Computer Science Mentors (CSM)

Junior Mentor $\longrightarrow$ Senior Mentor

$\searrow$ TA $\swarrow$

AI $\longrightarrow$ Reader $\longrightarrow$ TA

Normal / Gaussian Dist

$X_1, X_2, \dots, X_n$ indep & identically dist

$X_1 \sim \text{Bin}(3, 1/4)$

$X_2 \sim \text{Bin}(3, 1/4)$

$\vdots$

$X_n \sim \text{Bin}(3, 1/4)$

$A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\lim_{n\to\infty} A_n \xrightarrow{d} \mathcal{N}(\mathbb{E}[A], \text{Var}(A))$

$n=1$

$A_1 = X_1 \sim \text{Bin}(3, 1/4)$

$A_2 = \frac{1}{2}(X_1 + X_2)$

$A_{10} = \frac{1}{10}(X_1 + X_2 + \dots + X_{10})$

$X \sim \mathcal{N}(\mu, \sigma^2)$ $\quad \mathbb{E}[X] = \mu$, $\text{Var}(X) = \sigma^2$

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2\right)}$$

$e^x = \exp(x)$

$\int_a^b f_X(x)\,dx = \Pr[X \in [a,b]]$

$= F_X(b) - F_X(a)$

$= \int_{-\infty}^{b} f_X(x)\,dx - \int_{-\infty}^{a} f_X(x)\,dx$

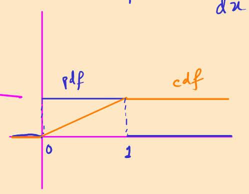
$F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx$

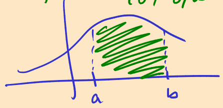
$f_X(x) = \frac{d}{dx} F_X(x)$

$U \sim \mathcal{U}[0,1]$

$F_U(x) = \begin{cases} 0 & x \le 0 \\ \frac{x}{1} & 0 \le x \le 1 \\ 1 & x \ge 1 \end{cases}$

$f_U(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & o/w \end{cases}$





$F_X(x) = \Pr[X \le x]$

$\Pr[X \in [a,b]] = F_X(b) - F_X(a)$

$\Pr[0.3 \le U \le 0.5] = 0.5 - 0.3 = 0.2$

$\int_{0.3}^{0.5} 1\,dx = 0.5 - 0.3 = 0.2$

$\Pr[V \le x] = \Pr[V \in (-\infty, x]]$

$= \int_{-\infty}^{x} f_V(x)\,dx$

$= F_V(x)$

$F_V(x) = \Pr[V \le x]$

$= \Pr[3U - 1 \le x]$

$= \Pr[U \le \frac{x+1}{3}]$

$= F_U\left(\frac{x+1}{3}\right)$

$f_V(x) = \frac{d}{dx} F_V(x)$

$= \frac{d}{dx} F_U\left(\frac{x+1}{3}\right)$

$= \frac{d}{dx}\left(\frac{x+1}{3}\right)$

$= \frac{1}{3}$

$V = 3U - 1$

$[-1, 2]$

$\frac{x+1}{3} \le 0 \Rightarrow x \le -1$

$f_V(x) = \begin{cases} 0 & x \le -1 \\ 1/3 & -1 \le x \le 2 \\ 0 & x \ge 2 \end{cases}$

$\frac{x+1}{3} \ge 1$

$\Rightarrow x \ge 2$

1 Interesting Gaussians

(a) If $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$ are independent, then what is $\mathbb{E}\left[(X+Y)^k\right]$ for any *odd* $k \in \mathbb{N}$?

$\mathbb{E}[(X+Y)^1] = 0 + 0 = 0$

$\int_{-\infty}^{\infty} \text{odd}(x)\, dx = 0$

$\mathbb{E}[(X+Y)^3] = 0$

$\mathrm{Var}(X^3) \neq \mathrm{Var}(X)$

$= \underbrace{\mathbb{E}[X^3]}_{0} + 3\,\underbrace{\mathbb{E}[X^2Y]}_{\mathbb{E}[X^2]\mathbb{E}[Y] = 0} + 3\,\underbrace{\mathbb{E}[XY^2]}_{\mathbb{E}[X]\mathbb{E}[Y^2] = 0} + \underbrace{\mathbb{E}[Y^3]}_{0} = 0$

$g(x) = x^3$

$\mathbb{E}[X^3] = \int_{-\infty}^{\infty} x^3 \cancel{\frac{1}{\sqrt{2\pi\sigma_X^2}}}\, \exp\left(\frac{-x^2}{\cancel{2\sigma_X^2}}\right)$ (C, γ)

$= \boxed{C \int_{-\infty}^{\infty} \underbrace{\underbrace{x^k}_{\text{odd}}\, \underbrace{\exp(-\gamma x^2)}_{\text{even}}}_{\text{odd}}\, dx} = 0$

$Z = X + Y \sim N(0, \sigma_X^2 + \sigma_Y^2)$

$X \sim N(\mu_1, \sigma_1^2)$

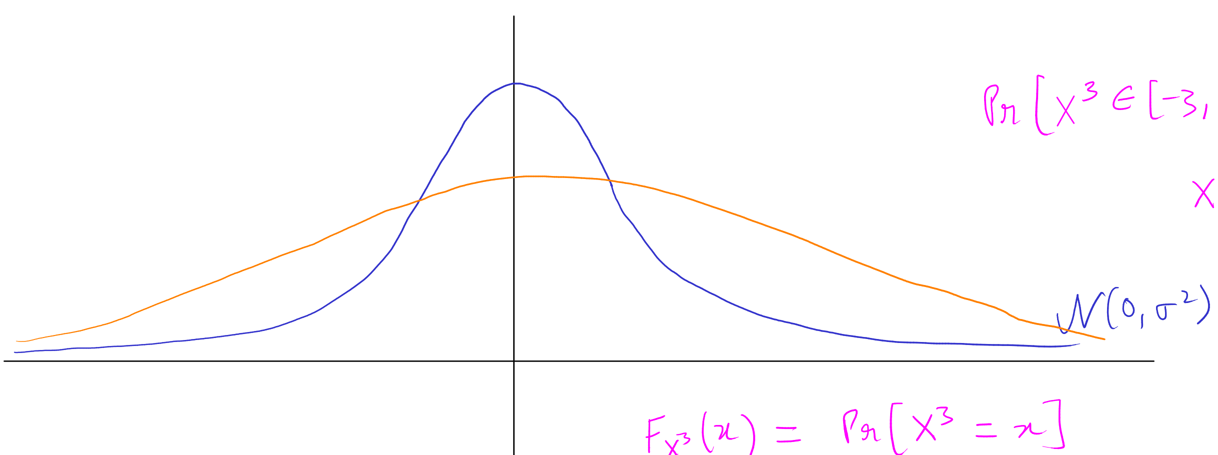
$Y \sim N(\mu_2, \sigma_2^2)$

$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$\mathbb{E}[(X+Y)^k]$
$= \mathbb{E}[Z^k]$
$= 0$

$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \underbrace{f_X(x)}_{\text{pdf}}\, dx$

$\sum_x g(x) \Pr[X = x]$



$\Pr[X^3 \in [-3, -1]] \geq 0$

X^3

$F_{X^3}(x) = \Pr[X^3 = x]$
$= \Pr[X = \sqrt[3]{x}] = F_X(\sqrt[3]{x})$

$f_{X^3} = \frac{d}{dx} F_{X^3}(x) = \frac{d}{dx} F_X(x^{1/3}) = \frac{1}{3} x^{-2/3} f_X(x^{1/3})$

$\neq$

$(f_X(x))^3$

$f_X(x)$

(b) Let $f_{\mu,\sigma}(x)$ be the density of a $N(\mu,\sigma^2)$ random variable, and let X be distributed according to $\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha) f_{\mu_2,\sigma_2}(x)$ for some $\alpha \in [0,1]$. Please compute $\mathbb{E}[X]$ and $\mathrm{Var}[X]$. Is X normally distributed?

$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$

$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

$\cancel{X = \alpha X_1 + (1-\alpha) X_2}$

$X = \alpha f_{\mu_1 \sigma_1}(x) + (1-\alpha) f_{\mu_2 \sigma_2}(x)$

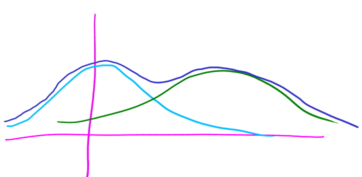
$\mathbb{E}[X] = \alpha \mu_1 + (1-\alpha)\mu_2$ plug into formula

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

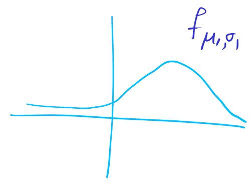
$\underbrace{\mathbb{E}[g(X)]} = \int_{-\infty}^{\infty} g(x) \underbrace{f_X(x)}_{\text{pdf}} dx$

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

$= \int_{-\infty}^{\infty} x^2 \left(\alpha f_{\mu_1 \sigma_1}(x) + (1-\alpha) f_{\mu_2 \sigma_2}(x)\right) dx$



$c \exp\left(c_2 (x - c_3)^2\right)$, $c > 0$, $c_2 < 0$



$f_{\mu_1,\sigma_1} + f_{\mu_2,\sigma_2} \not\Rightarrow f_{\mu_3,\sigma_3}$

$f_X(x) = \frac{1}{\sqrt{2\pi (3z)^2}} \exp\left(-\frac{1}{2(3z)^2}(x - (m^3 - nz))^2\right)$

$\mathcal{N}(m^3 - nz, (3z)^2)$

expo(λ)
$X \in \mathbb{R}_{\geq 0}$

pdf $\quad f_X(x) \quad \int_{-\infty}^{\infty} f_X(x)\,dx$

$\mathcal{N}(\mu, \sigma^2)$
$X \in \mathbb{R}$

cdf $\quad F_X(x) = \int_{-\infty}^{x} f_X(x)\,dx = \Pr[X \le x] = \Pr[X \in (-\infty, x)]$

$x \in [0, \infty) \quad \Pr[X = x] = e^{-\lambda} \frac{\lambda^x}{x!}$

$\Pr[X \in [a,b]]$
$= \int_a^b f_X(x)\,dx$ $= F_X(b) - F_X(a)$

$\Pr[Y = y] = e^{-(1-m^2)} \frac{(1-m^2)^x}{x!}$

$Y \sim \text{Pois}(1-m^2)$

2 Binomial Concentration

Here, we will prove that the binomial distribution is *concentrated* about its mean as the number of trials tends to ∞. Suppose we have i.i.d. trials, each with a probability of success $1/2$. Let S_n be the number of successes in the first n trials (n is a positive integer), and define

$$Z_n := \frac{S_n - n/2}{\sqrt{n}/2}.$$

(a) What are the mean and variance of Z_n?

(b) What is the distribution of Z_n as $n \to \infty$?

(c) Use the bound $\mathbb{P}[Z > z] \le (\sqrt{2\pi}z)^{-1}\,\mathrm{e}^{-z^2/2}$ when Z is a standard normal in order to approximately bound $\mathbb{P}[S_n/n > 1/2 + \delta]$, where $\delta > 0$.

3 Erasures, Bounds, and Probabilities

Alice is sending 1000 bits to Bob. The probability that a bit gets erased is p, and the erasure of each bit is independent of the others.

Alice is using a scheme that can tolerate up to one-fifth of the bits being erased. That is, as long as Bob receives at least 801 of the 1000 bits correctly, he can decode Alice's message.

In other words, Bob becomes unable to decode Alice's message only if 200 or more bits are erased. We call this a "communication breakdown", and we want the probability of a communication breakdown to be at most 10^{-6}.

(a) Use Chebyshev's inequality to upper bound p such that the probability of a communications breakdown is at most 10^{-6}.

$$\Pr[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2} \iff \Pr[|X-\mu| \geq \sigma k] \leq \frac{1}{k^2}$$

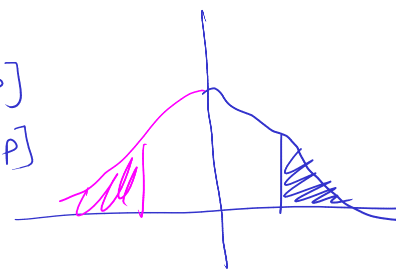
$X_i \quad 1 \leq i \leq 1000 \qquad \sum_{i=1}^{1000} X_i = X$

$\mathbb{E}[X] = 1000p$

$X \sim Bin(1000, p)$

$1000p(1-p) = Var(X)$

$$\Pr[X \geq 200] \leq 10^{-6}$$
$$= \Pr[X - 1000p \geq 200 - 1000p]$$
$$\leq \Pr[|X - 1000p| \geq 200 - 1000p]$$
$$\leq \frac{1000p(1-p)}{(200-1000p)^2} \leq 10^{-6}$$



(b) As the CLT would suggest, approximate the fraction of erasures by a Gaussian random variable (with suitable mean and variance). Use this to find an approximate bound for p such that the probability of a communications breakdown is at most 10^{-6}.

You may use that $\Phi^{-1}(1-10^{-6}) \approx 4.753$.