**Lecture Notes CS:5360** Randomized Algorithms
**Lecture 7**: Sept 11, 2018
**Scribe**: Tanmay Inamdar

# 1 Markov's Inequality

Recall that our general theme is to upper bound tail probabilities, i.e., probabilities of the form $\Pr(X \geq c \cdot E[X])$ or $\Pr(X \leq c \cdot E[X])$. The first tool towards that end is Markov's Inequality.
**Note.** This is a simple tool, but it is usually quite weak. It is mainly used to derive stronger tail bounds, such as Chebyshev's Inequality.

**Theorem 1 (Markov's Inequality)** *Let $X$ be a non-negative random variable. Then,*

$$\Pr(X \geq a) \leq \frac{E[X]}{a}, \qquad \textit{for any } a > 0.$$

Before we discuss the proof of Markov's Inequality, first let's look at a picture that illustrates the event that we are looking at.

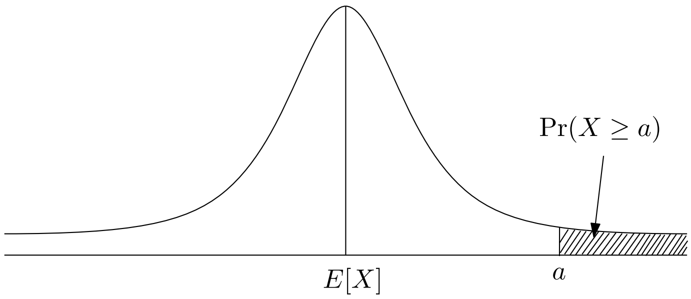


Figure 1: Markov's Inequality bounds the probability of the shaded region.

**Proof:**[1] Suppose $X$ is a discrete random variable, for simplicity.

$$\begin{aligned} E[X] &= \sum_x x \cdot \Pr(X = x) \\ &\geq \sum_{x \geq a} x \cdot \Pr(X = x) \\ &\geq a \cdot \sum_{x \geq a} \cdot \Pr(X = x) \\ &= a \cdot \Pr(X \geq a) \end{aligned}$$

Rearranging, we get

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$

$\square$

**Proof:**[2] Define a random variable $Y$ as follows. $Y = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise.} \end{cases}$

Now, if $X < a$, $Y = 0$. Otherwise, $X \geq a$, in which case $Y = 1$. In both cases, we have that $Y \leq \frac{X}{a}$. Note that we use the fact that $X$ is a non-negative random variable in the first case.

Therefore, $E[Y] \leq \frac{E[X]}{a}$. However, since $Y$ is an indicator random variable, $E[Y] = \Pr(Y = 1) = \Pr(X \geq a)$. This implies that $\Pr(X \geq a) \leq \frac{E[X]}{a}$. $\square$

**Example.** Let $X$ be a random variable that denotes the number of heads, when $n$ *fair* coins are tossed independently. Using Linearity of Expectation, we get that $E[X] = \frac{n}{2}$.

Plugging in $a = \frac{3n}{4}$ in Markov's Inequality, we get that $\Pr(X \geq \frac{3n}{4}) \leq \frac{n/2}{3n/4} = \frac{2}{3}$. This is a quite weak bound on the tail probability using Markov's Inequality, since we intuitively know that $X$ should be concentrated very tightly around its mean. (If we toss 10,000 fair coins, we have a sense that the probability of getting 7,500 or more heads is going to be very small.)

To illustrate this point further, consider $\Pr(X \geq n)$. Plugging in $a = n$, we get $\Pr(X \geq n) \leq \frac{n/2}{n} = \frac{1}{2}$. However, we know that $\Pr(X \geq n) = \Pr(X = n) = \frac{1}{2^n}$, since outcomes of *all* $n$ coin tosses must be heads, when $X = n$.

$\square$

The example above illustrates that often, the bounds given by Markov's Inequality are quite weak. This should not be surprising, however, since this bound only makes use of the expected value of a random variable.

## 2 Chebyshev's Inequality

In order to get more information about a random variable, we can use *moments* of a random variable.

**Definition 2 (Moment)** *The $k^{th}$* moment *of a random variable $X$ is $E[X^k]$.*

Higher moments often reveal more information about a random variable, which, in turn helps us derive better bounds. However, there is a trade-off. It is often difficult to compute higher moments in practical cases, e.g., while analyzing randomized algorithms. Now, let us look at the variance of a random variable.

**Definition 3 (Variance)** *The* variance *of a random variable $X$, denoted as* $\mathrm{Var}[X]$, *is $E[(X - E[X])^2]$.*

The variance of a random variable can be seen as the expected square of the distance of $X$, from its expected value $E[X]$. Another way to look at $\mathrm{Var}[X]$ is as follows.

$$\begin{aligned}
\mathrm{Var}[X] &= E\left[(X - E[X])^2\right] \\
&= E\left[X^2 - 2 \cdot X \cdot E[X] + E[X]^2\right] \\
&= E[X^2] - E[2E[X] \cdot X] + E\left[E[X]^2\right] && \text{(Linearity of Expectation)} \\
&= E[X^2] - 2E[X] \cdot E[X] + E[X]^2 && \text{(2 and } E[X] \text{ are constants)} \\
&= E[X^2] - E[X]^2.
\end{aligned}$$

That is, the variance of $X$ equals the difference between the second moment of $X$, and the square of the expected value of $X$ (i.e., the square of the first moment of $X$).

Now, we can derive Chebyshev's Inequality, which often gives much stronger bounds than the Markov's Inequality.

**Theorem 4 (Chebyshev's Inequality)** *For any $a > 0$,*

$$\Pr(|X - E[X]| \geq a) \leq \frac{\mathrm{Var}[X]}{a^2}.$$

Again, let us look at a picture that illustrates Chebyshev's Inequality.

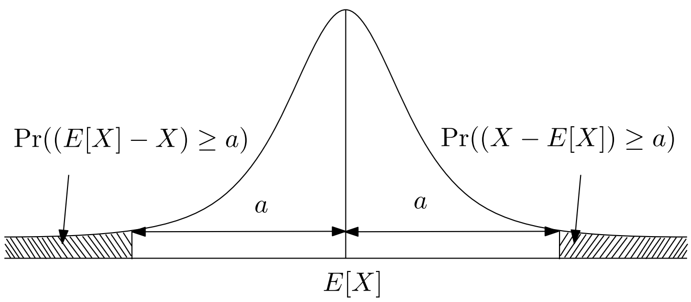


Figure 2: Chebyshev's Inequality bounds the probability of the shaded regions.

**Proof:**

$$\Pr\left(|X - E[X]| \geq a\right) = \Pr\left((X - E[X])^2 \geq a^2\right) = \Pr(Y \geq a^2)$$

Where, $Y = (X - E[X])^2$. Note that $Y$ is a non-negative random variable. Therefore, using Markov's Inequality,

$$\Pr(Y \geq a^2) \leq \frac{E[Y]}{a^2} = \frac{E\left((X - E[X])^2\right)}{a^2} = \frac{\mathrm{Var}[X]}{a^2}.$$

□

**Example.** Again consider the fair coin example. Recall that $X$ denotes the number of heads, when $n$ *fair* coins are tossed independently. We saw that $\Pr(X \geq \frac{3n}{4}) \leq \frac{2}{3}$, using Markov's Inequality. Let us see how Chebyshev's Inequality can be used to give a much stronger bound on this probability. First, notice that:

$$\Pr\left(X \geq \frac{3n}{4}\right) = \Pr\left(X - \frac{n}{2} \geq \frac{n}{4}\right) \leq \Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) = \Pr\left(|X - E[X]| \geq \frac{n}{4}\right).$$

That is, we are interested in bounding the upper tail probability. However, as seen before, Chebyshev's Inequality upper bounds probabilities of both tails. In order to use Chebyshev's Inequality, we must first calculate $\mathrm{Var}[X]$. First, we must characterize what kind of random variable $X$ is.

**Definition 5 (Binomial Random Variable)** *A random variable $X$ is* Binomial *with parameters $n$ and $p$ (denoted as $X \sim \mathrm{Bin}(n,p)$) if $X$ takes on values $0, 1, \ldots, n-1, n$, with the following distribution.*

$$\Pr(X = j) = \binom{n}{j} p^j (1-p)^{n-j}.$$

A binomial random variable $X \sim \mathrm{Bin}(n,p)$ denotes the number of successes (heads), when $n$ independent coins are tossed, with each coin having success (heads) probability of $p$. In our example, $X$ is a binomial random variable with parameters $n$ and $\frac{1}{2}$. However, we consider the more general case.

To compute $\mathrm{Var}[X]$, we need $E[X]$ and $E[X^2]$. For the case of Binomial Random variable, $E[X] = np$ can be computed easily, as seen before. However, computing $E[X^2]$ directly is quite tedious. Therefore, we decompose $X$ is the following manner.

For each coin toss $i = 1, \ldots, n$, define an indicator r.v. $X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$

That is, $X_i$ is 1 if the $i^{\text{th}}$ coin toss is heads, and 0 otherwise. It is easy to see that $X = \sum_{i=1}^{n} X_i$. Before we show how the variance of $X$ can be decomposed, we need the following definition.

**Definition 6 (Covariance)** *The Covariance of random variables $X_i$ and $X_j$, denoted as* $\mathrm{Cov}(X_i, X_j)$, *is* $E\left[(X_i - E[X_i]) \cdot (X_j - E[X_j])\right]$.

$\mathrm{Cov}(X_i, X_j)$ is a measure of correlation between $X_i$ and $X_j$. It immediately follows from the definition that $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$. Another way to look at $\mathrm{Cov}(X_i, X_j)$ is as follows.

$$\begin{aligned}
\mathrm{Cov}(X_i, X_j) &= E\left[(X_i - E[X_i]) \cdot (X_j - E[X_j])\right] \\
&= E\left[X_i X_j - X_i E[X_j] - X_j E[X_i] + E[X_i] E[X_j]\right] \\
&= E[X_i X_j] - E[X_j] \cdot E[X_i] - E[X_i] \cdot E[X_i] + E[E[X_i] \cdot E[X_j]] \\
&\hspace{16em} \text{(Using Linearity of Expectation)} \\
&= E[X_i X_j] - E[X_i] \cdot E[X_j]
\end{aligned}$$

Now, we state the following theorem without proof.

**Theorem 7**

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] + \sum_{\substack{i,j \\ i \neq j}} \mathrm{Cov}(X_i, X_j)$$

Consider the case where all $X_i$'s are mutually independent. Then, for any $X_i, X_j$, $E[X_i X_j] = E[X_i] \cdot E[X_j]$, which implies that $\mathrm{Cov}(X_i, X_j) = 0$. That is,

**Theorem 8 (Linearity of Variance)** *If $X_1, X_2, \ldots, X_n$ are all mutually independent, then*

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i].$$

**Notes.**

1. Linearity of Variance requires the independence of the random variables, whereas Linearity of Expectation does not.

2. We do not need mutual independence between the random variables for Linearity of Variance. A weaker notion called *pairwise independence* suffices. That is, for any distinct $X_i, X_j$, it is sufficient to require $X_i, X_j$ be independent.

**Example**: (Continued) In our coin toss example, all $X_i$'s are in fact mutually independent. Therefore, $\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$.

For any $X_i$, $\mathrm{Var}[X_i] = E[X_i^2] - E[X_i]^2$. $E[X_i] = \Pr(X_i = 1) = p$. Note that $X_i^2$ also has the same distribution as $X_i$, and therefore, $E[X_i^2] = p$. So, $\mathrm{Var}[X_i] = p - p^2 = p(1-p)$.

And therefore, $\mathrm{Var}[X] = np(1-p)$.

**Theorem 9 (Variance of a Binomial Random Variable)** *If* $X \sim \mathrm{Bin}(n, p)$*, then* $\mathrm{Var}[X] = np(1-p)$.

Therefore, in the case of *fair* coin tosses, $\mathrm{Var}[X] = \frac{n}{4}$. By Chebyshev's Inequality,

$$\Pr\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq \frac{(n/4)}{(n/4)^2} = \frac{4}{n}.$$

Recall that Markov's Inequality gave us a much weaker bound of $\frac{2}{3}$ on the same tail probability. Later on, we will discover that using Chernoff Bounds, we can get an even stronger bound of $O\left(\frac{1}{\exp(n)}\right)$ on the same probability. However, Chernoff Bounds require mutual independence, whereas even the weaker notion of pairwise independence suffices for an application of Chebyshev's Inequality.

□