STAB52 Final Review

- $P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B)$
- $P^n_K = \frac{n!}{(n-k)!}$ $\quad C^n_K = \frac{n!}{k!(n-k)!}$
- n elements $\Rightarrow 2^n$ subset
- $(x+y)^n = (x+y)\cdots(x+y) = \sum_{i=0}^{n} \binom{n}{i} \cdot x^i \cdot y^{n-i}$
- $\binom{n}{k_1 k_2 \cdots k_n} = \frac{n!}{k_1! k_2! \cdots k_n!}$
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$ $\quad P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ $\quad P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$

$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$

- If independent

  $\Rightarrow P(A \cap B) = P(A) \cdot P(B) \quad P(A|B) = P(A) \quad P(B|A) = P(B)$

- If independent with $A \subseteq B$, then

  $\Rightarrow P(A \cap B) = P(A) \cdot P(B)$ but $A \subseteq B$ then $P(A \cap B) = P(A)$

  $\Rightarrow P(A) = P(A) \cdot P(B)$

  $\Rightarrow P(A) = 0$ or $P(B) = P(S) = 1$

Sequential probability problems

$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_2, A_1) \times \cdots$

Conditional independent if $P(A \cap B|C) = P(A|C)P(B|C)$ but mutually independent $\Rightarrow$ conditionally independent

$I_A \times I_B$: $A \cap B$ indicator CDF: $F_X(x) = P(X \leq x)$

Distribution $P(a < x \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$

1、discrete

① Bernoulli Distribution: 1/0 (building blocks) $X \sim Bernoulli(p)$

② Binomial Distribution: RV X be # of success in n trials $\binom{n}{x} p^x q^{n-x}$ $X \sim Binomial(n, p)$

③ Geometric Distribution: RV x be # of failures until 1st success pq^{x-1} $X \sim Geometric(p)$

⑤ Negative Binomial Distribution: RV x be # of trials until r^{th} success $\binom{x-1}{r-1} p^r q^{x-r}$ $X \sim NegBinom(r, p)$

⑥ Possion Distribution: RV x be # of success in cont interval (time etc) $X \sim Poisson(\lambda)$

2、Continuous (using CDF: $P(a < x \leq b) = F_X(b) - F_X(a)$ or PDF: $P(a < x \leq b)$ $\int_a^b f_X(x)dx$)

(i) uniform

① uniform Distribution: $\frac{b-a}{u-l}$ $\forall l \leq a < b \leq u$ $X \sim Uniform(l, u)$

② Exponential Distribution: $X \sim Exponential(\lambda)$ ...

③ Gamma Distribution (sum of Exponentials): $X \sim Gamma(\alpha, \lambda)$

④ Normal Distribution (RV $X \in \mathbb{R}$ distribution around its center with some spread): $X \sim Normal(\underbrace{\mu}_{center}, \underbrace{\sigma}_{width})$

if we have CDF: $F_X(x)$
then PDF $= f_X(x) = \frac{d}{dx} F_X(x)$

if we have PDF: $f_X(x)$
then CDF $= F_X(x) = \int_{-\infty}^{x} f_X(x)dx$

- Gamma Distribution

For $\alpha > 1$: $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

α is int: $\Gamma(\alpha) = (\alpha-1)!$

- Normal Distribution

μ: center of distribution

σ: width of distribution

Expectation

- $E(X)$ doesn't have to be one of possible X value
- $E(I_A)=P(A)$

$$\begin{cases} E(X)=\sum_{\text{all } x} x\,p_X(x) & \text{(dis)} \\ E(X)=\int_{-\infty}^{\infty} x f_X(x)\,dx & \text{(cont)} \end{cases}$$

- $h(X)=a+bX$
  then $E(a+bX)=a+bE(X)$

- $V(X)=E((X-E(X))^2)=E(X^2)-\mu^2=E(X^2)-E(X)^2$
- Variance typically denoted by σ^2

Change of variable

3 methods $\begin{cases} \text{① General method} \\ \text{② CDF method} \\ \text{③ PDF method (cont RVs)} \end{cases}$

① if $P(X\in B)$ and $Y=h(X)$, then $P(Y\in A)=P(X\in \overset{\text{inverse image}}{\overbrace{h^{-1}}}(A))$

② if h is $\nearrow$ $F_Y(y)=P(Y\le y)=P(Y\in[-\infty,y])=P(X\in[-\infty,h^{-1}(y)])=P(X\le h^{-1}(y))=F_X(h^{-1}(y))$

if h is $\searrow$ $F_Y(y)=P(Y\le y)=P(Y\in[-\infty,y])=P(X\in[h^{-1}(y),\infty])=P(X>h^{-1}(y))=1-P(X\le h^{-1}(y))$
$=1-F_X(h^{-1}(y))$

③ $f_Y(y)=\dfrac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|}$

2019 Midterm

$E(I) = P(I=1)$

$$F_T(t) = \begin{cases} 0 & , t \le 2 \\ 1 - e^{-(\frac{t}{2}-1)^3} & , t > 2 \end{cases}, \quad S = \frac{42.2}{T}$$

since $t \in (2, \infty)$, then $S \in (0, 21.2)$

$$F_S(s) = P(S \le s) = P\left(\frac{42.2}{T} \le s\right) = P\left(T \ge \frac{42.2}{s}\right) = 1 - P\left(T < \frac{42.2}{s}\right) = 1 - F_T\left(\frac{42.2}{s}\right) = 1 - \left(1 - e^{-\left(\frac{\frac{42.2}{s}}{2} - 1\right)^3}\right)$$

$$= e^{-\left(\frac{21.1}{s} - 1\right)^3}$$

6.

a) WTS: $F(b) = \dfrac{F(a+b) - F(a)}{1 - F(a)}$

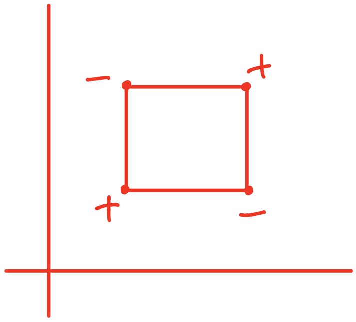
$$F_X(b) = P(X \le b) = P(X \le a+b \mid X > a) = \frac{P(X \le a+b \cap X > a)}{P(X > a)} = \frac{P(a < X \le a+b)}{1 - F(a)} = \frac{F(a+b) - F(a)}{1 - F(a)}$$

b) WTS: $\dfrac{P_X(k+1)}{P_X(k)} = \dfrac{1 - F(k)}{1 - F(k-1)}$

from part a) we have $F(b) = \dfrac{F(a+b) - F(a)}{1 - F(a)} \Rightarrow F(b)(1 - F(a)) = F(a+b) - F(a)$

where $\underbrace{F(a+b) - F(a)}_{P_X(b)} = F(b)(1 - F(a))$

then $$\frac{P_X(k+1) = F(k+1) - F(k) = F(1)(1 - F(k))}{P_X(k) = F(k) - F(k-1) = F(1)(1 - F(k-1))} = \frac{1 - F(k)}{1 - F(k-1)}$$



Joint PMF: $P_{X,Y}(x,y) = P(X=x, Y=y) = P(\{X=x\}\cap\{Y=y\})$

Marginal PMF: $\sum_y P(X=x, Y=y)$ (fix one, sum up others)

Joint CDF: $F_{X,Y}(x,y) = P(X\le x, Y\le y)$

Marginal CDF: $F_X(x)=F_{X,Y}(x,\infty)$ & $F_Y(y)=F_{X,Y}(\infty,y)$ $\lim_{x/y\to\infty}$ $\quad \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)\,ds\,dt \quad \frac{\partial^2}{\partial x\,\partial y}F_{X,Y}(x,y)$

Joint PDF: $P((X,Y)\in R)=\iint_R f_{X,Y}(x,y)\,dx\,dy$

Marginal PDF: $f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$ $\quad f_Y(y)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$

* independent RVs $\Rightarrow P(X\in A, Y\in B)=P(X\in A)P(Y\in B)$ CDF: $F_{X,Y}(x,y)=F_X(x)F_Y(y)$

iid: independent & identically distributed

RVs cannot be independent if their joint one is not rectangular □ ✓ △ ✗

determine if $X_1 \perp X_2$ we can check if $f_{X_1,X_2}(x_1,x_2) \overset{?}{=} \underbrace{f_{X_1}(x_1)f_{X_2}(x_2)}_{\text{marginal}}$

Conditional distribution

$$P((X,Y)\in A \mid (X,Y)\in B) = \frac{P((X,Y)\in A \cap (X,Y)\in B)}{P((X,Y)\in B)}$$

conditional PMF: $P_{X|Y}(x|y) = P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{P_{X,Y}(x,y)}{P_Y(y)}$

conditional CDF: $F_{X|Y}(x|y) = P(X\le x \mid Y=y)$

conditional PDF: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ $\nearrow \int_{-\infty}^{x}$

2D change of RVs is similar to 1D $\Longrightarrow$ Ex. $f_{X,Y}(x,y)=\begin{cases}4x^3y+2y^5, & 0\le x\le 1,\ 0\le y\le 1\\ 0, & \text{o/w}\end{cases}$

$Z=X+Y^2$ $W=X-Y^2$ what is $f_{Z,W}(z,w)$?

PDF method

$$f_{Z,W}(z,w)=\frac{f_{X,Y}(h^{-1}(z,w))}{|J(h^{-1}(z,w))|}$$

Sum of RVs: $X_1+X_2+\dots$

order statistics of RVs: largest value of $X_1, X_2, \dots$

Convolution Method: $Z=X+Y$

$$f_Z(z)=\int_{-\infty}^{\infty} f_{X,Y}(x,\underbrace{z-x}_{y})\,dx \text{ or } \int_{-\infty}^{\infty} f_{X,Y}(\underbrace{z-y}_{x},y)\,dy$$

same for discrete

For RVs $X_1,\dots,X_n$ the k^{th} order statistic ($X_{(k)}$) is the k^{th} smallest variable

$X_{(1)}=\min(X_1,\dots,X_n)$

$X_{(n)}=\max(X_1,\dots,X_n)$

we know that $f_{Z,W}(z,w)=\dfrac{f_{X,Y}(h^{-1}(z,w))}{|J(h^{-1}(z,w))|}$ and

$h(z,w)=(h_1(x,y),h_2(x,y))=(x+y^2,x-y^2)$

so $J(x,y)=\begin{vmatrix}\frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y}\\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y}\end{vmatrix}=\begin{vmatrix}1 & 2y\\ 1 & -2y\end{vmatrix}=-2y-2y=-4y$

and $\begin{cases}Z=X+Y^2\\ W=X-Y^2\end{cases}\Rightarrow \begin{cases}Z+W=2X\\ \Downarrow\\ X=\frac{Z+W}{2}\end{cases}$ and $\begin{matrix}Z-W=2Y^2\\ \Downarrow\\ Y=\sqrt{\frac{Z-W}{2}}\end{matrix}$ $\therefore h^{-1}(z,w)=\left(\frac{z+w}{2},\sqrt{\frac{z-w}{2}}\right)$

$$\therefore f_{Z,W}(z,w)=\frac{f_{X,Y}(h^{-1}(z,w))}{|J(h^{-1}(z,w))|}=\frac{4\left(\frac{z+w}{2}\right)^3\sqrt{\frac{z-w}{2}}+2\left(\sqrt{\frac{z-w}{2}}\right)^5}{4\sqrt{\frac{z-w}{2}}}$$

and $0\le x\le 1,\ 0\le y\le 1 \Rightarrow 0\le \frac{z+w}{2}\le 1,\ 0\le \sqrt{\frac{z-w}{2}}\le 1$

For i.i.d. RVs $X_1,\dots,X_n$, the distribution of max $X_{(n)}$ is

$$F_{(n)}(x)=[F(x)]^n \ \&\ f_{(n)}(x)=\frac{d}{dx}[F(x)]^n=n[F(x)]^{n-1}f(x)$$

the distribution of min $X_{(1)}$ is

$$F_{(1)}(x)=1-[1-F(x)]^n \ \&\ f_{(1)}(x)=n[1-F(x)]^{n-1}f(x)$$

if $X\perp Y$, then $E[g(X)h(Y)]=E(g(X))\times E(h(Y))$

rth moment of X is $E(X^r)$

rth central moment is $E((X-\mu)^r)$

covariance of X and Y is $\text{cov}(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]$

Correlation $=\rho_{X,Y}=\dfrac{\text{cov}(X,Y)}{\sigma_X\sigma_Y}$

if $\text{cov}(X,Y)>0\Rightarrow X\uparrow Y\uparrow$ /

$\text{cov}(X,Y)<0\Rightarrow X\uparrow Y\downarrow$ \

if $=0$ ○

Cov(X,Y) properties

① $Cov(X,X)=V(X)$

② $Cov(X,Y)=E(XY)-\mu_x\mu_y$

③ $X\perp Y \Rightarrow Cov(X,Y)=0$ but $Cov(X,Y)=0 \not\Rightarrow X\perp Y$

$V(X_1+X_2)=V(X_1)+V(X_2)+2Cov(X_1,X_2)$

eg.

- Two processes run *serially*, with i.i.d. Exponential(1) completion times. Find the variance of the time until both processes complete.

$Z=X_1+X_2$

$V(Z)\overset{\text{i.i.d}}{=}V(X_1)+V(X_2)+2\underbrace{Cov(X_1,X_2)}_{\text{i.i.d}\ 0}=1+1=2$

- Repeat, assuming completion times are correlated with ρ=.5

$V(Z)=V(X_1)+V(X_2)+2Cov(X_1,X_2)$

$=1+1+2\cdot\frac{1}{2}=3$

$\Rightarrow Cov(X_1,X_2)=\frac{1}{2}$

$=Corr(X_1,X_2)\cdot\underbrace{\sqrt{V(X_1)}}_{=1}\cdot\underbrace{\sqrt{V(X_2)}}_{=1}$

$m_X(t)=E(e^{tX})$

$E(X^k)=m_X^{(k)}(0)$ i.e want to find $E(X^k)$, we have $m_X(t)$, then diff k times of $m_X(t)$

$\Rightarrow m_X^{(k)}(t)$ and let $t=0$

and $E(e^{tX})=\int_{-\infty}^{\infty}e^{tx}f_X(x)dx$ or $\sum_{\text{all }x}e^{tx}P_X(x)$

$m_X(t)=m_Y(t)\Leftrightarrow X\sim Y$

Let $Y=g(X_1,\ldots,X_n)$ with $m_Y(t)=E(e^{tY})=E(e^{t\cdot g(X_1,\ldots,X_n)})$

If $m_Y(t)$ is MGF of some known distribution $\rightarrow$ Y follows that distribution

Let $Y=a_1X_1+\ldots+a_nX_n$, where $X_1,\ldots,X_n$ are independent with MGFs

Then $m_Y(t)=m_{X_1}(a_1t)\times\ldots\times m_{X_n}(a_nt)=\prod_{i=1}^{n}m_{X_i}(a_it)$

conditional Expectation

Given RV Y and event A, g(Y) is a function

Discrete: $E[g(Y)|A]=\sum_y g(y)P(y|A)$

Continuous: $E[g(Y)|A]=\int_{-\infty}^{\infty}g(y)f(y|A)dy$

$\Rightarrow$ ① find $f(y|A)$ or $P(y|A)$ where $f(y|A)=\frac{d}{dx}F(Y|A)=\frac{d}{dx}P(Y\le y|Y>1)$

$=\frac{d}{dx}\frac{P(1\le Y\le y)}{P(Y>1)}$

② $E(Y|A)=\int_{1}^{\infty}y\cdot f(y|A)dx$

A: {Y>1}

Given RV Y and X , g(Y) is a function

Discrete: $E[g(Y)|X=x] = \sum_x g(Y) p(y|x)$

Continuous: $E[g(Y)|X=x] = \int_{-\infty}^{\infty} g(Y) f(y|x) dy$ and $f_{X|Y}(x|y) = \dfrac{f_{XY}(x,y)}{f_Y(y)}$

$$\boxed{V(X+Y) = V(X) + V(Y) + \overbrace{2\,Cov(X,Y)}^{\text{if indep} \Rightarrow 0} \qquad \Rightarrow \text{indep } V(X+Y) = V(X) + V(Y)}$$

$$\boxed{V(aX) = a^2 V(X)}$$

$$\bar{X}_n \sim^{appr} N(\overset{E(\bar{X}_n)}{\mu}, \overset{V(\bar{X}_n)}{\sigma^2 / n})$$

$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$ where $E(\bar{X}_n) = \mu$

but $V(\bar{X}_n) = \frac{\sigma^2}{n}$

$\Phi(z)$ be standard normal distribution

Example

- Express the following probabilities in terms of the standard Normal CDF $\Phi(\cdot)$ $= F_Z(z)$ (i.e CDF of standard normal $Z \sim N(0,1)$)
  - $P(X>1)$ where $X \sim N(\mu=5, \sigma^2=4)$

    $P(X>1) = P\left(\frac{X-\mu}{\sigma} > \frac{1-\mu}{\sigma}\right) = P\left(Z > \frac{1-5}{2}\right) = P(Z > -2) = 1 - P(Z \le -2) = 1 - \Phi(-2)$

  - $P(|X|>1)$ where $X \sim N(\mu=5, \sigma^2=4)$

    $P(|X|>1) = P(\{X<-1 \cup X>1\}) = P(X<-1) + \underbrace{P(X>1)}_{1-\Phi(-2)}$

    $P(X<-1) = P\left(\frac{X-\mu}{\sigma} < \frac{-1-\mu}{\sigma}\right) = P\left(Z < \frac{-1-5}{2}\right)$
    $= P(Z<-3) = \Phi(-3)$

    $\therefore$ $= \Phi(-3) + 1 - \Phi(-2)$

- We can *estimate* σ^2 by *sample variance*

$$\boxed{S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

replace μ with $\bar{X}_n$

  - Note: if we knew mean μ we could use $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$, but since μ is unknown we must rely only on sample $X_1,\ldots,X_n$

- Gamma($n/2, 1/2$) called *chi-square* distribution with parameter n
  - Parameter n a.k.a. *degrees of freedom*
- Sampling distribution of sample variance given by

$$\frac{(n-1)S_n^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{\sigma^2} \sim \chi^2(n-1)$$

  - Result derived from $\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n)$

- If $Z \sim N(0,1)$ & $V \sim \chi^2(n)$ with $Z \perp V$, then $T = \frac{Z}{\sqrt{V/n}}$ follows a

t distribution, with parameter (*degrees of freedom*) n

- For sample mean & variance, we have

$$\frac{(\bar{X}_n - \mu)\big/\sqrt{\sigma^2/n}}{\sqrt{S_n^2/\sigma^2}} = \boxed{\frac{\bar{X}_n - \mu}{\sqrt{S_n^2/n}} \sim t(n-1)}$$

  - *t* distribution can be used to find accuracy of $\bar{X}_n$ estimation, even if variance is unknown (i.e. using sample estimate of σ^2)