STAB52 Notes

Probability Models

Measure of how likely is an event to happen values range: $[0,1]$

S - sample space $\Leftrightarrow$ universal set

$s_1, s_2, \cdots \in S$ - outcomes $\Leftrightarrow$ elements

$A, B, C, \cdots = \{s_1, s_2 \cdots\}$ - events $\Leftrightarrow$ set

Set ~~Prober~~ Properties

Commutative $A \cap B \Leftrightarrow B \cap A$ $\quad A \cup B \Leftrightarrow B \cup A$

Associative $A \cup (B \cup C) \Leftrightarrow (A \cup B) \cup C$ $\quad A \cap (B \cap C) \Leftrightarrow (A \cap B) \cap C$

Distributive $A \cup (B \cap C) \Leftrightarrow (A \cup B) \cap (B \cup C)$ $\quad A \cap (B \cup C) \Leftrightarrow \cancel{A \cup} (A \cap B) \cup (A \cap C)$

$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$ $\quad \bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n$

De Morgan $(A \cap B)^c \Leftrightarrow A^c \cup B^c$ $\quad (A \cup B)^c \Leftrightarrow A^c \cap B^c$

$A \cap A^c \Leftrightarrow \emptyset$ $\quad A \cup A^c \Leftrightarrow S$ $\quad S^c \Leftrightarrow \emptyset$

Event Relations

$A \subseteq B$ A is a subset of B (B has all elements of A)

$\emptyset \subseteq$ any set $\quad A \subseteq B \cap B \subseteq A \Rightarrow A = B$

Disjoint / Mutually Exclusive Two or more events have no elements in common

Partition Two or more events are disjoint and $\bigcup_{i=1}^{n} A_i = S$

Probability Axioms

P( ) Probability Function / Measures a map from set/events to ~~$\mathbb{R}$~~ $[0,1] \in \mathbb{R}$

- $0 \le P(A) \le 1 \quad \forall A \subseteq S$
- $P(S) = 1$
- $A_1, A_2, A_3, \cdots$ are mutually exclusive $\Leftrightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ $\quad (P(A_1 \cup A_2 \cup \cdots A_n \cup \cdots)) =$
  $= P(A_1) + P(A_2) + \cdots$

Probability Model consists of $(S, \{A, B, \cdots\}, P)$

Law of Total Probability

If $A_1, A_2, \cdots$ is a partition and $B \in S$ then $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots$

* Useful for $P(B) = P(A \cap B) + P(A^c \cap B)$ $\quad A \setminus B = A \cap B^c$

Inclusion Exclusion Principle

$P(A \setminus B) = P(A) - P(A \cap B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

For union of n events: + events of ~~even~~ odd number of elements

− events of even number of elements

e.g. $P(A \cup B \cup C \cup D) = P(A) + P(B) + P(C) + P(D)$

$- P(AB) - P(AC) - P(AD) - P(BC) - P(BD) - P(CD)$

$+ P(ABC) + P(ABD) + P(ACD) + P(BCD) - P(ABCD)$

Discrete Sample Space - sample spaces with distinct outcomes

can be finite : finite number of elements, ~~can~~ countably infinite : elements ~~can~~ has corresponding $\mathbb{N}$

if $A = \{s_1, s_2, \cdots\}$ then $P(A) = P(s_1) + P(s_2) + \cdots$

Uniform Probability Space - finite space whose outcomes have equal probability

if $S = \{s_1, \cdots, s_n\}$ where $n = |S|$, then $P(s_i) = \frac{1}{n} = \frac{1}{|S|} > 0$, $\forall i = 1, \cdots n$

if $A \subseteq S$ then $P(A) = \frac{|A|}{|S|}$

Counting

Multiplication m elements and n elements - $m \times n$ possible ordered pairs

Permutation Ordered arrangement of k elements, chosen WITHOUT replacement from n possible elements $= P_k^n = \frac{n!}{k!}$

Combination Unordered collection of k elements, chosen WITHOUT repetition from n possible elements $= C_k^n = \binom{n}{k} = \frac{n!}{(n-k)!\,k!}$

Binomial Theorem

$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i} \quad \Rightarrow \quad 2^n = (1+1)^n = \sum_{i=0}^{n} \binom{n}{i}$

Properties of Binomial Coefficients

$\binom{n}{0} = \binom{n}{n} = 1 \qquad \binom{n}{1} = \binom{n}{n-1} = n \qquad \binom{n}{k} = \binom{n}{n-k}$

Multinomial Rule

\# ways to partition n elements into ~~k~~ m sets each with $k_1, \cdots k_m$ elements

$$= C^n_{k_1, k_2, \cdots k_m} = \binom{n}{k_1, k_2 \cdots k_m} = \frac{n!}{k_1!\,k_2!\cdots k_m!} \qquad \text{where } \sum_{i=1}^{m} k_i = n$$

Multinomial Theorem

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{\substack{k_1, k_2 \cdots k_m \\ \sum k_i = n}} \binom{n}{k_1, k_2 \cdots k_m} x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m}$$

any combination of $k_1, k_2 \ldots k_m$ that adds up to n

Conditional Probability - the probability of A given B has occured

$P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(B) > 0$ and follows probability axioms

Baye's Rule

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

law of total probability

$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots$ if $A_1, A_2, \cdots$ forms a partition

* use: $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$

$$P(A_j|B) = \frac{P(B|A_j)\,P(A_j)}{\underbrace{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}_{P(B)}}$$

Independence

$P(A \cap B) = P(A)P(B) \iff P(A|B) = P(A)$ and $P(B|A) = P(B)$ if $P(A) > 0$, $P(B) > 0$

this is a PROPERTY of events

Mutual Independence

finite collection $A_1, A_2, \cdots A_n$ is mutually independent if for any sub-collection $A_{k_1}, A_{k_2}, \cdots A_{k_m}$

$$P(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_m}) = P(A_{k_1})P(A_{k_2}) \cdots P(A_{k_m}) \iff P\left(\bigcap_{i=1}^{m} A_{k_i}\right) = \prod_{i=1}^{m} P(A_{k_i})$$

Conditioning ~~can~~ on Multiple Events

$$P(A|B,C) = P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} \quad P(B \cap C) > 0$$

if A, B, C are mutually independent, then

$$P(A|B,C) = P(A), \quad P(B|A,C) = P(B), \quad P(A \cap C|B) = P(A \cap C)$$

pairwise independence : $P(A_i \cap A_j) = P(A_i)P(A_j) \quad \forall i<j;\ i,j = 1, 2, \cdots, n$

$\not\Rightarrow A_1, A_2, \ldots, A_n$ are mutually independent

General Multiplication Rule of Probability

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2,A_1) \cdots P(A_n|A_{n-1}, A_{n-2}, \cdots A_1)$$

Conditional Independence

$P(A \cap B|C) = P(A|C)P(B|C)$ — A, B conditionally independent given C

$\Rightarrow P(A|B \cap C) = P(A|C)$ and $P(B|A \cap C) = P(B|C)$

conditional independence $\not\Leftrightarrow$ independence

$P(A \cap B|C) = P(A|C)P(B|C) \not\Leftrightarrow P(A \cap B) = P(A)P(B)$

BUT mutually independent $\Rightarrow$ conditionally independent

$P(A \cap B|C) = P(A \cap B)$ ~~P(C)~~ $= P(A)P(B) = P(A|C)P(B|C)$

RANDOM VARIABLES & DISTRIBUTIONS

A function from the sample space ~~G~~ (S) to the real line ($\mathbb{R}$)

$$\underset{\text{event}}{A} = \{\underset{\text{set of events } s \text{ in } S}{s \in S} : \underset{\text{map from } S \text{ to } \mathbb{R}}{X(s) = b}\} = \{\underset{RV}{X} = \underset{\mathbb{R}}{b}\}$$

Discrete - if the RV X can assume a finite $\{x_1, x_2, \cdots, x_n\}$ or countably infinite $\{x_1, x_2, \cdots\}$ number of values

Indicator RVs

$$I_A(s) = \begin{cases} 1, & s \in A \\ 0, & s \notin A \end{cases} \qquad I_A \times I_B = I_{A \cap B}$$

Distribution The collection of probabilities $P(X \in B)$ for all subsets B of real line

Cumulative Distribution Function

$F_X(x) = P(X \leq x) = P(\{s \in S : X(s) \leq x\}) \quad \forall x \in \mathbb{R}$

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

- $F_X(-\infty) \equiv \lim_{x \to -\infty} F_X(x) = 0$
- $F_X(\infty) \equiv \lim_{x \to \infty} F_X(x) = 1$
- $\forall x_1 < x_2 \in \mathbb{R} \Rightarrow F_X(x_1) \leq F_X(x_2)$

Discrete Distributions - ~~collection of all probabilities of the form~~

$P(X = x_i) = P(\{s \in S : X(s) = x_i\}) = p_X(x_i) \quad , \forall i = 1, 2, \cdots$

- $\sum_{\forall i} p_X(x_i) = 1$
- $p_X(x_i) \geq 0 \quad \forall i$

~~CPF~~ CDF of discrete distribution : $F_X(x) = P(X \leq x) = \sum_{x_i \leq x} p_X(x_i) \quad \forall x \in \mathbb{R}$

Bernoulli Distribution # successes

$X \sim \text{Bernoulli}(p)$ probability of success

$\equiv X \sim \text{Binomial}(1, p)$

$$p_X(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

Binomial Distribution # successes

$X \sim \text{Binomial}(n, p)$ # of trials, probability of success

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$

Geometric Distribution # failure before success

$X \sim \text{Geometric}(p)$

$\equiv X \sim \text{Negative Binomial}(1, p)$

$$p_X(x) = \begin{cases} (1-p)^x p & x = 0, 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}$$

Negative Binomial Distribution # failure before r^{th} success

$X \sim \text{Negative Binomial}(r, p)$ # success, probability of success

$$p_X(x) = \begin{cases} \binom{x+r-1}{x} \cancel{q^x p} \; (1-p)^x p^r & x = 0, 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}$$

Poisson Distribution # success

$X \sim \text{Poisson}(\lambda)$ average # of successes over an interval

$$p_X(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & x = 0, 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}$$

Hypergeometric Distribution

$X \sim \text{Hypergeometric}(N, M, n)$

$$p_X(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \quad \max(0, n+M-N) \leq x \leq \min(n, M)$$

Continuous Distributions

$P(X = x) = 0 \;, \forall x \in \mathbb{R}$, but $P(S) = 1$ as X takes on uncountably many values

$\hookrightarrow P(X \in [a, b]) > 0$ for $[a, b] \in \mathbb{R}$

Density Function $f: \mathbb{R} \to \mathbb{R}$ $f(x) \geq 0$ $\forall x \in \mathbb{R}$ and
$\int_{-\infty}^{\infty} f(x)dx = 1$

Absolutely Continuous RV

A RV is absolutely continuous if there is a density function f, such that

$$P(a \leq x \leq b) = \int_a^b f(x)dx \quad , a \leq b$$

CDF for Continuous Distribution

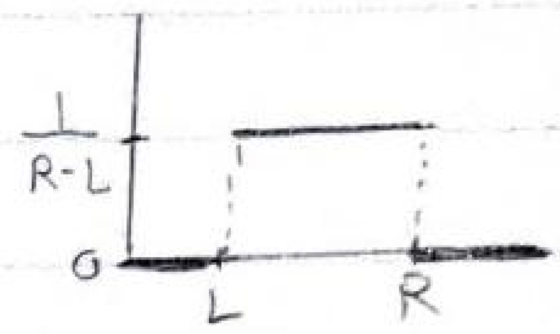
$F_X: \mathbb{R} \to [0,1]$ $F_X(x) = P(X \leq x)$ $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$

$P(a < x \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(t)dt$

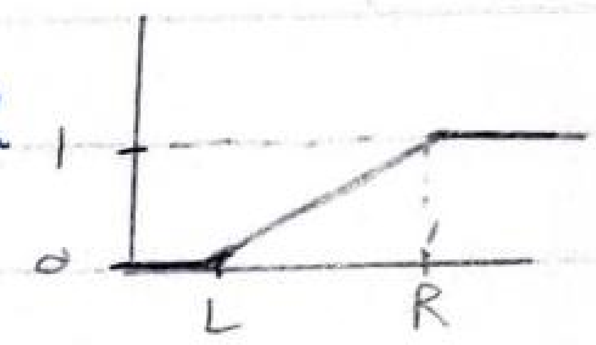
Uniform Distribution

$X \sim \text{Uniform}[L,R]$ interval $[L,R]$

$$f(x) = \begin{cases} \frac{1}{R-L} & , L \leq x < R \\ 0 & \text{otherwise} \end{cases}$$



$$F(x) = \begin{cases} 0 & , x < L \\ \frac{x-L}{R-L} & , L \leq x \leq R \\ 1 & , x \geq R \end{cases}$$



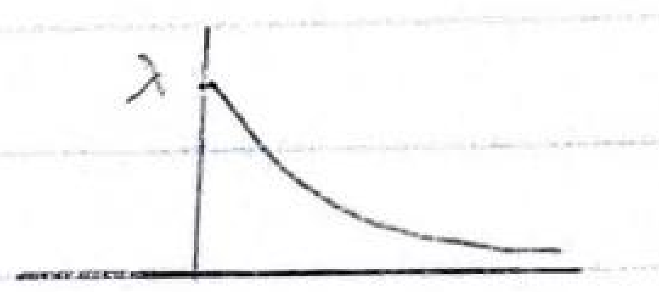
Exponential Distribution

$X \sim \text{Exponential}(\lambda)$ $\lambda > 0$

- exponential distribution is memoryless

$\equiv X \sim \text{Gamma}(1, \lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Gamma Distribution

$X \sim \text{Gamma}(\alpha, \lambda)$ $\lambda, \alpha > 0$

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & , x > 0 \\ 0 & x \leq 0 \end{cases} \qquad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$

for $\alpha > 1 \Rightarrow \Gamma(\alpha) = \alpha \Gamma(\alpha - 1)$

for $\alpha \in \mathbb{N} \Rightarrow \Gamma(\alpha) = (\alpha-1)!$

Normal Distribution

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

$\Phi(z) = 1 - \Phi(-z) = \cancel{P(X \leq x)}\ P(Z \leq z)$

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} = \phi(x)$ $x \in \mathbb{R}$

$P(a < x < b) = \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$ $z = \frac{x-\mu}{\sigma}$

Beta Distribution

$X \sim \text{Beta}(a, b)$ $a > 0$, $b > 0$

$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

$$f_X(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} & , 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

EXPECTATION

Discrete $E(X) = \sum_{x_i} x P(X=x) = \sum_{x_i} x p_X(x) = \mu_X$

Continuous $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$

if $Y = h(X)$ $E(Y) = E(h(X)) = \begin{cases} \sum_x h(x) p_X(x) & \text{discrete} \\ \int_{-\infty}^{\infty} h(x) f_X(x) & \text{continuous} \end{cases}$

Linearity of Expectation

$E(a+bX) = a + bE(X)$

$E(g(X) + h(X)) = E(g(X)) + E(h(X))$

Independence

If X and Y are independent,

$E(XY) = E(X)E(Y)$

Variance

$V(X) = Var(X) = \sigma^2 = E((X-E(X))^2) = E(X^2) - (E(X))^2$

Standard Deviation / SD/σ

positive $\sqrt{\sigma^2}$

$E(X^2) = Var(X) + (E(X))^2$

1D CHANGE OF VARIABLE

X is a RV and $Y = h(X)$ $h: \mathbb{R} \to \mathbb{R}$, ~~If~~ X is discrete $\Rightarrow$ Y is discrete

$p_Y(y) = \sum_{x \in h^{-1}\{y\}} p_X(x)$

set of X for which $h(X) = y$

X is continuous => Y is continuous

Finding $f_Y(y)$ given $Y = h(X)$ and $f_X(x)$

METHOD 1:

1. $F_Y(y) = P(Y \leq y)$ --- definition
$= P(h(X) \leq y)$
$= P(X \leq h^{-1}(y))$
$= F_X(h^{-1}(y))$
usually depends on how second line works out

2. $f_Y(y) = \frac{d}{dy} F_Y(y)$
$= \frac{d}{dy} F_X(h^{-1}(y)) = f_X(\cdots)\cdots$ (chain rule)

answer you got in [1]

METHOD 2: $f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|}$

1. Check $h(x)$ strictly increasing or decreasing with ~~and~~ support
2. Find $h^{-1}(y)$
3. Find $h'(x)$
4. $f_Y(y) = \frac{f_X(h^{-1}(y))}{|h'(h^{-1}(y))|}$

Simulation

Let F be a CDF, let $U \sim \text{Uniform}[0,1]$, let RV $Y = F^{-1}(U)$ They $P(Y \leq y) = F(y)$

e.g. Generate Exponential(λ) based on Uniform [0,1]

1. Let $U \sim$ Uniform $[0,1]$
2. $f_X(x) = \lambda e^{-\lambda x} \Rightarrow F_X(x) = \int_0^\infty \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}$ } given
3. $y = 1 - e^{-\lambda x}$

$1 - y = e^{-\lambda x}$ if you have $F_X(x) = y$

$\ln(1-y) = -\lambda x$ find $F^{-1}(y) = x$

$\frac{-\ln(1-y)}{\lambda} = x \quad \therefore F^{-1}(y) = \frac{-\ln(1-y)}{\lambda}$

4. $F_X^{-1}(u) = \frac{-1}{\lambda}\ln(1-u) \sim$ Exponential(λ)

2D DISTRIBUTIONS

For any RVs X, Y, their joint (bivariate) distribution is collection of all probabilities of the form

mapped point lies in B

$$P((X,Y) \in B) = P(\{s \in S : (X(s), Y(s)) \in B\}) \quad \forall B \in \mathbb{R}^2$$

events in Sample Space

discrete $p_{X,Y}(x,y) = P(X=x, Y=y) = P(\{X=x\} \cap \{Y=y\})$

① $p_{X,Y}(x,y) \geq 0 \quad \forall x, y$

② $\sum_{x,y} p_{X,Y}(x,y) = 1$

$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = P(\{X \leq x\} \cap \{Y \leq y\})$

① $F(-\infty, -\infty) = F(x, -\infty) = F(-\infty, y) = 0$

② $F(\infty, \infty) = 1$

Marginal CDF $F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y) \quad \forall x \in \mathbb{R}$, $F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x,y) \quad \forall y \in \mathbb{R}$

Marginal PMF $p_X(x) = P(X=x) = \sum_y P(X=x, Y=y) = \sum_y p_{X,Y}(x,y)$, $p_Y(y) = \sum_x P(X=x, Y=y) = \sum_x p_{X,Y}(x,y)$

n^{th} sum $\frac{a(1-r^n)}{1-r}$

Theorem

binomial expansion $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^{n-x} b^x$

X, Y RV, CDF $F_{X,Y}$, let $a \leq b$ and $c \leq d$. Then $P(a < X \leq b, c < Y \leq d)$
$= F_{X,Y}(b,d) - F_{X,Y}(a,d) - F_{X,Y}(b,c) + F_{X,Y}(a,c)$

$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$

$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$

continuous $P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)\, dx\, dy \quad \forall a \leq b, c \leq d$

if $f_{X,Y}(x,y)$ is a constant, $P(a \leq X \leq b, c \leq Y \leq d)$ is basically area of graph × c

① $f_{X,Y}(x,y) \geq 0$

② $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$

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

Conditional Distribution

~~continuous~~ $\frac{P(Y \in B, X=x)}{P(X=x)} = P(Y \in B \mid X=x)$ / $\frac{P(a < Y \leq b, X=x)}{P(X=x)} = P(a < Y \leq b \mid X=x)$

discrete $p_{Y|X}(y|x) = \frac{P(Y=y, X=x)}{P(X=x)} = \frac{p_{X,Y}(x,y)}{p_X(x)}$ *given* $p_X(x) > 0$

continuous $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$

$P(a \le Y \le b \mid X=x) = \int_a^b f_{Y|X}(y|x)\,dy$ *whenever* $a \le b$

$= \int_a^b \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy$

Independent RVs

$P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2)$

$P(a \le X \le b, c \le Y \le d) = P(a \le X \le b)P(c \le Y \le d)$ *whenever* $a \le b$, $c \le d$

discrete $p_{X,Y}(x,y) = p_X(x)\,p_Y(y)$ $\forall x,y \in \mathbb{R}$, $p_{Y|X}(y|x) = p_Y(y)$ *given* $p_X(x) > 0$

continuous (jointly absolutely continuous) $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ $\forall x,y \in \mathbb{R}$, $f_{Y|X}(y|x) = f_Y(y)$ *given* $f(x)$

$\int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}$

if $\begin{cases} p_{X,Y}(x,y) = g(x)h(y) & \forall x,y \in \mathbb{R}^2 \\ f_{X,Y}(x,y) = g(x)h(y) \end{cases}$, *then* $X \perp Y$

some function of x and some function of y

if $X \perp Y$, *then* $f(x) \perp g(Y)$ *for* $f, g: \mathbb{R} \to \mathbb{R}$

Expectation of functions of 2 RV

discrete $E(h(X,Y)) = \sum_{x,y} h(x,y) P(X=x, Y=y) = \sum_{x,y} h(x,y)\, p_{X,Y}(x,y)$

continuous $E(h(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y)\, f_{X,Y}(x,y)\,dx\,dy$

$E(aX + bY) = aE(X) + bE(Y)$

$E(g(X)h(Y)) = E(g(X))E(h(Y))$

VARIANCE, COVARIANCE, CORRELAT

Properties

$Var(X) \geq 0$

$Var(X) = E(X^2) - (E(X))^2$

$Var(X) \le E(X^2) \Leftarrow +ve$

$Var(aX+b) = a^2 Var(X)$

$Var(c) = 0 \quad \forall c \in \mathbb{R}$

$SD(X) = +\sqrt{Var(X)}$

$SD(aX+b) = |a| SD(X)$

$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$

Covariance

$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E($

Correlation

$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ $\rho_{X,Y} \in [-1, 1]$

$Cov(X,Y) = Cov(Y,X)$

$Cov(X, a+Y) = Cov(X,Y)$

$Cov(X,X) = Var(X)$

$Cov(X,Y) = E(XY) - E(X)E(Y)$

$X \perp Y \Rightarrow Cov(X,Y) = 0$

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

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

For any RV $X_1, \cdots, X_n$

possible combinations

$$Var(\sum_i X_i) = \sum_i Var(X_i) + 2\sum_{i<j} cov(X_i, X_j)$$

* if $X \perp Y$, then $Var(X+Y) = Var(X) + Var(Y)$

GENERATING FUNCTION

Moment Generating Function $\quad m(t) = E(e^{tX}) \quad , t \in \mathbb{R}$

$m_X(S) < \infty$ whenever $S \in (-S_0, S_0)$, $S_0 > 0$ $\qquad m_X(t) = r_X(e^t)$

$\Rightarrow \quad m_X(0) = 1 \qquad m_X^{(k)}(0) = E(X^k)$

$m'_X(0) = E(X)$

$m''_X(0) = E(X^2)$

Independence

$X_1, \cdots, X_n$ independent $\Rightarrow m_{\sum X_i}(t) = \prod m_{X_i}(t)$ $\quad$ simplest case:

Uniqueness Theorem $\qquad X \perp Y \Rightarrow m_{X+Y}(t) = m_X(t) m_Y(t)$

X be ~~an~~ RV s.t $\exists S_0 > 0$, $m_X(S) < \infty$ whenever $S \in (-S_0, S_0)$ if Y is another RV with $m_Y(S) = m_X(S)$ whenever $S \in (-S_0, S_0)$ then X and Y have same distribution

$m_X(t) = m_Y(t) \quad \forall t \in (-S_0, S_0) \Rightarrow X \overset{d}{=} Y$

Probability Generating Function $\quad r_X(t) = E(t^X) \quad , t \in \mathbb{R}$

assume ~~$\forall X$~~ $\forall x \in X$, $X \geq 0$, $r_X(t_0) < \infty$ for some $t_0 > 0$. $\qquad r_X(t) = m_X(\ln t)$

$\Rightarrow \quad r_X(0) = P(X=0) \qquad r_X^{(k)}(0) = k!\, P(X=k)$

$r'_X(0) = P(X=1)$

$r''_X(0) = 2P(X=2)$

Conditional Expectation

discrete $\quad E[g(X) \mid X=x] = \sum_y g(y)\, p_{Y|X}(y|x)$

continuous $\quad E[g(Y) \mid X=x] = \int_{-\infty}^{\infty} g(y)\, f_{Y|X}(y|x)\, dy$

Conditional Variance

$V(X \mid Y=y) = E(X^2 \mid Y=y) - (E(X \mid Y=y))^2$

Theorem of Total ~~Probability~~ Expectation

$E(E(X|Y)) = E(X)$

Theorem of Total Variance / Double Expectation

$Var(X) = Var(E(X|Y)) + E(Var(X|Y))$

Markov's Inequality

$X \geq 0 \ \forall x \in X$, X is a RV ; $\forall \alpha > 0$, $P(X \geq \alpha) \leq \frac{E(X)}{\alpha}$ = Markov's upper bound for $P(X \geq a)$

Chebyshev's Inequality

Y be an arbitrary RV with finite mean μ_Y. $\forall \alpha > 0$, $P(|Y - \mu_Y| \geq \alpha) \leq \frac{Var(Y)}{\alpha^2}$
= Chebyshev's upper bound for $P(|Y - \mu_Y| \geq a)$

SAMPLING DISTRIBUTION

Random Sample

Collection of iid RV $(X_1, X_2, \cdots, X_n)$ from some distribution

Statistic

$h(X_1, X_2, \cdots, X_n)$ e.g. $\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$ is a statistic

Sampling Distribution

distribution of a statistic

Averages of RVs

iid $X_1, X_2, \cdots, X_n$ $\mu = E(X_i) \ \forall x_i$ $V(X_i) = \sigma^2 \ \forall x_i$

Define their average $\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$

Mean of Sample mean $(\bar{X}_n)$ = $E(\bar{X}_n) = \mu$

Variance of sample mean $(\bar{X}_n)$ = $V(\bar{X}_n) = \frac{\sigma^2}{n}$

Weak Law of Large Numbers (WLLN)

The more samples, the closer it will be from common mean

$\bar{X}_n$ with finite σ^2 converges to μ i.e. $P(|\bar{X}_n - \mu| \geq \varepsilon) \to 0$ as $n \to \infty$, $\forall \varepsilon > 0$

Types Of Convergence

Z_n WLLN, $\bar{X} \to \mu$

Converges in Probability $X_n \xrightarrow{P} X$ $\lim_{n \to \infty} P(|X_n - X| \geq \varepsilon) = 0$, $\forall \varepsilon > 0$

Converges in Distribution $F_{X_n}(x) \to F_X(x)$ $X_n \xrightarrow{D} X$ $\lim_{n \to \infty} P(X_n \leq x) = P(X \leq x)$ for X s.t. $P(X = x) = 0$

Central Limit Theorem

$X_1, X_2, \cdots, X_n$ be iid with common finite mean μ and finite variance σ^2

$\Rightarrow \bar{X}_n \overset{appr}{\sim} N(\mu, \frac{\sigma^2}{n})$ or $Z_n \sim N(0,1)$ where $Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$

(i) Find ↗ then find Z_n in terms of $\bar{X}_n$

Normal Approximation to Binomial Distribution

$X \sim Binomial(n, p) \Rightarrow X \overset{appr}{\sim} N(np, npq)$

R BASICS

assign variables   x <- 2   x = 2   + enter → runs command and display output

combination   choose(10, 3) <=> $\binom{10}{3}$

Distributions

discrete

d - density/mass (P(X=x))

p - probability (P(X≤x))

X~Binomial(n,p) binom(X, size=n, prob=p)

X~Poisson(λ) ~~poi~~ pois (x, λ)

X~Geometric (p) geom(x, p)

continuous   d - density/mass (P(X=x))

q - quantile (P(X≤x)=θ) we give θ to find x

p - probability (P(X≤x))

X~Normal(μ, σ²)   norm(x, μ, σ)

qnorm(0.975) => P(Z≤?) = 0.975

X~Gamma(α, λ)   gamma(x, α, λ)

other distributions   library(distr)

f <- function(x) 3 * x^2   <=> $f(x) = 3x^2$

X <- AbscontDistribution(d=f, lowl=0, upl=1) <=> $f_X(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & o/w \end{cases}$

p(X)(0.7) <=> P(X≤0.7)

p(X)(0.7) − p(X)(0.2) <=> P(0.2≤X≤0.7) <=> integrate(f, 0.2, 0.7)

other useful examples

Sequence   X <- seq(start, end, step_size)

vector   d <- dgamma(x, α, λ)   where d[i] = P(X = X[i])

first 5 values   head(d,5)

plotting   plot(x, d, type="l")   line

curve(function, start, end, ylab=" ")   add to previous graph   y-values   ylabel

curve(function2, start, end, add=TRUE, col=" ")   colour

-Inf ← -∞

a <- integrate(f, -Inf, 1.645)

output is vector [value, error]

to get value, a$value

DISCRETE DISTRIBUTION

Bernoulli (p)

pmf $p_x(x) = \begin{cases} p & x=1 \\ 1-p & x=0 \\ 0 & o/w \end{cases}$ cdf $F_x(x) = \begin{cases} 0 & x<0 \\ 1-p & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$

E(x) $E(x) = p$ $m_x(t) = 1-p+p(e^t)$

V(x) $V(x) = pq$ $r_x(t) = \cancel{q+pt}\ 1-p+pt$

Binomial (n, p)

pmf $p_x(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x=0,1,\cdots,n \\ 0 & o/w \end{cases}$ $F_x(x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1-p)^{n-i}$

$E(x) = np$ $m_x(t) = (1-p+p(e^t))^n$

$V(x) = npq$ $r_x(t) = (1-p+pt)^n$

Geometric (p)

$p_x(x) = \begin{cases} (1-p)^x p & x=0,1,\cdots \\ 0 & o/w \end{cases}$

$E(x) = \frac{1-p}{p}$ $m_x(t) = \frac{p}{1-e^t(1-p)}$ $t < -\ln(1-p)$

$V(x) = \frac{1-p}{p^2}$ $r_x(t) = \frac{p}{1-t(1-p)}$ $t \in (\frac{-1}{1-p}, \frac{1}{1-p})$

Negative Binomial (r, p)

$p_x(x) = \begin{cases} \binom{x+r-1}{x} (1-p)^x p^r & x=0,1,\cdots \\ 0 & o/w \end{cases}$

$E(x) = \frac{r(1-p)}{p}$ $m_x(t) = \left(\frac{p}{1-e^t(1-p)}\right)^r$

$V(x) = \frac{r(1-p)}{p^2}$ $r_x(t) = \left(\frac{p}{1-t(1-p)}\right)^r$

Poisson (λ)

if $X_i \sim$ Poisson (λ) $Y = X_1 + X_2 + \cdots + X_n$
$Y \sim$ Poisson ($n\lambda$)

$p_x(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & x=0,1,\cdots \\ 0 & o/w \end{cases}$

$E(x) = \lambda$ $m_x(t) = e^{\lambda(e^t-1)}$ $t \in \mathbb{R}$

$V(x) = \lambda$ $r_x(t) = e^{\lambda(t-1)}$

Hypergeometric (N, m, n)

$p_x(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \max(0, n+M-N) \leq x \leq \min(n, M) \\ 0 & o/w \end{cases}$

$E(x) = n\left(\frac{M}{N}\right)$

$V(x) = n\left(\frac{M}{N}\right)\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$

CONTINOUS DISTRIBUTION

Uniform $[L,R]$

$$\cancel{p_X(x} f_X(x) = \begin{cases} \frac{1}{R-L} & L \le x < R \\ 0 & o/w \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < L \\ \frac{x-L}{R-L} & L \le x < R \\ 1 & x \ge R \end{cases}$$

$$E(X) = \frac{L+R}{2}$$

$$m_X(t) = \begin{cases} \frac{e^{tR} - e^{tL}}{t(R-L)} & t \neq 0 \\ 1 & t = 0 \end{cases}$$

$$V(X) = \frac{(R-L)^2}{12}$$

Exponential(λ)

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & o/w \end{cases}$$

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & o/w \end{cases}$$

$$E(X) = \frac{1}{\lambda}$$

$$m_X(t) = \frac{\lambda}{\lambda - t} \quad t < \lambda$$

$$V(X) = \frac{1}{\lambda^2}$$

Gamma (α, λ)

$$f_X(x) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} & x > 0 \\ 0 & o/w \end{cases}$$

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx \qquad \frac{\Gamma(\alpha)}{\lambda^\alpha} = \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx$$

$$E(X) = \frac{\alpha}{\lambda}$$

$$m_X(t) = \frac{\lambda^\alpha}{(\lambda - t)^\alpha} \quad t < \lambda \qquad \Gamma(0.5) = \sqrt{\pi}$$

$$V(X) = \frac{\alpha}{\lambda^2}$$

Normal (μ, σ^2)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty$$

if $Z \sim N(0,1)$ $\quad \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$

$$m_Z(t) = e^{\frac{1}{2}t^2}$$

$$m_X(t) = e^{\mu t + \frac{1}{2}(\sigma t)^2}$$

$$E(X) = \mu$$

$$V(X) = \sigma^2$$

Beta (α, β)

$$f_X(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \\ 0 & o/w \end{cases}$$

$$\beta(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$$E(X) = \frac{\alpha}{\alpha+\beta}$$

$$V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$