Lecture 1

1 Concentration Inequalities

Theorem 1.1 (Markov's inequality). *Let X be a non-negative random variable. Then, for every real number $t > 0$, it holds that*

$$\mathbb{P}\left(X \geq t\right) \leq \frac{\mathbb{E}\left(X\right)}{t}.$$

Proof. Fix some $t > 0$ and let I_t denote the indicator random variable for the event "$X \geq t$", i.e., $I_t = 1$ if $X \geq t$ and $I_t = 0$ if $X < t$. Observe that, by definition,

$$X \geq t \cdot I_t \tag{1}$$

Hence

$$t \cdot \mathbb{P}\left(X \geq t\right) = t \cdot \mathbb{P}\left(I_t = 1\right) = t \cdot \mathbb{E}\left(I_t\right) = \mathbb{E}\left(t \cdot I_t\right) \leq \mathbb{E}\left(X\right),$$

where the last equality holds by the linearity of expectation and the inequality holds by (1) and by the monotonicity of expectation. $\square$

Remark 1.2. *In general, Markov's inequality is best possible. Indeed, for any real number $k \geq 1$, let*

$$X_k \sim \begin{cases} k & 1/k \\ 0 & 1 - 1/k \end{cases}$$

Then

$$\mathbb{E}\left(X_k\right) = k \cdot 1/k + 0 \cdot (1 - 1/k) = 1$$

and

$$\mathbb{P}\left(X_k \geq k\right) = \mathbb{P}\left(X_k = k\right) = \frac{1}{k} = \frac{\mathbb{E}\left(X\right)}{k}.$$

Theorem 1.3 (Chebyshev's inequality). *Let X be a random variable with finite variance. Then, for every real number $t > 0$, it holds that*

$$\mathbb{P}\left(\left|X - \mathbb{E}\left(X\right)\right| \geq t\right) \leq \frac{\mathrm{Var}\left(X\right)}{t^2}.$$

Proof. Since $\left(X - \mathbb{E}\left(X\right)\right)^2$ is a non-negative random variable, we can apply Markov's inequality to obtain

$$\mathbb{P}\left(\left|X - \mathbb{E}\left(X\right)\right| \geq t\right) = \mathbb{P}\left(\left(X - \mathbb{E}\left(X\right)\right)^2 \geq t^2\right) \leq \frac{\mathbb{E}\left(\left(X - \mathbb{E}\left(X\right)\right)^2\right)}{t^2} = \frac{\mathrm{Var}\left(X\right)}{t^2},$$

where the first equality holds since $\left|X - \mathbb{E}\left(X\right)\right| \geq t$ if and only if $\left(X - \mathbb{E}\left(X\right)\right)^2 \geq t^2$. $\square$

Remark 1.4. *For $t = \lambda\sigma_X$, where $\lambda > 0$ is a real number, Chebyshev's inequality implies that*

$$\mathbb{P}\left(|X - \mathbb{E}\left(X\right)| \geq \lambda\sigma_X\right) \leq \frac{\mathrm{Var}\left(X\right)}{\lambda^2 \cdot \mathrm{Var}\left(X\right)} = \frac{1}{\lambda^2}.$$

That is, the probability that X deviates from its expectation by λ standard deviations decreases quadratically in λ.

Example 1: The number of tables manufactured in some factory at any given month is a random variable X with expected value 100 and variance 100. We would like to establish upper bounds on $\mathbb{P}\left(X \geq 120\right)$. We first apply Markov's inequality to obtain

$$\mathbb{P}\left(X \geq 120\right) \leq \frac{\mathbb{E}(X)}{120} = \frac{100}{120} = \frac{5}{6}.$$

Next, we apply Chebyshev's inequality to obtain

$$\mathbb{P}\left(X \geq 120\right) = \mathbb{P}\left(X - 100 \geq 20\right) \leq \mathbb{P}\left(|X - 100| \geq 20\right) \leq \frac{Var(X)}{20^2} = \frac{100}{400} = \frac{1}{4}.$$

Since Chebyshev's inequality uses more information on X (i.e., its expectation and variance), it is not surprising that it yields a better bound.

Example 2: Toss a fair coin 1000 times, all coin tosses being mutually independent. Let X be the total number of coin tosses (of these 1000 tosses) whose outcome is heads. Intuitively, we expect X to be roughly 500. This intuition is made precise by Chebyshev's inequality. Observe that $X \sim \mathrm{Bin}\left(1000, 1/2\right)$ and thus $\mathbb{E}\left(X\right) = 1000 \cdot 1/2 = 500$ and $\mathrm{Var}\left(X\right) = 1000 \cdot 1/2 \cdot (1 - 1/2) = 250$. Hence

$$\begin{aligned}\mathbb{P}\left(450 < X < 550\right) &= 1 - \mathbb{P}\left(X \leq 450 \text{ or } X \geq 550\right) = 1 - \mathbb{P}\left(|X - 500| \geq 50\right) \\ &= 1 - \mathbb{P}\left(|X - \mathbb{E}\left(X\right)| \geq 50\right) \geq 1 - \frac{\mathrm{Var}\left(X\right)}{50^2} = 1 - \frac{250}{2500} = 0.9,\end{aligned}$$

where the inequality holds by Chebyshev's inequality (Theorem 1.3).

Example 3: Given a coin with probability p to come up heads in any single toss, we would like to determine p with high probability up to some small error. That is, we want to have an algorithm which, given the coin and two parameters $\varepsilon > 0$ and $\delta > 0$, outputs a real number q such that $|q - p| < \varepsilon$ holds with probability at least $1 - \delta$.

Our algorithm is very simple. We will flip the coin n times for some $n(\varepsilon, \delta)$ which will be determined later. If the number of coin flips whose outcome is heads will be m, then the algorithm will output m/n. Let X denote the number of coin flips whose outcome is heads.

Clearly $X \sim Bin(n,p)$; in particular $\mathbb{E}(X) = np$ and $Var(X) = np(1-p)$. It thus follows by Chebyshev's inequality that

$$\mathbb{P}\left(|X/n - p| \geq \varepsilon\right) = \mathbb{P}\left(|X - np| \geq \varepsilon n\right) = \mathbb{P}\left(|X - \mathbb{E}(X)| \geq \varepsilon n\right) \leq \frac{Var(X)}{\varepsilon^2 n^2} = \frac{p(1-p)}{\varepsilon^2 n}.$$

Since $p(1-p) \leq \min\{p, 1-p\} \leq 1/2$, by taking n to be any integer satisfying $n > \frac{1}{2\varepsilon^2 \delta}$, we obtain $\mathbb{P}\left(|X/n - p| \geq \varepsilon\right) < \delta$ as required.

The following concentration inequalities are less general but much stronger than Theorems 1.1 and 1.3.

Theorem 1.5 (Chernoff inequalities)**.** *Let $X_1, \ldots, X_n$ be independent random variables that return values in $[0,1]$, and let $X = \sum_{i=1}^n X_i$. Then*

(a) *For every $t > 0$ it holds that*

$$\mathbb{P}\left(X \geq \mathbb{E}(X) + t\right) \leq e^{-2t^2/n} \text{ and } \mathbb{P}\left(X \leq \mathbb{E}(X) - t\right) \leq e^{-2t^2/n}.$$

(b) *For every $\varepsilon > 0$ it holds that*

$$\mathbb{P}\left(X \leq (1-\varepsilon)\mathbb{E}(X)\right) \leq e^{-\varepsilon^2 \mathbb{E}(X)/2}.$$

(c) *If, moreover, $\varepsilon \leq 3/2$, then*

$$\mathbb{P}\left(X \geq (1+\varepsilon)\mathbb{E}(X)\right) \leq e^{-\varepsilon^2 \mathbb{E}(X)/3}.$$

Let's take another look at Example 2, this time using Chernoff's inequalities. We have

$$\mathbb{P}\left(X \leq 450\right) = \mathbb{P}\left(X \leq 500 - 50\right) = \mathbb{P}\left(X \leq \mathbb{E}(X) - 50\right) \leq e^{-2 \cdot 50^2/1000} = e^{-5}$$

and, similarly,

$$\mathbb{P}\left(X \geq 550\right) = \mathbb{P}\left(X \geq 500 + 50\right) = \mathbb{P}\left(X \geq \mathbb{E}(X) + 50\right) \leq e^{-2 \cdot 50^2/1000} = e^{-5}.$$

Hence

$$\mathbb{P}\left(450 < X < 550\right) = 1 - \left(\mathbb{P}\left(X \leq 450\right) + \mathbb{P}\left(X \geq 550\right)\right) \geq 1 - 2e^{-5} \approx 0.98652.$$

We see that the bound we obtained using Theorem 1.5 is much better than the bound we obtained using Theorem 1.3.

We will not prove Theorem 1.5 in this course. Instead we will prove a simpler version (known as the Chernoff-Hoeffding inequalities) whose proof captures some of the main ideas of the proof of Theorem 1.5.

Theorem 1.6 (Chernoff-Hoeffding inequalities)**.** *Let $X_1, \ldots, X_n$ be independent random variables such that $\mathbb{P}\left(X_i = 1\right) = \mathbb{P}\left(X_i = -1\right) = 1/2$ for every $1 \leq i \leq n$, and let $X = \sum_{i=1}^n X_i$. Then, for every $t > 0$ it holds that*

$$\mathbb{P}\left(X \geq t\right) \leq e^{-t^2/(2n)},$$
$$\mathbb{P}\left(X \leq -t\right) \leq e^{-t^2/(2n)}.$$