University of Edinburgh

INFR11156: Algorithmic Foundations of Data Science (2025)

Lecture 1: Review on Probability Theory

1 The Law of Large Numbers

We first look at the fundamental probability inequalities that will be used in our course.

Theorem 1 (Markov Inequality)**.** *Let X be a non-negative random variable. Then, for any $c > 0$ it holds that*

$$\mathbf{Pr}\left[X \geq c \right] \leq \frac{\mathbf{E}\left[X \right]}{c}.$$

Proof. By definition, it holds that

$$\begin{aligned}\mathbf{E}\left[X \right] = \int_0^\infty x \cdot p(x)\mathrm{d}x &= \int_0^c x \cdot p(x)\mathrm{d}x + \int_c^\infty x \cdot p(x)\mathrm{d}x \\ &\geq \int_c^\infty x \cdot p(x)\mathrm{d}x \geq c \cdot \int_c^\infty p(x)\mathrm{d}x = c \cdot \mathbf{Pr}\left[X \geq c \right].\end{aligned}$$

This implies that $\mathbf{Pr}\left[X \geq c \right] \leq \frac{\mathbf{E}[X]}{c}$. □

Theorem 2 (Chebyshev's Inequality)**.** *Let X be a random variable. Then for any $c > 0$ it holds that*

$$\mathbf{Pr}\left[|X - \mathbf{E}\left[X \right]| > c \right] \leq \frac{\mathbf{Var}\left[X \right]}{c^2}.$$

Proof. Since $|X - \mathbf{E}\left[X \right]| \geq c$ if and only if $|X - \mathbf{E}\left[X \right]|^2 \geq c^2$, we have that

$$\begin{aligned}\mathbf{Pr}\left[|X - \mathbf{E}\left[X \right]| \geq c \right] &= \mathbf{Pr}\left[|X - \mathbf{E}\left[X \right]|^2 \geq c^2 \right] \\ &\leq \frac{\mathbf{E}\left[|X - \mathbf{E}\left[X \right]|^2 \right]}{c^2} \\ &= \frac{\mathbf{Var}\left[X \right]}{c^2}\end{aligned}$$

where the first inequality follows from the Markov inequality. □

Theorem 3 (Law of Large Numbers)**.** *Let $x_1, \cdots x_n$ be n independent samples of a random variable X. Then, it holds that*

$$\mathbf{Pr}\left[\left| \frac{x_1 + \cdots + x_n}{n} - \mathbf{E}\left[X \right] \right| \geq \varepsilon \right] \leq \frac{\mathbf{Var}\left[X \right]}{n\varepsilon^2}. \tag{1}$$

Proof. By Chebyshev's Inequality, it holds that

$$\mathbf{Pr}\left[\left| \frac{x_1 + \cdots + x_n}{n} - \mathbf{E}\left[X \right] \right| \geq \varepsilon \right] \leq \frac{\mathbf{Var}\left(\frac{x_1 + \cdots + x_n}{n} \right)}{\varepsilon^2} = \frac{\mathbf{Var}\left(x_1 + \cdots + x_n \right)}{n^2\varepsilon^2} = \frac{\mathbf{Var}\left[X \right]}{n\varepsilon^2}.$$

□

Remarks on the Law of Large Numbers.

1. Number of samples (n) is in the denominator of right hand side in (1), which means that the more samples we take, the smaller error we have.

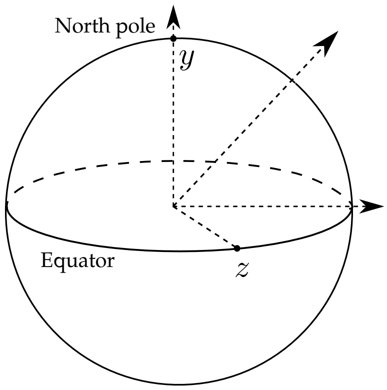

Figure 1: Let y and z be two random points at unit distance from origin, and let y be the north pole. Then it is very likely that z is near the equator.

2. Parameter ε is also in the denominator of the right hand side, which means the larger ε is, the smaller the error is.

By the Law of Large Numbers, we know that with high probability the average of samples will be close to the expectation of the random variable. Now we look at an application of the Law of Large Numbers. Assume that $y = (y_1, \cdots, y_d)$ and $z = (z_1, \cdots, z_d)$ are two random points drawn from d-dimensional random Gaussian with unit variance in each direction. Then, it holds that

$$\mathbf{E}\left[\, y_i^2 \,\right] = \mathbf{E}\left[\, |y_i - \mathbf{E}[\, y_i \,]|^2 \,\right] = \mathbf{Var}[y_i] = 1$$

and $\mathbf{E}\left[\, z_i^2 \,\right] = 1$ for the same reason. By linearity of expectation, we have $\mathbf{E}\left[\, \|y\|^2 \,\right] = d$ and $\mathbf{E}\left[\, \|z\|^2 \,\right] = d$. We apply the Law of Large Numbers and know that $\|y\|^2 \approx d$ and $\|z\|^2 \approx d$ with high probability. On the other hand, we have that

$$\mathbf{E}\left[\, (y_i - z_i)^2 \,\right] = \mathbf{E}\left[\, y_i^2 - 2 \cdot y_i \cdot z_i + z_i^2 \,\right] = \mathbf{E}\left[\, y_i^2 \,\right] - 2 \cdot \mathbf{E}[\, y_i \,] \cdot \mathbf{E}[\, z_i \,] + \mathbf{E}\left[\, z_i^2 \,\right] = 2.$$

This gives us that

$$\|y - z\|^2 \approx 2d \approx \|y\|^2 + \|z\|^2 \,,$$

i.e., y and z must be approximately orthogonal, see Figure 1 for illustration.