COMP 480/580 — Probabilistic Algorithms and Data Structure Jan 15, 2019

Lecture 3: Tail Bounds

Lecturer: Luay Nakhleh *Scribe By: Will Koh*

1 Motivation

- How large can a large variable be?
- How far can a value that the random variable takes be from its mean?

Say X is the number of steps an algorithm takes. Then, $\mathbb{E}[X]$ is the average-case running-time of the algorithm. In COMP 182, 382, and 582, the worst-case running-time of algorithms are studied, but what if an algorithm, on average, takes $2n$ steps but on certain inputs, it takes $500n^2$ steps?

2 Background

A random variable is a function from the sample space of an experiment / process to the set of real numbers. The expected value (also called the expectation or mean) of a discrete random variable X on the sample space S is

$$\mathbb{E}[X] = \sum_{s \in S} P(s) \cdot X(s)$$

or equivalently,

$$\mathbb{E}[X] = \sum_{x} x \cdot P(X = x)$$

The Linearity of Expectations is a property that states that the expectation of a sum of random variables is equivalent to the sum of the expectation of the random variables. That is,

$$\mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]$$

For constants within the expectations,

$$\mathbb{E}[aX_i + b] = a\mathbb{E}[X_i] + b$$

The variance of a random variable X on a sample space S is

$$\mathbb{V}[X] = \mathbb{E}\left[(X - E(X))^2\right]$$

or equivalently

$$\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2$$

Bienayme's Formula states that if X_i for $i \in 1, 2, ..., n$ are pairwise independent random variables on S, then

$$\mathbb{V}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{V}\left[X_i\right]$$

Side note: If the random variables are not pairwise independent, then you must account for the covariance. Thus,

$$\mathbb{V}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i} \mathbb{V}\left[X_i\right] + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$$

3 Markov's Inequality

Let X be a random variable that takes only nonnegative values. Then, for every real number $a > 0$,

$$P(X \geq a) \leq \frac{\mathbb{E}\left[X\right]}{a}$$

Proof:

$$\begin{aligned}
\mathbb{E}\left[X\right] &= \sum_{x} x \cdot P(X = x) \\
&= \left(\sum_{x \geq a} x \cdot P(X = x)\right) + \left(\sum_{x < a} x \cdot P(X = x)\right) \\
&\geq \left(\sum_{x \geq a} a \cdot P(X = x)\right) + (0) \qquad (1) \\
&= a\left(\sum_{x \geq a} P(X = x)\right) \\
&= a\big(P(X \geq a)\big)
\end{aligned}$$

$$\therefore P(X \geq a) \leq \frac{\mathbb{E}\left[X\right]}{a} \qquad (2)$$

QED.

3.1 Discussion

Markov's Inequality answers the question of how large a value X can take. Unfortunately, for distributions encountered in practice, Markov's inequality gives a very loose bound. Take, for example, a uniform distribution with a mean of 0 and a normal distribution with a mean of 0. In this scenario, the uniform distribution is more like to have a tale value than the normal distribution is. Unfortunately, Markov's Inequality does not take this into account.

3.2 Example: Graph Traversal Time

Assume the expected time Algorithm A takes to traverse a graph with n nodes is $2n$. What is the probability that the algorithm takes more than 10 times that?

Solution:

$$P(X \geq 10 \cdot \mathbb{E}[X]) \leq \frac{\mathbb{E}[X]}{10 \cdot \mathbb{E}[X]} = \frac{1}{10}$$

4 Chebyshev's Inequality

Let X be a random variable. For every real number $r > 0$,

$$P\big(|X - \mathbb{E}[X]| \geq a\big) \leq \frac{\mathbb{V}[X]}{a^2}$$

The proof is as follows:

$$P\big(|X - \mathbb{E}[X]| \geq a\big) = P\Big(\big(X - \mathbb{E}[X]\big)^2 \geq a^2\Big) \leq \frac{\mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big]}{a^2} \tag{3}$$

Note that $\mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \mathbb{V}[X]$.

$$P\big(|X - \mathbb{E}[X]| \geq a\big) \leq \frac{\mathbb{V}[X]}{a^2} \tag{4}$$

QED.

4.1 Discussion

A quick comparison of Markov's Inequality and Chebyshev's Inequality is shown below, respectively:

$$P(X \geq k\mu) \geq \frac{1}{k}$$

and

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

As it can be seen, Markov's Inequality scales linearly with k, whereas Chebyshev's Inequality scales quadratically with k, providing a tighter bound as k increases.

4.2 Example

Assume we have a distribution whose mean is 80 and standard deviation is 10. What is a lower bound on the percentage of values that fall exclusively between 60 and 100 in this distribution?

Solution:
We know $\mathbb{E}[X] = 80$, $\mathbb{V}[X] = \sigma^2 = 10^2 = 100$, and $a = \frac{100-60}{2} = 20$. Therefore,

$$P\big(|X(s) - 80| \geq 20\big) \leq \frac{100}{20^2} = \frac{1}{4}$$

and the lower bound is 75%.

5 Illustration: Estimating π Using the Monte Carlo Method

Assume that there is square board whose area is 1, and inside it is a circle whose radius is 1/2. Now, throw darts at this board. The probability that it lands inside the circle equals the ratio of the circle area to the square area ($\frac{\pi r^2}{4r^2} = \pi/4$). Therefore, calculate the proportion of times that the dart landed inside the circle and multiply it by 4. (*Pseudocode can be found on lecture slides.*)

Let X_i be the random variable that denotes whether the i^{th} dart landed inside the circle (1 if it did and 0 otherwise). Then,

$$\hat{\pi}(n) = 4\frac{\sum_{i=1}^{n} X_i}{n}$$

where $\hat{\pi}(n)$ is the estimate of π after n darts thrown.

The calculation of expectation and variance for X_i is shown below, using the properties of a Bernoulli distribution with $p = \frac{\pi}{4}$.

$$\mathbb{E}\left[X_i\right] = \frac{\pi}{4} \times 1 + \left(1 - \frac{\pi}{4}\right) \times 0 = \frac{\pi}{4}$$

$$\mathbb{V}\left[X_i\right] = \frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)$$

Using the Linearity of Expectation, calculate the expectation of $\hat{\pi}$.

$$\mathbb{E}\left[\hat{\pi}\right] = \mathbb{E}\left[\frac{4}{n}\sum_{i=1}^{n} X_i\right] = \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i\right] = \frac{4}{n} \cdot \frac{n\pi}{4} = \pi$$

Since the random variables X_i are pairwise independent (e.g. throwing a dart does not affect the result of another dart), we can use Bienayme's Formula to calculate the variance of $\hat{\pi}$. Note that multiplying a random variable by a constant increases the variance by the square of that constant.

$$\mathbb{V}\left[\hat{\pi}\right] = \mathbb{V}\left[\frac{4}{n}\sum_{i=1}^{n} X_i\right] = \frac{16}{n^2}\mathbb{V}\left[\sum_{i=1}^{n} X_i\right] = \frac{16}{n^2}\sum_{i=1}^{n}\mathbb{V}\left[X_i\right] = \frac{16}{n^2} \cdot \frac{n\pi}{4}(1 - \frac{\pi}{4}) = \frac{\pi(4-\pi)}{n}$$

5.1 Accuracy

The question that is raised at this point is: how big should n be for us to get a good estimate? In a probabilistic setting, the question can be reworded as: **what should the value of n be such that the estimation error of π is within δ with probability at least ϵ?**

$$P\Big(|\hat{\pi}(n) - \pi| < \delta\Big) > \epsilon$$

or equivalently

$$P\big(|\hat{\pi}(n) - \pi| \geq \delta\big) \leq 1 - \epsilon$$

To achieve a good estimate, δ should be very small and ϵ to be as close to 1 as possible. For this example, $\delta = 0.001$ and $\epsilon = 0.95$. The problem becomes solving for n in

$$P\big(|\hat{\pi} - \pi| \geq 0.001\big) \leq 0.05$$

The above equation is an application of Chebyshev's Inequality in the form of

$$P\big(|\hat{\pi} - \mathbb{E}\left[\hat{\pi}\right]| \geq a\big) \leq \frac{\mathbb{V}\left[\hat{\pi}\right]}{a^2}$$

From above,

$$\mathbb{V}\left[\hat{\pi}(n)\right] = \frac{\pi(4-\pi)}{n}$$

so we solve for n such that

$$\frac{\mathbb{V}\left[\hat{\pi}\right]}{a^2} = \frac{\pi(4-\pi)}{n(0.001)^2} \leq 0.05$$

Since we are trying to estimate the value of π, we cannot leave the variable in the answer. Therefore, we can use the following inequality to substitute away π:

$$\pi(4-\pi) \leq 4$$

Then, the problem can be written as

$$\frac{\pi(4-\pi)}{n(0.001)^2} \leq \frac{4}{n(0.001)^2} \leq 0.05$$

Solving for n in the above inequality gives

$$n \geq 80{,}000{,}000$$

to have an accuracy of $\delta = 0.001$ with probability $\epsilon = 0.95$.

6 The Weak Law of Large Numbers

6.1 Corollary of Chebyshev's Inequality

Let X_1, X_2, ..., X_n be independent random variables with

$$\mathbb{E}\left[X_i\right] = \mu_i \text{ and } V(X_i) = \sigma_i^2$$

Then, for any $a > 0$,

$$P\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right| \geq a\right) \leq \frac{\sum_{i=1}^{n} \sigma_i^2}{a^2}$$

The above is a corollary of the Chebyshev's Inequality. It extends Chebyshev's Inequality to multiple random variables using Linearity of Expectations and Bienayme's Formula. The proof is shown below:

$$P\big(|X - \mathbb{E}[X]| \geq a\big) \leq \frac{\mathbb{V}[X]}{a^2} \text{ (Chebyshev's Inequality)} \tag{5}$$

Let $X = \sum_{i=1}^{n} X_i$

$$\begin{aligned}
P\left(\left|\sum_{i=1}^{n} X_i - \mathbb{E}\left[\sum_{i=1}^{n} X_i\right]\right| \geq a\right) &\leq \frac{\mathbb{V}\left[\sum_{i=1}^{n} X_i\right]}{a^2} \\
P\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mathbb{E}[X_i]\right| \geq a\right) &\leq \frac{\sum_{i=1}^{n} \mathbb{V}[X_i]}{a^2} \\
P\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right| \geq a\right) &\leq \frac{\sum_{i=1}^{n} \sigma_i^2}{a^2}
\end{aligned} \tag{6}$$

QED.

6.2 The Weak Law of Large Numbers

Let X_1, X_2, ..., X_n be independently and identically distributed (i.i.d) random variables, where the unknown expected value μ is the same for all variables and their variance is finite. Then, for any $\epsilon > 0$, we have

$$P\left(\left|\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right| \geq \epsilon\right) \xrightarrow{n\to\infty} 0$$

This follows from the corollary above. The proof is as follows:

$$\begin{aligned}
P\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\right| \geq a\right) &\leq \frac{\sum_{i=1}^{n} \sigma_i^2}{a^2} \\
P\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mathbb{E}[X_i]\right| \geq a\right) &\leq \frac{\sum_{i=1}^{n} \mathbb{V}[X_i]}{a^2} \\
P\left(\left|\sum_{i=1}^{n} \frac{1}{n} X_i - \sum_{i=1}^{n} \mathbb{E}\left[\frac{1}{n} X_i\right]\right| \geq a'\right) &\leq \frac{\sum_{i=1}^{n} \mathbb{V}\left[\frac{1}{n} X_i\right]}{a'^2} \\
P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i]\right| \geq a'\right) &\leq \frac{\frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}[X_i]}{a'^2}
\end{aligned} \tag{7}$$

Let $a' = \epsilon$, where ϵ is a positive number. Also, note that $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}[X_i] = \sigma^2$ for all i.

$$
\begin{aligned}
P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu\right| \geq \epsilon\right) &\leq \frac{\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2}{\epsilon^2} \\
P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}n\mu\right| \geq \epsilon\right) &\leq \frac{\frac{1}{n^2}n\sigma^2}{\epsilon^2} \qquad (8) \\
P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) &\leq \frac{\sigma^2}{n\epsilon^2}
\end{aligned}
$$

As $n \to \infty$, the right hand side of the inequality approaches 0. Since the left hand side is a probability, it must be between 0 and 1 inclusive.

$$
\therefore P\left(\left|\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right| \geq \epsilon\right) \xrightarrow{n\to\infty} 0 \qquad (9)
$$

QED.

7 Chernoff Bounds

Let $X = X_1 + X_2 + \cdots + X_n$, where all the X_i's are independent and $X_i \sim Bernoulli(p_i)$, and let $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$. Then, for $\delta > 0$,

$$
P\big(|X - \mu| \geq \delta\mu\big) \leq 2e^{-\frac{\delta^2\mu}{2+\delta}}
$$

The bound can also be separated and written as

$$
P\big(X \geq (1+\delta)\mu\big) \leq e^{-\frac{\delta^2\mu}{2+\delta}} \text{ for } \delta > 0
$$

$$
P\big(X \leq (1-\delta)\mu\big) \leq e^{-\frac{\delta^2\mu}{2}} \text{ for } 1 > \delta > 0
$$

7.1 Lemmas

Before proving the two bounds, we must prove the following three lemmas.

7.1.1 Lemma 1

Given a random variable $Y \sim Bernoulli(p)$, $\forall s \in \mathbb{R}$,

$$
\mathbb{E}\left[e^{sY}\right] \leq e^{p(e^s - 1)}
$$

The proof is as follows:

$$\begin{aligned} \mathbb{E}[Y] &= p \times 1 + (1-p) \times 0 \\ \mathbb{E}\left[e^{sY}\right] &= e^{s\times 1} \times p + e^{s\times 0}(1-p) \\ &= e^s \times p + (1-p) \\ &= 1 + p(e^s - 1) \end{aligned} \tag{10}$$

Note that $1 + y \leq e^y$.

$$\mathbb{E}\left[e^{sY}\right] \leq e^{p(e^s-1)} \tag{11}$$

QED.

7.1.2 Lemma 2

Let X_1, ..., X_n be independent random variables, and $X = \sum_{i=1}^n X_i$. Then, for $s \in \mathbb{R}$,

$$\mathbb{E}\left[e^{sX}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{sX_i}\right]$$

The proof is as follows:

$$\begin{aligned} \mathbb{E}\left[e^{sX}\right] &= \mathbb{E}\left[e^{s\sum_{i=1}^n X_i}\right] \\ &= \mathbb{E}\left[e^{sX_1+sX_2+\cdots+sX_n}\right] \\ &= \mathbb{E}\left[e^{sX_1} \times e^{sX_2} \times \cdots \times e^{sX_n}\right] \\ &= \mathbb{E}\left[\prod_{i=1}^n e^{sX_i}\right] \end{aligned} \tag{12}$$

Note that the expectation of a product of independent random variables is equal to the product of the expectatino of the random variables.

$$\mathbb{E}\left[e^{sX}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{sX_i}\right] \tag{13}$$

QED.

7.1.3 Lemma 3

Let X_1, ..., X_n be independent random variables with Bernoulli distributions, and $X = \sum_{i=1}^n X_i$ and $\mathbb{E}[X] = \sum_{i=1}^n p_i = \mu$. Then, for $s \in \mathbb{R}$,

$$\mathbb{E}\left[e^{sX}\right] \leq e^{(e^s-1)\mu}$$

The proof falls from Lemma 1 and 2 and is as follows:

$$
\begin{aligned}
\mathbb{E}\left[e^{sX}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{sX_i}\right] &\leq e^{p_i(e^s-1)} \text{ (Lemma 1 \& 2)} \\
&= \prod_{i=1}^{n} e^{p_i(e^s-1)} \\
&= e^{p_1(e^s-1)} \times e^{p_2(e^s-1)} \times \cdots \times e^{p_n(e^s-1)} \\
&= e^{p_1(e^s-1)+p_2(e^s-1)+\cdots+p_n(e^s-1)} \\
&= e^{(e^s-1)(p_1+p_2+\cdots+p_n)} \\
&= e^{(e^s-1)\mu}
\end{aligned}
\tag{14}
$$

QED.

7.2 Proof of Chernoff Bounds

7.2.1 Proof of Upper Bound

The proof for the upper bound of Chernoff Bounds is shown below:

$$
\begin{aligned}
P(X \geq a) = P\left(e^{sX} \geq e^{sa}\right) &\leq \frac{\mathbb{E}\left[e^{sX}\right]}{e^{sa}} \text{ (Markov's Inequality)} \\
&= \frac{e^{(e^s-1)\mu}}{e^{sa}} \text{ (Lemma 3)}
\end{aligned}
\tag{15}
$$

We want to choose an s that will minimize the the upper bound. Therefore, we can derive the right hand side of the equation with respect to s and find the minimum by setting it to zero.

$$
\begin{aligned}
\frac{e^{(e^s-1)\mu}}{e^{sa}} &= e^{(e^s-1)\mu-sa} \\
\frac{d}{ds} e^{(e^s-1)\mu-sa} &= e^{(e^s-1)\mu-sa}\left(e^s\mu - a\right) = 0
\end{aligned}
\tag{16}
$$

The only term that can be 0 is the $e^s\mu - a$ so set that to zero and solve for s.

$$
\begin{aligned}
e^s\mu - a &= 0 \\
e^s &= \frac{a}{\mu} \\
s &= \ln \frac{a}{\mu}
\end{aligned}
\tag{17}
$$

Let $a = (1+\delta)\mu$.

$$
\begin{aligned}
s &= \ln \frac{(1+\delta)\mu}{\mu} \\
&= \ln(1+\delta)
\end{aligned}
\tag{18}
$$

This is a minimum since for values $s < \ln 1 + \delta$, the derivative is below 0, and vice versa for greater values. So, set $s = \ln(1+\delta)$ for $\delta > 0$.

$$\begin{aligned} P\big(X \geq (1+\delta)\mu\big) &\leq \frac{e^{(e^{\ln(1+\delta)}-1)\mu}}{e^{\ln(1+\delta)(1+\delta)\mu}} \\ &= \frac{e^{\delta\mu}}{(1+\delta)^{(1+\delta)\mu}} \\ &= \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \end{aligned} \tag{19}$$

By taking the natural logarithm of the expression for the upper bound, we get

$$\mu\left(\delta - (1+\delta)\ln(1+\delta)\right) \tag{20}$$

Knowing that $\ln(1+x) \geq \frac{x}{1+x/2}$,

$$\begin{aligned} \mu\left(\delta - (1+\delta)\ln(1+\delta)\right) &\leq \mu\left(\delta - (1+\delta)\frac{\delta}{1+\delta/2}\right) \\ &= \mu\delta\left(1 - \frac{1+\delta}{1+\delta/2}\right) \\ &= \mu\delta\left(\frac{1+\delta/2-1-\delta}{1+\delta/2}\right) \\ &= \mu\delta\left(-\frac{\delta/2}{1+\delta/2}\right) \\ &= -\frac{\delta^2}{2+\delta}\mu \end{aligned} \tag{21}$$

From equation 21, we can get to the expected upper bound.

$$P\big(X \geq (1+\delta)\mu\big) \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \leq e^{-\frac{\delta^2}{2+\delta}\mu} \text{ for } \delta > 0 \tag{22}$$

QED.

7.2.2 Proof of Lower Bound

The proof for the lower bound of Chernoff Bounds is shown below:

$$\begin{aligned} P(X \leq a) = P\left(e^{-sX} \geq e^{-sa}\right) &\leq \frac{\mathbb{E}\left[e^{-sX}\right]}{e^{-sa}} \text{ (Markov's Inequality)} \\ &= \frac{e^{(e^{-s}-1)\mu}}{e^{-sa}} \text{ (Lemma 3)} \end{aligned} \tag{23}$$

Same reasoning for how we want to choose our value for s as the upper bound.

$$\begin{aligned} \frac{e^{(e^{-s}-1)\mu}}{e^{-sa}} &= e^{(e^{-s}-1)\mu+sa} \\ \frac{d}{ds}e^{(e^{-s}-1)\mu+sa} &= e^{(e^{-s}-1)\mu+sa}\left(-e^{-s}\mu + a\right) = 0 \end{aligned} \tag{24}$$

The only term that can equal 0 is $-e^{-s}\mu + a$.

$$\begin{aligned} -e^{-s}\mu + a &= 0 \\ e^{-s}\mu &= a \\ e^{-s} &= \frac{a}{\mu} \\ e^{s} &= \frac{\mu}{a} \\ s &= \ln\frac{\mu}{a} \end{aligned} \tag{25}$$

Let $a = (1-\delta)\mu$.

$$\begin{aligned} s &= \ln\frac{\mu}{(1-\delta)\mu} \\ &= \ln\frac{1}{1-\delta} = -\ln(1-\delta) \end{aligned} \tag{26}$$

Set $s = -\ln(1-\delta)$ for $0 < \delta < 1$.

$$\begin{aligned} P\big(X \le (1-\delta)\mu\big) &\le \frac{e^{(e^{-(-\ln(1-\delta))}-1)\mu}}{e^{-(-\ln(1-\delta))(1-\delta)\mu}} \\ &= \frac{e^{(e^{\ln(1-\delta)}-1)\mu}}{e^{\ln(1-\delta)(1-\delta)\mu}} \\ &= \frac{e^{-\delta\mu}}{(1-\delta)^{(1-\delta)\mu}} \\ &= \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \end{aligned} \tag{27}$$

By taking the natural logarithm of the expression for the lower bound, we get

$$\mu\left(-\delta - (1-\delta)\ln(1-\delta)\right) \tag{28}$$

Knowing that $(1-\delta)\ln(1-\delta) \ge -\delta + \frac{\delta^2}{2}$ from the Taylor expansion of $\ln(1-x)$,

$$\begin{aligned} \mu\left(-\delta - (1-\delta)\ln(1-\delta)\right) &\le \mu\left(-\delta + \delta - \frac{\delta^2}{2}\right) \\ &= -\frac{\mu\delta^2}{2} \end{aligned} \tag{29}$$

From equation 29, we can get to the expected lower bound.

$$P\big(X \le (1-\delta)\mu\big) \le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \le e^{-\frac{\mu\delta^2}{2}} \text{ for } 0 < \delta < 1 \tag{30}$$

QED.

7.3 Discussion

Assume an experiment where a fair coin is tossed 200 times. How likely is it to observe at least 150 heads?

In this situation, note that $\mathbb{E}[X_i] = \frac{1}{2}$ and $\mathbb{V}[X_i] = \frac{1}{4}$. Each flip is also independent of each other.

7.3.1 Markov's Inequality

$$P(X \geq 150) \leq \frac{\mathbb{E}[X]}{150} = \frac{\sum_{i=1}^{200}(0.5)}{150} = 0.6666 \tag{31}$$

7.3.2 Chebyshev's Inequality

$$P\big(|X - 100| \geq 50\big) \leq \frac{\mathbb{V}[X]}{50^2} = \frac{\sum_{i=1}^{200}\mathbb{V}[X]}{50^2} = \frac{50}{50^2} = 0.02 \tag{32}$$

7.3.3 Chernoff Bounds

$$P\big(X \geq (1+\delta)\mu\big) \leq e^{-\frac{\delta^2\mu}{2+\delta}} \tag{33}$$

We know $(1+\delta)\mu = 150$ and $\mu = 100$, so $\delta = 0.5$.

$$P(X \geq 150) \leq e^{-\frac{(0.5)^2(100)}{2+0.5}} = 0.0000453999 \tag{34}$$

7.4 Alternate Form

Let $X = X_1 + X_2 + \cdots + X_n$, where all the X_i are independent random variables and $a \leq X_i \leq b$ for all i, and let $\mu = \mathbb{E}[X]$. Then, for $\delta > 0$

$$P\big(X \geq (1+\delta)\mu\big) \leq e^{-\frac{2\delta^2\mu^2}{n(b-a)^2}}$$

and

$$P\big(X \leq (1-\delta)\mu\big) \leq e^{-\frac{\delta^2\mu^2}{n(b-a)^2}}$$

References

[1] cs.rice.edu/ nakhleh/TailBounds.pdf
[2] math.mit.edu/ goemans/18310S15/chernoff-notes.pdf
[3] people.eecs.berkeley.edu/ jfc/cs174/lecs/lec10/lec10.pdf