Chernoff Bound

Kaulik Poddar (MD2207)

Saheli Datta (MD2213)

Tiyasa Dutta (MD2225)

ISI- Delhi

24.04.2023

From where did we get the notion of Chernoff Bound ?

While proving Glievenko Cantelli Lemma we arrived at a stage that

$$\mathbb{P}(|\hat{F}_n(x) - F(x)| > \epsilon) \leq \frac{\mathbb{V}\mathrm{ar}(\hat{F_n(x)})}{\epsilon^2}$$

where

$$\mathbb{V}\mathrm{ar}(\hat{F_n(x)}) = \frac{F(x)(1 - F(x))}{n\epsilon^2}$$

and this upper bound is not summable so we cant prove almost sure convergence in this way so from here we were introduced to the concept of Chernoff Bound.

Statement :

For a r.v X and for any ϵ we have

$$\mathbb{P}(X > \epsilon) \leq \inf_{\lambda > 0} \frac{\mathbb{E}e^{\lambda x}}{e^{\lambda \epsilon}} \qquad - (i)$$

and similarly

$$\mathbb{P}(X < \epsilon) \leq \inf_{\lambda > 0} \frac{\mathbb{E}[e^{-\lambda X}]}{e^{-\lambda \epsilon}} \qquad - (ii)$$

where both (i) and (ii) gives an upper bound to the tail probability.

This was named after Herman Chernoff in 1952.

In Probability theory a Chernoff Bound is an exponentially decreasing upper bound on the tail of a r.v based on its MGF.

Properties :

- If $\epsilon < \mathbb{E}[X]$ then the upper bound is trivially 1.
- Similarly if $\epsilon > \mathbb{E}[X]$ then also the upper bound is trivially 1 .
- Let us denote the bound by $C(\epsilon) = \inf_{\lambda > 0} \frac{\mathbb{E}e^{\lambda x}}{e^{\lambda \epsilon}}$ then

$$C_{X+k}(\epsilon) = C_X(\epsilon - k)$$

- The bound is exact iff X is a degenerated r.v.
- The bound is tight only at or beyond the extremes of a bounded r.v where the infima are attained for infinite λ.
- For unbounded r.v the bound is nowhere tight.

Comparison between Markov Chebyshev's and Chernoff Bound

We will do this through an example.

Set up : X~Bin(n,p), we will bound

$$\mathbb{P}(X > \alpha n)$$

where p< α <1.

$$\mathbb{P}(X > \alpha n) \leq \frac{p}{\alpha} \quad - Markov's Bound$$

$$\mathbb{P}(X > \alpha n) \leq \frac{p(1-p)}{n(\alpha - p)^2} \quad - Chebyshev's Bound$$

$$\mathbb{P}(X > \alpha n) \leq (\frac{p}{\alpha})^{\alpha n} (\frac{1-p}{1-\alpha})^{n(1-\alpha)} \quad - Chernoff's Bound$$

Now we will see what happens if we specify a value of p and α. Moreover in general Chernoff's perform better than Markov and it assumes more than that of the assumptions done in Chebyshev.

Hoeffding's Lemma:

Suppose X is a random Variable such that $X \in [a, b]$ almost surely. Then,

$$\mathbb{E}[\mathrm{e}^{\mathrm{s}(\mathbb{X}-\mathbb{E}[\mathbb{X}])}] \leq e^{\frac{s^2(b-a)^2}{8}}$$

Proof:

WLG, replace X by $X - \mathbb{E}[\mathbb{X}]$.

We can assume $\mathbb{E}[\mathbb{X}] = 0$, so that a $\leq 0 \leq$ b.

Since, $\mathrm{e}^{\mathrm{s}\mathbb{X}}$ is convex function of x we have for all $x \in [a, b]$,

$$f(\lambda a + (1 - \lambda b)) \leq \lambda f(a) + (1 - \lambda) f(b), \lambda \in (0, 1)$$

$$\Rightarrow e^{sx} \leq \frac{b - x}{b - a} e^{sa} + \frac{x - a}{b - a} e^{sb}$$

$$\Rightarrow \mathbb{E}[\mathrm{e}^{\mathrm{s}\mathbb{X}}] \leq \frac{b}{b - a} e^{sa} + \frac{-a}{b - a} e^{sb}$$

$$= e^{L(s(b-a))}$$

where,

$$L(h) = \frac{ha}{b - a} + \log(1 + \frac{a - ae^{h}}{b - a})$$

$$L'(h) = \frac{a}{b-a} - \frac{ae^h}{b-ae^h}$$

$$L''(h) = -\frac{abe^h}{(b-ae^h)^2}$$

Now,

$$L(0) = 0$$

$$L'(0) = 0$$

and,

$$(b+ae^h)^2 \geq 0$$

$$\Rightarrow (b-ae^h)^2 + 4ae^hb \geq 0$$

$$\Rightarrow -\frac{abe^h}{(b-a)^2} \leq \frac{1}{4}$$

By Taylor's Series Expansion,

$$L(h) = L(0) + hL'(0) + \frac{h^2}{2} L''(\theta h)$$

for some $\theta \in (0, 1)$

$$= \frac{h^2}{2} L''(\theta h) \leq \frac{h^2}{8}$$

Hence,

$$\mathbb{E}[\mathrm{e}^{\mathrm{s}\mathbb{X}}] \leq e^{\frac{s^2(b-a)^2}{8}}$$

Application

Corollary: Multiplicative form of Chernoff Bound

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables such that X_i always lies in the interval [0,1]. Define $X = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E[X]}$. Let $p_i = \mathbb{E[X_i]}$.

Then, for any $0 \leq \delta < 1$,

$$\mathbb{P}[X < (1-\delta)\mu] \leq (\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}})^{\mu}$$

Now, we can further bound this probability,

$$\mathbb{P}[X < (1-\delta)\mu] \leq (\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}})^{\mu} \leq e^{\frac{-\delta^2\mu}{2}}$$

Randomised Algorithms

What is a Randomized Algorithm ?

A randomized algorithm is a technique that uses a source of randomness as part of its logic. It is used to reduce runtime, or time complexity in a standard algorithm. The algorithm works by generating a random number, r, within a specified range of numbers and making decisions based of the value of r.

Let us look at an Example.

Example

Suppose we want to estimate the number of divisors of a number M.

One way is to try out all the numbers from 1 to M and count them which are the divisors of M. This will give us the exact value of the total number of divisors of M. **The run time will be O(M)**.

Otherwise, if we consider the following:

- Take a number n.
- At each step of n iterations, pick a random number from 1 to n and see whether it is a divisor of M.
- Repeat the first 2 steps till nth iteration and see the total count, and divide it by n and Multiply with M. This will give an Estimate of the total number of divisors of M.

The run time will be O(n) which is lesser than **O(M)**

But anyone would ask the question "But how good is the algorithm? How far from the actual value is the answer? If I were to try out the algorithm with larger n, how much better would my estimate be?"

Here comes the application of the corollary we just have proved.

Application of the Multiplicative form of the Chebychev

Suppose that these algorithm runs are independent and each algorithm run takes a correct decision with probability p

Let, $X_1, X_2, \ldots, X_n$ be IID random variables following Bernoulli(p).

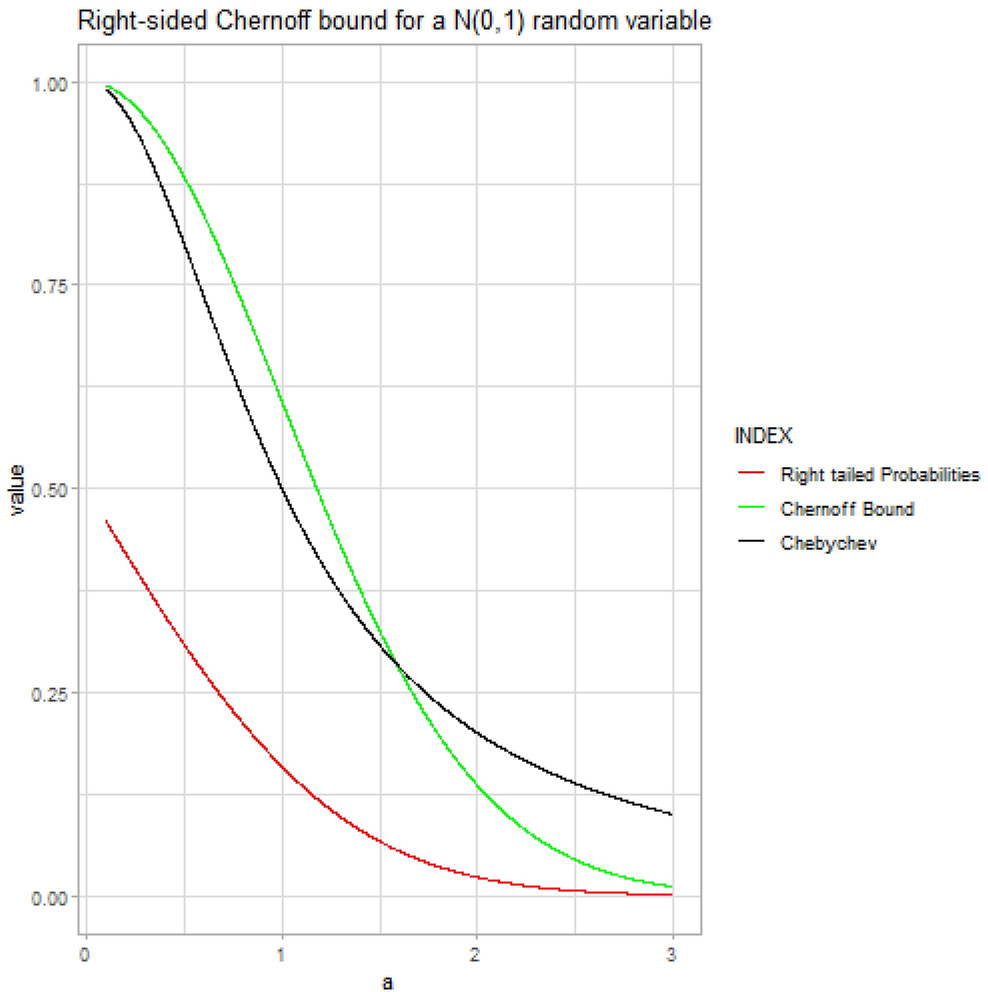
Then,

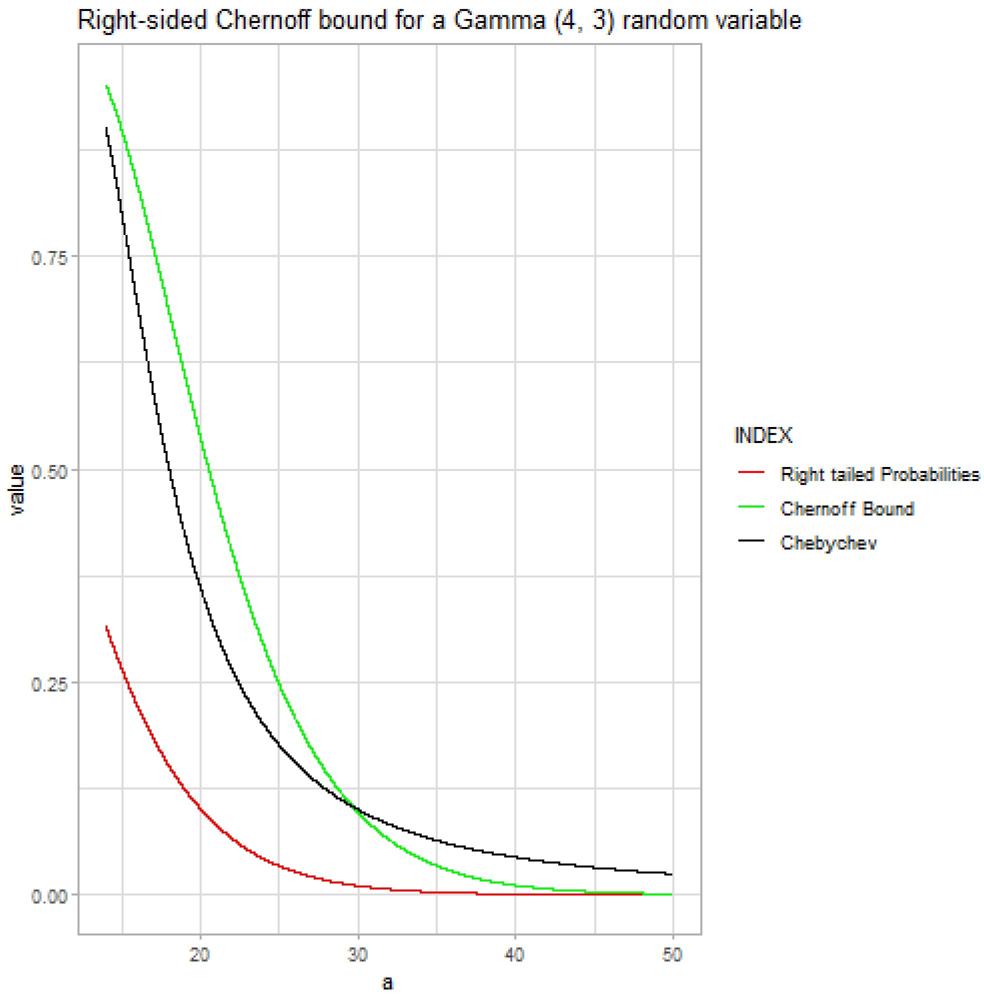
P[More than $\frac{n}{2}$ decisions are correct] = $\mathbb{P}[\mathrm{x} > \frac{\mathrm{n}}{2}] \geq 1 - e^{-n(p-(1/2))^2/(2p)}$ which is equal to $1-\delta$ if we choose

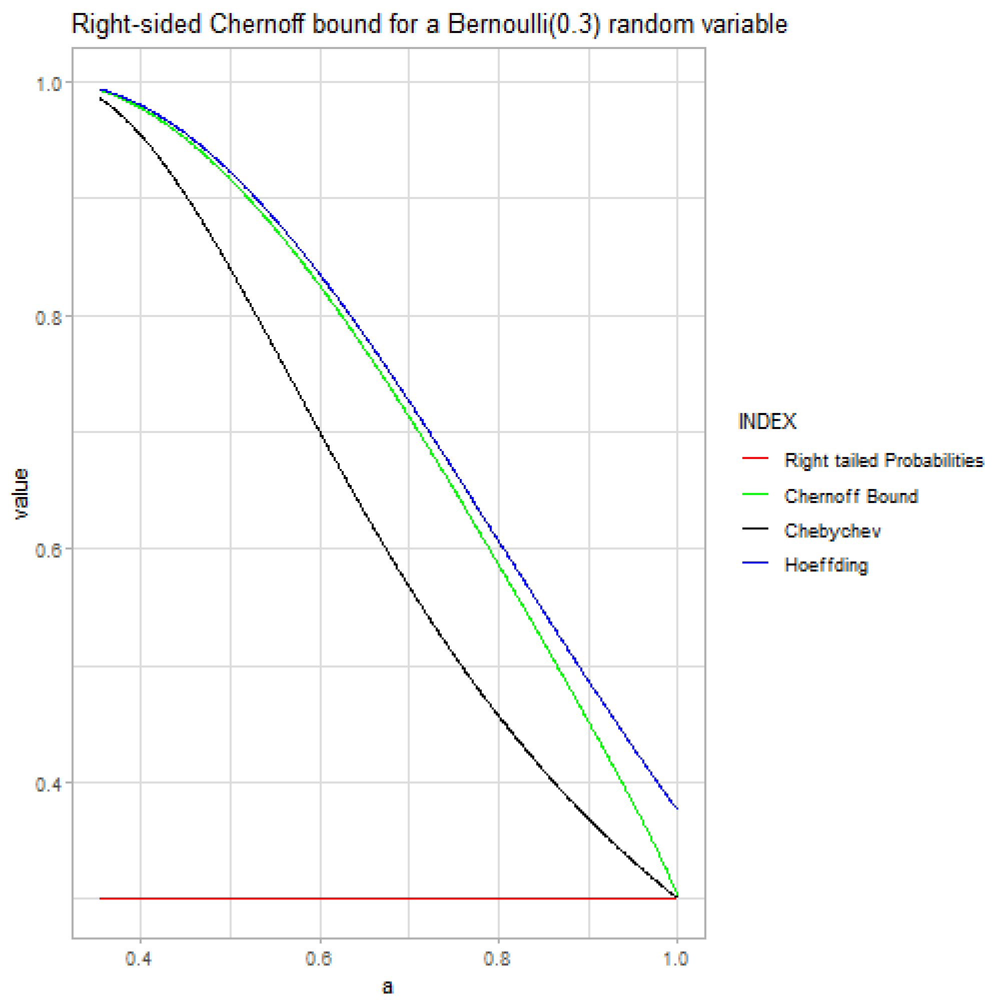
$$n = \log(1/\delta)2p/(p-1/2)^2$$

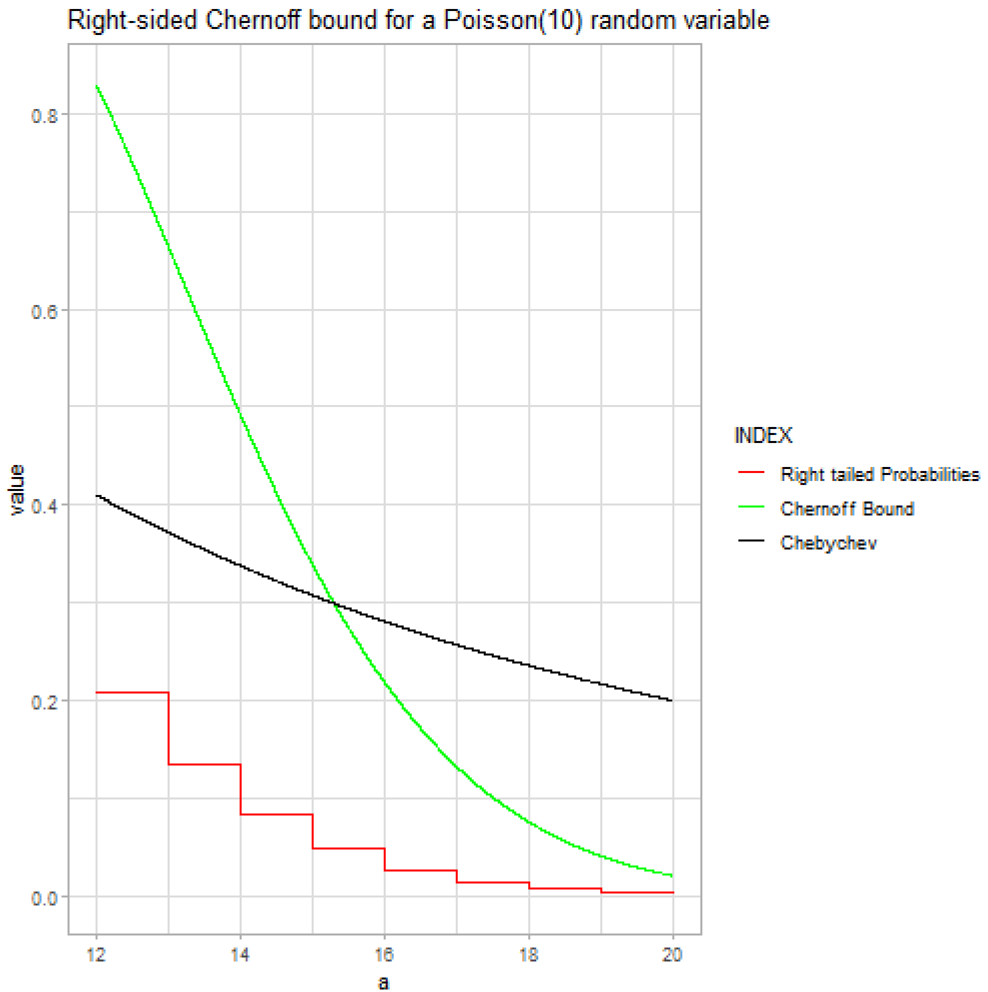
Therefore, under the above assumptions, if we pre specify the error δ we can always find a simulation number, using which we can increase the success rate.

Simulations




Right-sided Chernoff bound for a Gamma (4, 3) random variable
value
0.75
0.50
0.25
0.00
20
30
40
50
a
INDEX
Right tailed Probabilities
Chernoff Bound
Chebychev


Right-sided Chernoff bound for a Bernoulli(0.3) random variable
1.0
0.8
0.6
0.4
value
0.4
0.6
0.8
1.0
a
INDEX
Right tailed Probabilities
Chernoff Bound
Chebychev
Hoeffding


Right-sided Chernoff bound for a Poisson(10) random variable
value
0.0
0.2
0.4
0.6
0.8
12
14
16
18
20
a
INDEX
Right tailed Probabilities
Chernoff Bound
Chebychev