Discussion 6

Zhichao Wang

8–10 AM Monday

HSS 5072

zhw036@ucsd.edu

$$0 \le \mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \ge 0$$

$$\mathbb{E}[X^2] \ge \mathbb{E}[X]^2$$

$$\mathbb{E}[f(X)] \ge f(\mathbb{E}[X])$$

convex function

Jensen's inequality

$$\mathbb{E}[aX + bY] = a\,\mathbb{E}[X] + b\,\mathbb{E}[Y]$$

$$\mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X)$$

constants

- $S_N = X_1 + X_2 + \cdots + X_N$

$$\mathbb{E}[S_N] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_N]$$

- If $X_1, X_2, \ldots, X_N$ $\boxed{\text{independent}}$

$$\mathrm{Var}(S_N) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_N)$$

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\,\mathbb{E}[X_2]$$

independence

X & Y independent

$$\begin{aligned}
&\mathrm{Var}(X+Y) \\
&= \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X+Y])^2 \\
&= \mathbb{E}[X^2 + Y^2 + 2XY] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \\
&= \mathbb{E}[X^2] + \mathbb{E}[Y^2] + \boxed{2\mathbb{E}[XY]} - (\mathbb{E}[X])^2 - (\mathbb{E}[Y])^2 - \boxed{2\,\mathbb{E}[X]\cdot\mathbb{E}[Y]} \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y)
\end{aligned}$$

$$\text{EX}: \quad X \sim N(3, 4)$$

$$E[X] = 3$$

$$Var(X) = 4 = E[X^2] - E[X]^2$$

$$E[X^2] = 9 + 4 = 13$$

find $P(2 < X < 6)$

$P(X > c) = 0.33$

$E[X^2]$ and $E[X^3]$

$$\Phi(x) = P(Z \le x)$$

$Z \sim N(0,1)$ $\boxed{x > 0}$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

$$\frac{X-3}{2} = \frac{X-3}{\sqrt{4}} \sim N(0,1)$$

$$P(2 < X < 6)$$

$$= P(-1 < X-3 < 3)$$

$$= P\left(-\frac{1}{2} < \frac{X-3}{2} < \frac{3}{2}\right)$$

$$= \Phi\left(\frac{3}{2}\right) - \Phi\left(-\frac{1}{2}\right)$$

$$= \Phi\left(\frac{3}{2}\right) - \left(1 - \Phi\left(\frac{1}{2}\right)\right)$$

$$= \Phi\left(\frac{3}{2}\right) - 1 + \Phi\left(\frac{1}{2}\right)$$

$$= 0.9332 + 0.6915 - 1$$

$$\Phi(-3) = 1 - \Phi(3)$$

$$P(X \le c) = 1 - 0.33 = 0.67$$

$$c = 0.44$$

$$Z = \frac{X-3}{2} \qquad Z \sim N(0,1) \qquad E[Z^3] = 0$$

$$E\left[\frac{(X-3)^3}{8}\right] = 0 \Rightarrow E[(X-3)^3] = 0$$

$$\| $$

$$E[X^3 + (-3)^3 + 3X(-3)^2 - 3X^2 \cdot 3] = 0$$

$$\mathbb{E}[X^3] - 27 + 27\underbrace{\mathbb{E}[X]}_{3} - 9\underbrace{\mathbb{E}[X^2]}_{13} = 0$$

Normal approximation of the binomial distribution. Suppose that $S_n \sim \text{Bin}(n, p)$ with n large and p not too close to 0 or 1. Then

$$P\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) \quad \text{is close to} \quad \Phi(b) - \Phi(a). \tag{4.3}$$

As a rule of thumb, the approximation is good if $\boxed{np(1-p) > 10.}$

Poisson approximation for counting rare events. Assume that the random variable X counts the occurrences of rare events that are not strongly dependent on each other. Then the distribution of X can be approximated with a Poisson(λ) distribution for $\lambda = E[X]$. That is,

$\boxed{np^2 \ll 1}$

$$P(X = k) \quad \text{is close to} \quad e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for} \quad k \in \{0, 1, 2, \dots\}. \tag{4.14}$$

$S_n \sim \text{Bin}(n, p) \longrightarrow \text{Poisson}(np)$

$X \sim \text{Poisson}(\lambda)$

$$P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad k=0,1,2\ldots$$

$$\mathbb{E}[X] = \text{Var}(X) = \lambda$$

Section 4.4

Exercise 4.9. Let $X \sim \text{Poisson}(10)$.

(a) Find $P(X \geq 7)$.
(b) Find $P(X \leq 13 \mid X \geq 7)$.

$X = \#$ goals for one game $\sim \text{Poisson}(\lambda)$

$P(X \geq 1) = 0.5 \Rightarrow$

$P(X=3) = ?$

Exercise 4.10. A hockey player scores at least one goal in roughly half of his games. How would you estimate the percentage of games where he scores a hat-trick (three goals)?

Exercise 4.11. On the first 300 pages of a book, you notice that there are, on average, 6 typos per page. What is the probability that there will be at least 4 typos on page 301? State clearly the assumptions you are making.

$X = \#$ typos in this page

$X = 0, 1, 2, \ldots$

$X \sim \text{Poisson}(\lambda) \qquad \mathbb{E}[X] = 6 = \lambda$

$$P(X \geq 4) = 1 - P(X < 4)$$
$$= 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3)$$

$$P(X=k) = e^{-6} \frac{6^k}{k!}$$

EX: flip a fair Coin 10000 times

$$P(\,|\#H - \#T| \le 100) \approx ?$$

$p = \frac{1}{2}$ $n = 10000$ $S_n = \#$ Heads

normal: $np(1-p) > 10$

$$\frac{10000}{2}\left(1-\frac{1}{2}\right) = 2500 > 10. \checkmark$$

poisson: $np^2 = 10000 \cdot \left(\frac{1}{2}\right)^2 = 2500$ ✗ not small

$\#H = S_n$, $\#T = n - S_n$

$$P(\,|S_n - (n - S_n)| \le 100)$$

$$= P(-100 \le 2S_n - n \le 100)$$

$$= P(n - 100 \le 2S_n \le \overset{10000}{\underbrace{n + 100}_{10100}})$$

$$4950 \le S_n \le 5050$$

$$= P\left(\frac{-50}{50} \le \boxed{\frac{S_n - 5000}{50}} \le \frac{50}{50}\right)$$

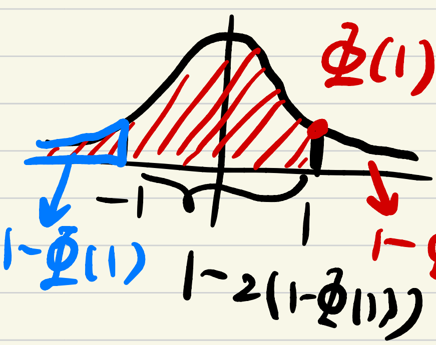
$$\approx 1 - 2(1 - \Phi(1))$$

normal $\approx \dfrac{S_n - np}{\sqrt{np(1-p)}}$

$np = 5000$

$np(1-p) = 2500$

$\sqrt{np(1-p)} = 50$



EX: How many randomly chosen guests should I invite to my party s.t. probability of having a guest with the same birthday as mine is at least 2/3.

$X \sim Bin(n, \frac{1}{365}) \Rightarrow P(X \geq 1) = 1 - P(X=0) = 1 - \left(\frac{364}{365}\right)^n \geq \frac{2}{3}$

X = # guests have the same birthday as mine.

$P(X \geq 1) \geq \frac{2}{3}$

Poisson Approx.

$\boxed{X \sim Poissn\left(\frac{n}{365}\right)} \Longrightarrow P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$

$\lambda = \frac{n}{365}$

$P(X \geq 1) = 1 - P(X=0)$

$= 1 - e^{-\lambda} \geq \frac{2}{3}$

$\Rightarrow \frac{1}{3} \geq e^{-\lambda} \Rightarrow \lambda \geq \ln 3$

$\frac{n}{365} \geq \ln 3$

$n \geq 365 \cdot \ln 3$

$n = \ln 3 \cdot 365 \approx 401$

$\boxed{\text{Check}}$ $n \cdot p^2 = 401 \cdot \left(\frac{1}{365}\right)^2 \ll 1$

Exact computation:

$$1-\left(\frac{364}{365}\right)^n \geq \frac{2}{3}$$

$$\Rightarrow \quad \frac{1}{3} \geq \left(\frac{364}{365}\right)^n$$

$$\ln\frac{1}{3} \geq n \cdot \ln\left(\frac{364}{365}\right)$$

$$n \geq \ln\left(\frac{1}{3}\right) / \ln\left(\frac{364}{365}\right) \approx 400$$

$$n \approx 400 \quad p = \frac{1}{365}.$$

① Poisson $\Rightarrow$ $n \cdot p^2 = \dfrac{400}{365^2} \ll 1$

small ✓

② normal $\Rightarrow$ $np(1-p) > 10$ ✗

$$= \frac{400}{365} \cdot \frac{364}{365} \approx 1$$