Lecture 4: Probabilistic Inequalities

September 5, 2025

1 Markov's Inequality

We have already seen Markov's inequality.

Inequality 1 (Markov's Inequality)**.** *Suppose X is a non-negative random variable. Then for any $t > 0$,*

$$Pr[X \geq t \cdot E[X]] \leq \frac{1}{t}$$

Alternatively, for any $\alpha > 0$,

$$Pr[X \geq \alpha] \leq \frac{E[X]}{\alpha}$$

Exercise 1. For any $t \geq 1$, give an example of a random variable where Markov's inequality is tight.

2 Variance

The expectation is the first moment and provides important information, but is often not sufficient to do more sophisticated analysis. Now we will discuss the second moment, or the variance.

Definition 1 (Variance)**.** The variance of a random variable X is defined as:

$$Var(X) = E[(X - E[X])^2]$$

The standard deviation is $stdev(X) = \sqrt{Var(X)}$.

An alternative and useful formula for variance is:

$$Var(X) = E[X^2] - (E[X])^2$$

Example 1 (Binary Random Variable)**.** Let $X = 1$ with probability p and $X = 0$ otherwise.

$$E[X] = p \cdot 1 + (1 - p) \cdot 0 = p$$
$$E[X^2] = p \cdot 1^2 + (1 - p) \cdot 0^2 = p$$
$$Var(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1 - p)$$

Example 2. Let $X = 1$ with probability $1/2$ and $X = -1$ with probability $1/2$.

$$E[X] = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0$$
$$E[X^2] = \frac{1}{2}(1^2) + \frac{1}{2}((-1)^2) = 1$$
$$Var(X) = E[X^2] - (E[X])^2 = 1 - 0^2 = 1$$

Example 3 (Geometric Random Variable)**.** Let X be a geometric random variable with parameter p. This represents the number of coin tosses to get the first head, where the probability of heads is p.

$$E[X] = \frac{1}{p}$$
$$Var(X) = \frac{1 - p}{p^2}$$

Example 4 (Poisson Random Variable)**.** Let X be a Poisson random variable with parameter λ. The probability distribution is $Pr[X = i] = e^{-\lambda}\frac{\lambda^i}{i!}$.

$$E[X] = \lambda$$
$$Var(X) = \lambda$$

Example 5 (Normal Distribution)**.** For a normal distribution with probability density function $f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$:

$$E[X] = \mu$$
$$Var(X) = \sigma^2$$

Lemma 1. *If X_1 and X_2 are independent random variables, then $Var(X_1+X_2) = Var(X_1)+Var(X_2)$.*

Proof Sketch. We know $E[X_1+X_2] = E[X_1]+E[X_2]$. The variance is $Var(X_1+X_2) = E[(X_1+X_2)^2] - (E[X_1+X_2])^2$. Expanding the terms:

$$E[(X_1 + X_2)^2] = E[X_1^2 + 2X_1X_2 + X_2^2] = E[X_1^2] + 2E[X_1X_2] + E[X_2^2]$$
$$(E[X_1 + X_2])^2 = (E[X_1] + E[X_2])^2 = (E[X_1])^2 + 2E[X_1]E[X_2] + (E[X_2])^2$$

Because X_1 and X_2 are independent, $E[X_1X_2] = E[X_1]E[X_2]$. The result follows from algebra. □

More generally, for a sum of n mutually independent random variables $X = \sum_{i=1}^{n} X_i$:

$$E[X] = \sum_{i=1}^{n} E[X_i] \quad \text{(by linearity of expectation)}$$
$$Var(X) = \sum_{i=1}^{n} Var(X_i) \quad \text{(due to independence)}$$

3 Chebyshev's Inequality

Inequality 2 (Chebyshev's Inequality)**.** *Suppose X is a random variable with variance σ_X^2. Then for any $t > 0$,*

$$Pr[|X - E[X]| \geq t\sigma_X] \leq \frac{1}{t^2}$$

Alternatively, for any $\gamma > 0$,

$$Pr[|X - E[X]| \geq \gamma] \leq \frac{\sigma_X^2}{\gamma^2}$$

Proof. Let $Y = (X - E[X])^2$. By definition, $E[Y] = \sigma_X^2$. Y is a non-negative random variable. We can apply Markov's inequality to Y.

$$\begin{aligned} Pr[|X - E[X]| \geq t\sigma_X] &= Pr[(X - E[X])^2 \geq t^2\sigma_X^2] \\ &= Pr[Y \geq t^2\sigma_X^2] \\ &\leq \frac{E[Y]}{t^2\sigma_X^2} \quad \text{(by Markov's inequality)} \\ &= \frac{\sigma_X^2}{t^2\sigma_X^2} = \frac{1}{t^2} \end{aligned}$$

□

4 Applications of Chebyshev's Inequality

4.1 Variance Reduction

Suppose we have a randomized procedure that outputs a random variable X to estimate a quantity of interest α, such that $E[X] = \alpha$. This may not be adequate if $Var(X)$ is large.

To improve the estimate, we can run the algorithm k times independently, obtaining estimates $X_1, X_2, \ldots, X_k$. The final estimate is the average:

$$\hat{X} = \frac{1}{k}\sum_{i=1}^{k} X_i$$

The expectation remains correct:

$$E[\hat{X}] = E\left[\frac{1}{k}\sum_{i=1}^{k} X_i\right] = \frac{1}{k}\sum_{i=1}^{k} E[X_i] = \frac{1}{k}(k\alpha) = \alpha$$

The variance is reduced:

$$Var(\hat{X}) = Var\left(\frac{1}{k}\sum_{i=1}^{k} X_i\right) = \frac{1}{k^2}\sum_{i=1}^{k} Var(X_i) = \frac{1}{k^2}(k \cdot Var(X)) = \frac{Var(X)}{k}$$

Repetition and averaging reduces variance.

4.2 Balls and Bins

Consider throwing m balls independently and uniformly at random into n bins. Let Y_i be the load in bin i (number of balls in bin i). Let X_{ij} be an indicator variable for ball j landing in bin i. $Pr[X_{ij} = 1] = 1/n$. Then $Y_i = \sum_{j=1}^{m} X_{ij}$. The X_{ij} for $j = 1, \ldots, m$ are independent.

$$E[Y_i] = \sum_{j=1}^{m} E[X_{ij}] = m \cdot \frac{1}{n} = \frac{m}{n}$$
$$Var(Y_i) = \sum_{j=1}^{m} Var(X_{ij}) = m \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right) \approx \frac{m}{n}$$

A key quantity of interest is the maximum load over all bins, $Z = \max_{i=1}^{n} Y_i$. How unevenly are the balls spread? What is $E[Z]$?

Let's focus on the case $m = n$. Then $E[Y_i] = 1$ and $Var(Y_i) = 1(1 - 1/n) < 1$. The variables Y_i are correlated, making direct analysis difficult. We can use a combination of a deviation inequality and the union bound.

Lemma 2 (Union Bound)**.** *Let $A_1, \ldots, A_n$ be events. Then $Pr[A_1 \cup A_2 \cup \cdots \cup A_n] \le \sum_{i=1}^{n} Pr[A_i]$.*

Let's apply Chebyshev's inequality to Y_i for $m = n$: $Pr[|Y_i - E[Y_i]| \ge \gamma] \le \frac{Var(Y_i)}{\gamma^2} < \frac{1}{\gamma^2}$. Let $\gamma = t\sqrt{n}$. We have $Pr[|Y_i - 1| \ge t\sqrt{n}] \le \frac{1}{t^2 n}$. This implies $Pr[Y_i \ge t\sqrt{n} + 1] \le Pr[|Y_i - 1| \ge t\sqrt{n}] \le \frac{1}{t^2 n}$.

Now, let's bound the probability of the maximum load.

$$\begin{aligned} Pr[Z \ge t\sqrt{n} + 1] &= Pr[\exists i, Y_i \ge t\sqrt{n} + 1] \\ &= Pr[\cup_{i=1}^{n}\{Y_i \ge t\sqrt{n} + 1\}] \\ &\le \sum_{i=1}^{n} Pr[Y_i \ge t\sqrt{n} + 1] \quad \text{(by Union Bound)} \\ &\le \sum_{i=1}^{n} \frac{1}{t^2 n} = n \cdot \frac{1}{t^2 n} = \frac{1}{t^2} \end{aligned}$$

Exercise 2. Show that $E[Z] \le c\sqrt{n}$ for some fixed constant c.

4.3 Random Walk on the Line

Start at the origin. At each step, walk one unit to the right or left with equal probability (1/2). Let $X_i \in \{-1, +1\}$ be the i-th step. Let $Y_n = \sum_{i=1}^{n} X_i$ be the position after n steps.

$$\begin{aligned} E[X_i] = 0, \quad Var(X_i) &= E[X_i^2] - (E[X_i])^2 = 1 - 0 = 1 \\ E[Y_n] &= \sum E[X_i] = 0 \\ Var(Y_n) &= \sum Var(X_i) = n \end{aligned}$$

The standard deviation is $\sigma_{Y_n} = \sqrt{n}$. Using Chebyshev's inequality:

$$Pr[|Y_n - E[Y_n]| \geq t\sqrt{n}] = Pr[|Y_n| \geq t\sqrt{n}] \leq \frac{1}{t^2}$$

This implies that the expected distance from the origin, $E[|Y_n|]$, is $O(\sqrt{n})$.

5 Chernoff-Hoeffding Bounds

Chernoff-Hoeffding bounds are very useful probabilistic inequalities that provide tight bounds in many settings. To motivate, they give much better results than Chebyshev's inequality for the previous examples:

- **Balls and Bins ($m = n$):** $E[Z] = O\left(\frac{\log n}{\log\log n}\right)$, much better than $O(\sqrt{n})$.
- **Random Walk:** $Pr[|Y_n| \geq t\sqrt{n}] \leq 2e^{-t^2/2}$, which is much stronger than $\leq \frac{1}{t^2}$.

5.1 Chernoff Bounds for Sums of Binary Variables

Theorem 1. *Let $X = \sum_{i=1}^{n} X_i$, where X_i are independent binary random variables ($X_i \in \{0,1\}$). Let $\mu = E[X]$. Then:*

1. ***Upper Tail:*** *For any $\delta > 0$,*

$$Pr[X \geq (1+\delta)\mu] \leq \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}$$

2. ***Lower Tail:*** *For $\delta \in (0,1)$,*

$$Pr[X \leq (1-\delta)\mu] \leq \left[\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right]^{\mu}$$

These are often called multiplicative Chernoff bounds.

5.1.1 Proof of Chernoff Bound (Upper Tail)

The key idea is to use the moment generating function and apply Markov's inequality. For any $t > 0$:

$$Pr[X \geq (1+\delta)\mu] = Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}}$$

Now we bound $E[e^{tX}]$:

$$\begin{aligned} E[e^{tX}] &= E\left[e^{t\sum X_i}\right] = E\left[\prod e^{tX_i}\right] \\ &= \prod E[e^{tX_i}] \quad \text{(by independence)} \end{aligned}$$

Let $E[X_i] = p_i$. Then $E[e^{tX_i}] = p_i e^{t\cdot 1} + (1-p_i)e^{t\cdot 0} = p_i e^t + 1 - p_i = 1 + p_i(e^t - 1)$. Using the inequality $1 + x \leq e^x$:

$$E[e^{tX_i}] \leq e^{p_i(e^t-1)}$$

So, $\prod E[e^{tX_i}] \leq \prod e^{p_i(e^t-1)} = e^{\sum p_i(e^t-1)} = e^{(e^t-1)\sum p_i} = e^{(e^t-1)\mu}$. Plugging this back into the Markov bound:

$$Pr[X \geq (1+\delta)\mu] \leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = \left[\frac{e^{e^t-1}}{e^{t(1+\delta)}}\right]^{\mu}$$

We choose t to minimize the expression. Let $g(t) = e^t - 1 - t(1+\delta)$. $g'(t) = e^t - (1+\delta) = 0 \implies e^t = 1+\delta \implies t = \ln(1+\delta)$. Substituting this value of t gives the exponent: $(1+\delta) - 1 - (1+\delta)\ln(1+\delta) = \delta - (1+\delta)\ln(1+\delta)$. The bound becomes:

$$\left[e^{\delta-(1+\delta)\ln(1+\delta)}\right]^{\mu} = \left[\frac{e^{\delta}}{e^{(1+\delta)\ln(1+\delta)}}\right]^{\mu} = \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}$$

The proof for the lower tail is similar, using $Pr[X \leq (1-\delta)\mu] = Pr[e^{-tX} \geq e^{-t(1-\delta)\mu}]$ for $t > 0$.

5.1.2 Simplified Chernoff Bounds

The standard forms can be simplified into more convenient bounds. For $\mu = E[X]$:

- **Lower Tail:** For $0 < \delta < 1$,
$$Pr[X \leq (1-\delta)\mu] \leq e^{-\mu\delta^2/2}$$
- **Upper Tail:** For $0 < \delta < 1$,
$$Pr[X \geq (1+\delta)\mu] \leq e^{-\mu\delta^2/3}$$
For $\delta \geq 1$ (large deviations),
$$Pr[X \geq (1+\delta)\mu] \leq e^{-\mu(1+\delta)\ln(1+\delta)/2} \approx e^{-c\delta\ln(\delta)\mu}$$

5.2 Hoeffding's Inequality (General Case)

This is an additive form of the bound for variables bounded in an interval.

Theorem 2 (Hoeffding's Inequality)**.** *Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \in [a_i, b_i]$. Let $X = \sum_{i=1}^{n} X_i$. Then for any $a > 0$,*

$$Pr[|X - E[X]| \geq a] \leq 2e^{-\frac{2a^2}{\sum_{i=1}^{n}(b_i-a_i)^2}}$$

A common special case is for $X_i \in \{-1, 1\}$, where $b_i - a_i = 2$.

$$Pr[|X - E[X]| \geq a] \leq 2e^{-\frac{a^2}{2n}}$$

Note that this bound depends on n and a, and is useful when $E[X]$ can be 0.

6 Applications of Chernoff-Hoeffding Bounds

6.1 Balls and Bins (revisited)

For $m = n$ balls into n bins, Y_i is the load on bin i. $E[Y_i] = 1$. Let's bound the probability that $Y_i \geq 10\frac{\ln n}{\ln\ln n}$. Here $\mu = 1$. Let $(1+\delta)\mu = 10\frac{\ln n}{\ln\ln n}$. This implies δ is large. We use the large deviation upper tail bound:

$$Pr[Y_i \geq (1+\delta)\mu] \leq e^{-\frac{(1+\delta)\ln(1+\delta)\mu}{2}}$$

We have $(1+\delta)\mu = 10\frac{\ln n}{\ln\ln n}$. For large n, $\ln(1+\delta) \approx \ln(10\frac{\ln n}{\ln\ln n}) \approx \ln\ln n$. So, $(1+\delta)\ln(1+\delta)\mu \approx (10\frac{\ln n}{\ln\ln n})(\ln\ln n) = 10\ln n$.

$$Pr\left[Y_i \geq 10\frac{\ln n}{\ln\ln n}\right] \leq e^{-\frac{10\ln n}{2}} = e^{-5\ln n} = \frac{1}{n^5}$$

By the union bound, the probability that the max load exceeds this value is:

$$Pr\left[Z \geq 10\frac{\ln n}{\ln\ln n}\right] \leq \sum_{i=1}^{n} Pr\left[Y_i \geq 10\frac{\ln n}{\ln\ln n}\right] \leq n \cdot \frac{1}{n^5} = \frac{1}{n^4}$$

This shows that with high probability, the maximum load is $O\left(\frac{\ln n}{\ln\ln n}\right)$.

6.2 Randomized Quicksort

We saw that randomized quicksort runs in expected time $O(n\ln n)$. We can show that the runtime is tightly concentrated around this mean. Let $Q(A)$ be the number of comparisons on an array A of size n. We will show that with high probability, the recursion depth is bounded.

Consider a fixed element a. At each level of recursion, a is in some sub-array S_i. We call level i "lucky" for a if the chosen pivot splits S_i such that $|S_{i+1}| \leq \frac{3}{4}|S_i|$. The probability of a lucky split is at least 1/2. After $h = 32\ln n$ levels, we want to know the probability that a is still in a sub-array of size > 1. For this to happen, a must have had very few lucky splits. The number of lucky splits needed to reduce an array of size n to size 1 is roughly $\log_{4/3} n \approx 3.4\ln n$. Let's consider $h = 32\ln n$ coin tosses (for lucky/unlucky splits). The expected number of heads (lucky splits) is $\mu = \frac{1}{2}h = 16 \ln n$. The recursion depth for element a is greater than h only if the number of lucky splits in h rounds is less than $\approx 4 \ln n$. Let Y be the number of lucky splits. We want to bound $Pr[Y < 4 \ln n]$. Here, $\mu = 16 \ln n$. $4 \ln n = (1 - \frac{3}{4})\mu$. So $\delta = 3/4$. Using the simplified lower tail bound:

$$Pr[Y \leq (1 - \delta)\mu] \leq e^{-\mu\delta^2/2}$$

$$Pr[Y < 4 \ln n] \leq e^{-(16 \ln n)(3/4)^2/2} = e^{-(16 \ln n)(9/16)/2} = e^{-9/2 \ln n} = \frac{1}{n^{4.5}}$$

(Note: the lecture notes use different constants leading to $1/n^9$, but the principle is the same). By the union bound over all n elements, the probability that any element has a recursion depth $> 32 \ln n$ is bounded by $n \cdot \frac{1}{n^{4.5}} = \frac{1}{n^{3.5}}$, which is very small. Since the number of comparisons at each level is at most n, the total number of comparisons is bounded by (depth) $\times\, n = O(n \ln n)$ with high probability.

6.3 Random Walk (revisited)

Using Hoeffding's inequality for $Y_n = \sum_{i=1}^{n} X_i$ where $X_i \in \{-1, 1\}$ and $E[Y_n] = 0$. Let $a = t\sqrt{n}$.

$$Pr[|Y_n| \geq t\sqrt{n}] \leq 2e^{-\frac{2(t\sqrt{n})^2}{\sum_{i=1}^{n}(1-(-1))^2}} = 2e^{-\frac{2t^2 n}{\sum 4}} = 2e^{-\frac{2t^2 n}{4n}} = 2e^{-t^2/2}$$

This is an exponential decay in t^2, which is significantly stronger than the $1/t^2$ decay from Chebyshev's inequality.