p/\epsilon^2 \ln(1/\delta) = O(\frac{1}{\epsilon^2} \ln(1/\delta)$ trials are sufficient to be δ-confident. The result here shows that in fact any prediction algorithm, in particular the above one requires at least $m = p/2\epsilon^2 \ln(1/4\delta) = \Omega(\frac{1}{\epsilon^2} \ln(1/\delta))$ samples.

3 Main

Let $X_1, X_2, \ldots, X_m$ be identical independent (i.i.d) random variables. The Chernoff bound gives exponentially decreasing bound on tail distributions of the sum $\sum_{i=1}^{m} X_i$. In particular we have the following result for the upper tail distribution.

Theorem 1 (Chernoff Bound)**.** *Let $X_1, X_2, \ldots, X_m$ be i.i.d random variables taking values 0 or 1, and $Pr[X_i = 1] = p$. Then*

1. *for any $t \geq 0$ [General bound]*

$$Pr\left[(\sum_{i=1}^{m} X_i - \mu) > t\right] \leq \exp(\frac{-2t^2}{m}) \tag{1}$$

2. *for any $0 \leq t \leq mp$ [Bound for small deviation]*

$$Pr\left[(\sum_{i=1}^{m} X_i - \mu) > t\right] \leq \exp(\frac{-t^2}{3\mu}) \tag{2}$$

Where $\mu = E[\sum_{i=1}^{m} X_i] = mp$.

(1) and (2) respectively imply standard deviations of $\sqrt{m}$ and $\sqrt{\mu}$ for upper tail distribution of sum of i.i.d binary random variables. There are various generalizations of the Chernoff Bounds that can be found for example in [1]. Next theorem shows that (2) is actually tight within a constant factor.

Theorem 2. *Let $X_1, X_2, \ldots, X_m$ be i.i.d random variables taking values 0 or 1, and $Pr[X_i = 1] = p$.*

- *If $p \leq \frac{1}{4}$, then for any $t \geq 0$*

$$Pr\left[(\sum_{i=1}^{m} X_i - \mu) > t\right] \geq \frac{1}{4}\exp(\frac{-2t^2}{\mu}) \tag{3}$$

- *If $p \leq \frac{1}{2}$, then for any $0 \leq t \leq m(1-2p)$*

$$Pr\left[(\sum_{i=1}^{m} X_i - \mu) > t\right] \geq \frac{1}{4}\exp(\frac{-2t^2}{\mu}) \tag{4}$$

Where $\mu = E[\sum_{i=1}^{m} X_i] = mp$.

Proof. We use Slud's Inequality [2]. Let X be sum of m Bernoulli trials with success-probability p. If either $p \leq \frac{1}{4}$ and $k \geq mp$ or $p \leq \frac{1}{2}$ and $mp \leq k \leq m(1-p)$, then

$$Pr[X \geq k] \geq Pr[Z \geq \frac{k - mp}{\sqrt{mp(1-p)}}] \tag{5}$$

where Z is a normal $(0,1)$ random variable [2]. Therefore if either $p \leq 1/4$ and $t \geq 0$ or $p \leq 1/2$ and $0 \leq t \leq m(1-2p)$

$$Pr\left[(\sum_{i=1}^{m} X_i - \mu) > t \right] \geq Pr[Z \geq \frac{t}{\sqrt{mp(1-p)}}] \tag{6}$$

Using standard lower bounds for upper tail of a normal random variables (see [3]) we get

$$Pr[Z > z] \geq \frac{1}{2}(1 - \sqrt{1 - e^{-z^2}}) \tag{7}$$

Thus,

$$\begin{aligned} Pr\left[(\sum_{i=1}^{m} X_i - \mu) > t \right] &\geq \frac{1}{2}(1 - \sqrt{1 - \exp(\frac{-t^2}{mp(1-p)})}) \\ &\geq \frac{1}{2}(1 - \sqrt{1 - \exp(\frac{-2t^2}{mp})}) \qquad (8) \\ &\geq \frac{1}{4}\exp(\frac{-2t^2}{mp}) \qquad (9) \end{aligned}$$

Where the last inequality follows from the fact that $1 - \sqrt{x} \geq \frac{1-x}{2}$ for all x.

□

References

[1] D. P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms, chapter 1*, Cambridge University Press, 2009.

[2] E. V. Slud, *Distribution Inequalities for the Binomial Law*, Ann. Probab. Volume 5, Number 3 (1977), 404-412.

[3] M. Anthony and P. L. Bartlett, *Neural Network Learning - Theoretical Foundations, Appendix 1*, Cambridge University Press, 2002.