Lecture 2: Basic probability

Let X be a random variable on probability space

Expectation and Variance

$$\mathbb{E}X, \quad \mathbb{E}[X|Y]$$

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

Moment Generating Function:

$$M_X(t) = \mathbb{E}e^{tX}, \quad t \in \mathbb{R}$$

L^p-norm $\quad \|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}, \quad p \in (0,\infty)$

L^p-space $\quad L^p = \{X : \|X\|_p < \infty\}$ Banach Space

$$\|XY\|_1 \leq \|X\|_p \cdot \|X\|_q \text{ where } \frac{1}{p} + \frac{1}{q} = 1$$

Hölder's Inequality, Cauchy-Schwartz $\left(\begin{smallmatrix} p=2 \\ q=2 \end{smallmatrix}\right)$

Remark

$$\langle X, Y \rangle_2 = \mathbb{E}[XY], \quad \|X\|_2 = \sqrt{\langle X, X \rangle} = \sqrt{\mathbb{E}[X^2]}$$

Then

$$\|X - \mathbb{E}X\|_2 = \sqrt{Var(X)} \quad \text{and}$$

$$Cov(X,Y) := \mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y) = \langle X-\mathbb{E}X, Y-\mathbb{E}Y \rangle_2$$

Important Distributions:

① Uniform on $[0,1]$:

② Gaussian/Normal $N(\mu, \sigma^2)$

Density $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

③ Rademacher:

$$p(x=1) = p(x=-1) = \frac{1}{2}$$

④ Bernoulli(p):

$$p(x=1) = p$$
$$p(x=0) = 1-p$$

⑤ Poisson (λ):

$$p(x=k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad \forall_{k=0,1,2,\dots}$$

A few basic facts:

① If r.v. X takes positive values, then

$$\mathbb{E}X = \int_0^\infty P[X \geq t]\, dt$$

② (Linearity of expectation)

$$\mathbb{E}\left[\sum_{i=1}^{k} c_i X_i\right] = \sum_{i=1}^{k} c_i\, \mathbb{E}X_i$$

③ $$Var(aX) = a^2\, Var(X)$$

④ (Jensen) If F is convex, then

$$\mathbb{E}F(X) \geq F(\mathbb{E}X)$$

Defn: A family $(X_1, \ldots, X_k)$ is independent, if

$$P[X_i \in E_i \ \forall_{i=1,\ldots,k}] = \prod_{i=1}^{k} P[X_i \in E_i]$$

⑤ (Linearity of Variance) If $X_1, \ldots, X_k$ are pairwise independent, then

$$\mathrm{Var}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} \mathrm{Var}(X_i)$$

⑥ (Tower Rule)

$$\mathbb{E}(X) = \mathbb{E}\left[\mathbb{E}[X \mid Y]\right]$$

⑦ (Markov) For any non-negative X and $t > 0$, we have

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}X}{t}$$

Pf: $\mathbb{E}X = \mathbb{E}\, X \mathbb{1}_{\{X \geq t\}} + \mathbb{E} X \mathbb{1}_{\{X < t\}}$

$$\geq t\, \mathbb{E} \mathbb{1}_{\{X \geq t\}} = t P[X \geq t] \quad \blacksquare$$

⑧ (Chebychev) For any r.v. X have

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

Pf: Apply Markov to $P[(X - \mathbb{E}X)^2 \geq t^2]$ $\blacksquare$

⑨ (Paley–Zymond) If $X \geq 0$ is a random variable, then for any $\theta \in [0,1]$:

$$P[X \geq \theta \mathbb{E} X] \geq (1-\theta)^2 \frac{[\mathbb{E} X]^2}{[\mathbb{E} X^2]}$$

Pf: Write

$$\mathbb{E} X = \mathbb{E}[X \mathbb{1}_{X \leq \theta \mathbb{E} X}] + \mathbb{E}[X \mathbb{1}_{X > \theta \mathbb{E} X}]$$

$$\leq \theta \mathbb{E} X + (\mathbb{E} X^2)^{1/2} \underbrace{(\mathbb{E} \mathbb{1}^2_{X \geq \theta \mathbb{E} X})}_{= P[X \geq \theta \mathbb{E} X]}{}^{1/2}$$

Cauchy Schwartz

Rearrange to get the result. ☒

⑩ (Gaussian Tails) $g \sim N(0,1)$ satisfies

$$P[g \geq t] \leq \frac{1}{2} e^{-t^2/2} \quad \forall t \geq 0$$

Pf: Recall

$$P[g \geq t] = \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$$

Change variables $x = t + y$ to get

$$P[g \geq t] = \int_0^\infty \underbrace{\frac{1}{\sqrt{2\pi}} e^{-t^2/2}} e^{-ty} e^{-y^2/2} \, dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \underbrace{\int_0^\infty \underbrace{e^{-ty}}_{\leq 1} e^{-y^2/2} \, dy}_{\leq \int_0^\infty e^{-y^2/2} dy = \sqrt{\frac{\pi}{2}}}$$

as needed ☒

Limit Theorems

<u>Thm:</u> (Strong Law of Large Numbers)
Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}X_i = \mu$. Then $S_n = \sum_{i=1}^{n} X_i$ satisfies

$$\frac{1}{n} S_n \to \mu \quad \text{almost surely}$$

$$\left[\lim_{n\to\infty} \frac{S_n}{n} = \mu \text{ w.p. } 1\right]$$

<u>Intuition:</u> $Var(S_n) = \frac{Var(X_1)}{n}$

<u>Thm:</u> (Central Limit Theorem)
Let $X_1, X_2, \ldots$ be i.i.d. with

$$\mathbb{E}X_i = \mu, \quad Var\, X_i = \sigma^2$$

Define $$Z_n = \frac{S_n - \mathbb{E}S_n}{\sqrt{Var(S_n)}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$$

Then $$Z_n \to N(0,1) \quad \text{in distribution}$$

$$\left[\mathbb{P}[Z_n \geq t] \to \frac{1}{\sqrt{2\pi}} \int_t^{\infty} e^{-x^2/2}\, dx\right]$$

So letting $g \sim N(0,1)$ we know

$$P[Z_n > t] \le \underbrace{\left|P[Z_n > t] - P[Z > t]\right|}_{=:\,\text{Error}(t)} + \underbrace{P[g > t]}_{\le e^{-t^2/2}}$$

Bad news from CLT:

$\sup_t \{\text{Error}(t)\}$ is usually $\Omega\left(\frac{1}{\sqrt{n}}\right)$

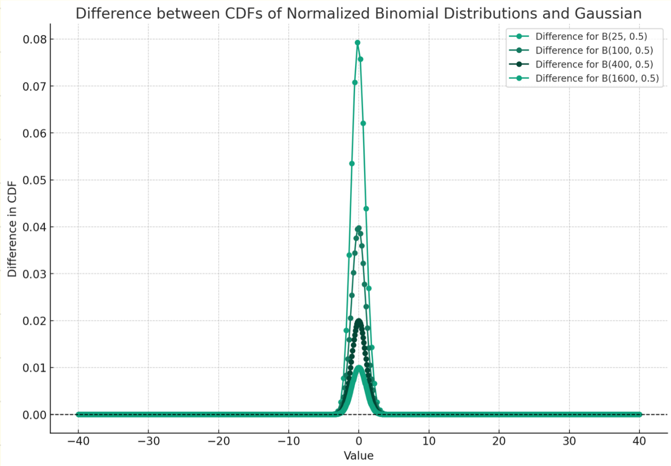
So from this argument

$$P[Z_n > t] \le \frac{c}{\sqrt{n}} + e^{-t^2/2}$$

<u>Ex:</u> Let $X_i \sim \text{Bernoulli}\left(\frac{1}{2}\right)$. What is $P\left(\frac{S_n}{n} > \frac{3}{4}\right)$? Let's compute

$$P\left(\frac{S_n}{n} \ge \frac{3}{4}\right) = P\left(\frac{S_n - \mathbb{E}S_n}{\sqrt{\text{Var}(S_n)}} \ge \frac{\sqrt{n}}{2}\right) \overset{\text{CLT}}{\le} \frac{c}{\sqrt{n}} + e^{-n/8}$$

Ex: Gaussian CDF − Binomial$(n, \frac{1}{2})$ CDF



The figures nonetheless suggest

$$P[Z_n > t] \leq \exp(-t^2/2) \quad !$$

We will prove such bounds for a large class of RVs directly <u>without</u> CLT!