COMP 480/580 — Probabilistic Algorithms and Data Structures Sep 4, 2025
Lecture 3
Scribes: Naga Shiva Harish Yedidi(ny20), Rahul Sai Telikicherla(rt71)
Lecturers: Anshumali Shrivastava, Agniva Chowdhury

1 Probability Basics

1.1 Random Variable

A random variable X is a function that assigns a numerical value to each outcome in a sample space Ω. Formally, if Ω is the sample space of a probabilistic experiment, then a random variable is a measurable function $X : \Omega \to \mathbb{R}$.

A random variable X can be either:

- **Discrete**, taking a countable number of distinct values.
- **Continuous**, taking an uncountable number of values, often represented as intervals on the real line.

Property: For a discrete random variable X with possible values $x_1, x_2, \dots$, the sum of all probabilities must equal 1:

$$\sum_i \Pr(X = x_i) = 1.$$

1.2 Probability Density Function (PDF)

For a continuous random variable X, the probability density function $f_X(x)$ satisfies

$$f_X(x) \geq 0, \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1,$$

and probabilities of intervals are computed as

$$\Pr(a \leq X \leq b) = \int_a^b f_X(x)\,dx.$$

1.3 Expectation and Linearity

The expected value (mean) of a random variable X is given by

$$\mathbb{E}[X] = \sum_x x \cdot p(x) \quad \text{(discrete)}$$
$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,dx \quad \text{(continuous)}.$$

Linearity of expectation holds for all random variables (no independence required):

$$\mathbb{E}\left[\sum_i X_i\right] = \sum_i \mathbb{E}[X_i].$$

1.4 Variance

The variance of X is

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] .$$

Useful properties:

- $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.
- If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
- If $X_1, \ldots, X_n$ are pairwise independent, then $\mathrm{Var}(\sum_i X_i) = \sum_i \mathrm{Var}(X_i)$.

2 Tail Bounds: Markov, Chebyshev, and Chernoff

We are interested in bounding the probability that a random variable deviates significantly from its mean. Expectation alone does not tell us about concentration; variance and independence information enable stronger tail bounds.

2.1 Markov's Inequality

Let X be a non-negative random variable. For any $a > 0$,

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

Proof idea. Define the indicator $I(X \geq a)$. Then $a\, I(X \geq a) \leq X$. Taking expectations gives $a \Pr(X \geq a) \leq \mathbb{E}[X]$.

Example: Geometric distribution (die until first 6)

Let X be the number of rolls until the first success for a fair 6-sided die ($p = 1/6$). Then $\mathbb{E}[X] = 1/p = 6$. Markov gives

$$\Pr(X \geq 12) \leq \frac{6}{12} = 0.5,$$

while the exact probability is $\Pr(X \geq 12) = 1 - \Pr(X \leq 11) \approx 0.1346$. Markov is valid but loose.

Example: Ads on a website

Suppose the average number of ads on a page is 25. Then

$$\Pr(X \geq 75) \leq \frac{25}{75} = \frac{1}{3},$$

and for a threshold below the mean (e.g., 20) Markov only gives a trivial bound ($25/20 > 1$).

2.2 Chebyshev's Inequality

Let X be any random variable with mean μ and variance σ^2. For any $a > 0$,

$$\Pr(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Proof sketch. Let $Z = X - \mu$. Then $\Pr(|Z| \geq a) = \Pr(Z^2 \geq a^2) \leq \frac{\mathbb{E}[Z^2]}{a^2} = \frac{\mathrm{Var}(X)}{a^2}$.

Example: Binomial tails for 200 coin tosses

Let $X \sim \mathrm{Binomial}(n = 200, p = 0.5)$. Then $\mu = 100$ and $\sigma = \sqrt{50} \approx 7.071$. We have

$$\Pr(X \geq 150) = \Pr(X - 100 \geq 50) \leq \Pr(|X - 100| \geq 50) \leq \frac{1}{(50/7.071)^2} \approx 0.02.$$

Example: Geometric again

For $X \sim \mathrm{Geom}(p)$ with $\mu = 1/p$ and $\mathrm{Var}(X) = (1 - p)/p^2$, Chebyshev can improve over Markov in some regimes, but may still be loose for extreme tails.

2.3 Chernoff Bounds

Let $X_1, \ldots, X_n$ be independent Bernoulli random variables and define $X = \sum_{i=1}^{n} X_i$ with mean $\mu = \mathbb{E}[X]$. Then for $\delta \geq 0$,

$$\Pr(X \geq (1 + \delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{2 + \delta}\right),$$

and for $0 \leq \delta \leq 1$,

$$\Pr(X \leq (1 - \delta)\mu) \leq \exp\left(-\frac{\delta^2 \mu}{2}\right).$$

Application: Hash table chain length

If each of n keys hashes independently and uniformly into m slots, the chain length at a slot is $X = \sum_{i=1}^{n} X_i$ where $X_i \sim \mathrm{Bernoulli}(1/m)$ and $\mu = n/m$. Chernoff bounds imply that the probability of observing a chain length much larger than μ decays exponentially in n.

2.4 Comparing the bounds

Inequality	**Inputs required**	**Tail strength**
Markov	Non-negativity + $\mathbb{E}[X]$	Loose (general)
Chebyshev	$\mathbb{E}[X]$, $\mathrm{Var}(X)$	Tighter (polynomial tail)
Chernoff	Independence + $\mathbb{E}[X]$	Strong (exponential tail)

Example At least 150 heads in 200 tosses

We want an upper bound for the probability of obtaining at least 150 heads, i.e., $\Pr(X \geq 150)$, where $X \sim \mathrm{Binomial}(200, 0.5)$. We compare Markov's, Chebyshev's, and Chernoff bounds, and then compute the actual probability.

Markov's Inequality

Markov's inequality states that for any non-negative random variable X and $a > 0$,

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

For $a = 150$ with $\mathbb{E}[X] = 100$,

$$\Pr(X \geq 150) \leq \frac{100}{150} = \frac{2}{3} \approx 0.6667.$$

This bound is quite loose, as Markov's inequality only uses the expectation of X.

Chebyshev's Inequality

Chebyshev's inequality states that for any random variable X with finite mean μ and variance σ^2, and for any $k > 0$,

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

Here, $\mu = 100$, $\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{50} \approx 7.071$. We want $\Pr(X \geq 150) = \Pr(X - 100 \geq 50)$. Define $k = 50/\sigma \approx 50/7.071 \approx 7.07$. Then

$$\Pr(X \geq 150) \leq \Pr(|X - 100| \geq 50) \leq \frac{1}{k^2} \approx \frac{1}{(7.07)^2} \approx \frac{1}{50} = 0.02.$$

Chernoff Bound

For a binomial random variable X, the upper tail Chernoff bound is

$$\Pr(X \geq (1+\delta)\mu) \leq \exp\left(-\frac{\delta^2\mu}{2+\delta}\right),$$

where $\mu = E[X] = 100$ and $\delta = (150 - 100)/100 = 0.5$. Substituting,

$$\Pr(X \geq 150) \leq \exp\left(-\frac{(0.5)^2 \cdot 100}{2+0.5}\right) = \exp\left(-\frac{25}{2.5}\right) = \exp(-10) \approx 0.000045.$$

Chernoff gives an exponentially small probability, much tighter than Markov and Chebyshev.

Conclusion

The upper bounds for $\Pr(X \geq 150)$ using the different inequalities are:

- Markov's Inequality: ≤ 0.6667
- Chebyshev's Inequality: ≤ 0.02
- Chernoff Bound: ≤ 0.000045

This comparison illustrates that Chernoff bounds are significantly tighter for large deviations, especially when the random variable is a sum of independent random variables. Chebyshev's inequality improves over Markov's by incorporating variance, but Chernoff's bound, leveraging independence, provides the sharpest estimate.

Actual Probability Calculation

To compute the actual probability of getting at least 150 heads in 200 fair coin tosses, use the binomial distribution. If $X \sim \mathrm{Binomial}(n = 200, p = 0.5)$, then

$$\Pr(X = k) = \binom{200}{k}(0.5)^{200},$$

and thus

$$\Pr(X \geq 150) = \sum_{k=150}^{200} \binom{200}{k}(0.5)^{200} \approx 2.753 \times 10^{-7}.$$