MA2216/ST2131 Probability

Lingjie, May 8, 2021

Counting

1. sample space, event space
2. number of possible outcome (multiplicative rule)
3. ordered (permutation) / not ordered (combination)

Sample Space S

mutually exclusive and collectively exhaustive set of random experiment outcomes
$s \in S$ is a sample point

Events E

single set of outcome	$\rightarrow E \subset S$
E occurs if sample point in event	$\rightarrow s \in E$

Set operations

$(E \cap F)^c = E^c \cup F^c$	$(E \cup F)^c = E^c \cap F^c$
$E \cup F = E \cup (F \cap E^c)$	$E = (E \cap F) \cup (E \cap F^c)$

Multiplication Rule

total number of possible outcome $= n_1 \times n_2 \times \ldots \times n_k$

Permutation

distinct ordered objs $\rightarrow \{1,2,3\} \neq \{3,2,1\}$

total: $n!$	only r objs in n objs: $\frac{n!}{(n-r)!}$
circle: $(n-1)!$	

n obj k cells (partition): $\frac{n!}{n_1!n_2!\ldots n_k!}$ = alike items: $\binom{n}{n_1, n_2 \ldots n_k}$

Combination

distinct unordered objs $\rightarrow \{1,2,3\} = \{3,2,1\}$
$\binom{n}{r} = \frac{n!}{(n-r)!r!}$

Binomial Expansion

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

Multinomial Expansion

$(x_1 + x_2 + \ldots + x_k)^n =$
$\sum_{0 \leq x_1+x_2+\ldots+x_k \leq n,\ i_1+i_2+\ldots+i_k=n} \binom{n}{i_1, i_2 \ldots, i_k} x^{i_1} \times x^{i_2} \times \ldots \times x^{i_k}$

Probability Measure/Distribution

a function that takes in an event and output probability
$P(E) \rightarrow [0,1], E \subset S$

Axioms of probability

1. $P(E) \in [0,1]$ | 2. $P(S) = 1$
3. countable additivity:
$P(\cup_{n=1}^\infty E_i) = \sum_{i=1}^\infty P(E_i)$ for $E_i \cup E_j = \emptyset \rightarrow$ disjoint events

Inclusion-Exclusion Principle

non disjoint events $\rightarrow E_i \cup E_j \neq \emptyset$

$$P(\cup_i^n E_i) = \sum_i^n P(E_i) - \sum_{1 \leq j < j \leq n} P(E_i \cap E_j) + \sum_{1 \leq i < j < k \leq n} P(E_i \cap E_j \cap E_k) + \ldots + (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)$$

alternatively

$$|\cup_{i=1}^n A_i| = \sum |singletons| - \sum |pairs| + \sum |triples| - \sum |quadruples| + \ldots + (-1)^{n+1} |n-tuples|$$

Conditional Probability

update posterior based on prior distribution
$P(B|A) = \frac{P(B \cap A)}{P(A)}$ and $P(\cup_i E_i | A) = \sum P(E_i|A)$ (axiom3)

Multiplication Rule for Successive Conditioning

sequential way of gathering info
$P(A \cap B) = P(B|A)P(A)$
$P(E_1 \cap \ldots \cap E_n) =$
$P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \ldots P(E_n|E_1 \cap E_2 \ldots \cap E_{n-1})$

Law of Total probability

$P(A) = \sum_i^n P(E_i)P(A|E_i)$

Bayes' Rule

$P(E_k|A) = \frac{P(E_k)P(A|E_k)}{\sum P(E_i)P(A|E_i)}$
if $n=2$: $P(E|A) = \frac{P(E)P(A|E)}{P(E)P(A|E) + P(E^c)P(A|E^c)}$

Independence of Events

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

jointly independence of multiple events

for any non-empty index $I \subset \{1, \ldots, n\}$
$P(\cap_{i \in I} E_i) = \prod_{i \in I} P(E_i)$

Random Variable

function that takes in sample space S, output image space H
$X : S \rightarrow H, H = \mathbb{R}$
$P(X \in A) = P(E), E = \{s \in S : X(s) \in A\}$

Probability Distribution

Given (S, P) and $X : S \rightarrow \mathbb{R}$
$P_X(A) = P(X \in A) = P(\{s \in S : X(s) \in A\})$
Pre-image of set A: $X^{-1}(A) = \{s \in S : X(s) \in A\}$

Cumulative Prob Distribution Function (F_X)

a function that takes in X and output P
$F_X : \mathbb{R} \rightarrow [0,1]$
$F_X(x) = P_X((-\infty, x]) = P(X \leq x)\ \forall x \in \mathbb{R}$

Discrete Random Variables

$X : S \rightarrow \mathbb{R}$, X is finite or countably infinite

Continuous Random Variables

if $P(X = x) = 0\ \forall\ x \in \mathbb{R}$

Prob. Mass Function $P(X = x)$ [Discrete]

function take takes in $\mathbb{R}$ and output P

$f_X : \mathbb{R} \rightarrow [0,1]$	$f_X(x) = P(X = x)$
$\sum f_X(x) = 1$	$P(X \in A) = \sum_{x \in A} f_X(x)$
$F_X(x) = \sum_{t \leq x} f_X(t)$	$f_X(x) = F_X(x) - F_X(x^-),\ \forall\ x \in \mathbb{R}$

Prob. Density Function (f_X) [Continuous]

absolutely continuous has a well defined f_X
$P(X \in (a,b]) = \int_a^b f_X(x)dx = P(X \in [a,b]) = F(b) - F(a)$

$f_X : \mathbb{R} \rightarrow [0, \infty)$	$\int_{-\infty}^\infty f_X(x)dx = 1$
$F_X(x) = \int_{-\infty}^x f_X(y)dy$	$f_X(x) = F_X'(x)$ if $F'(x)$ exist

Expectation and Covariance

Expectation

weighted sum of the outcome $\sum xP(X = x)$
note not all $E(g(\overrightarrow{X}))$ is well defined

Discrete:	$\sum_i g(x_i) f_X(x_i)$
Continuous:	$\int_{-\infty}^\infty g(x) f_X(x) dx$
Multiple RV:	$E[g(X,Y)]$
	$= \sum_{(x,y)} g(x,y) f_{(X,Y)}(x,y)$
	$= \int \int g(x,y) f_{(X,Y)}(x,y) dxdy$
$E(Y) =$	$E(Y 1_{x<0}) + E(Y 1_{x \geq 0})$

Theorem

Given a real-valued RV Y (e.g. $g(\overrightarrow{X})$)
let $Y^+ := max\{Y, 0\}, Y^- := max\{-Y, 0\}$

Expectation Properties

Comparison:	if $P(X \geq a) = 1 \Rightarrow E(X) \geq a$
Linearity:	$E[ag(x) + bh(x)] = aE[g(x)] + bE[h(x)]$
Monotonicity:	if $X_1 \geq X_2 \Rightarrow E(X_1) \geq E(X_2)$

Variance

$Var(X) = E([X - E(X)]^2) = E(X^2) - E(X)^2$
standard deviation $\sigma = \sqrt{Var(X)}$

- If $E(|Y|) = E(Y^+) + E(Y^-) < \infty$ we define $E(Y) := E(Y^+) - E(Y^-)$
- if $E(Y^+) = \infty, E(Y^-) < \infty \Rightarrow E(Y) := \infty$
- if $E(Y^-) = \infty, E(Y^+) < \infty \Rightarrow E(Y) := -\infty$
- if $E(Y^+) = E(Y^-) = \infty, E(Y)$ is undefined

Variance Properties

- $Var(X) = 0 \Leftrightarrow P(X = E(X)) = 1$
- $E(X^2) = E(X)^2 + Var(X) \Rightarrow E(X^2) \geq E(X)^2$
- $Var(aX + b) = a^2 Var(X)$
- $Var(X_1 + \cdots + X_n) = \sum_{i=1}^n Var(X_i) + 2\sum_{1\leq i<j\leq n} Cov(X_i, X_j)$

k^{th} moment

$E(X^k) = \int x_i^k f_X(x)dx$

$E(X)$ for non-negative integer value X

for changing of starting index

$$E(X) = \sum_{i=0} iP(X = i) = \sum_{i=1} iP(X = i) = \sum_{n=1} P(X \geq n)$$

Covariance

$Cov(X, Y) := E[(X - E(X))(Y - E(Y))]$

Properties

- $Cov(X, Y) = E(XY) - E(X)E(Y)$
- $Cov(\alpha X + a, \beta Y + b) = \alpha\beta Cov(X, Y)$
- independent $\Rightarrow Cov(X, Y) = 0$
- $Cov(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j) = \sum_{i=1}^m \sum_{j=1}^n a_i b_j Cov(X_i, Y_j)$

Correlation Coefficient

$\rho(X, Y) := \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

Properties

- if $\hat{X} := \frac{X-E(X)}{\sqrt{Var(X)}}, \hat{Y} := \frac{Y-E(Y)}{\sqrt{Var(Y)}}$ then $\rho(X, Y) = Cov(\hat{X}, \hat{Y})$
- for any $a, b > 0$, $\rho(aX, bY) = \rho(X, Y)$
- $\rho(X, Y) \in [-1, 1]$
- $\rho(X, Y) = 1$ if $Y = aX$

Mean and Variance of Sums of RV

$E(X_1 + X_2 + \cdots X_n) = E(X_1) + E(X_2) + \cdots E(X_n)$

$Var(X_1 + \cdots + X_n) = \sum_{i=1}^n Var(X_i) + 2\sum_{1\leq i<j\leq n} Cov(X_i, X_j)$
$= Var(X_1) + \cdots + Var(X_n)$ for independent RV

Cauchy-Schwarz Inequality and Correlation

$|\overrightarrow{a} \cdot \overrightarrow{b}| := |\overrightarrow{a}^T \overrightarrow{b}| = |\sum_{i=1}^n a_i b_i| \leq \sqrt{\sum_i a_i^2}\sqrt{\sum_i b_i^2} =: ||a||_2||b||_2$

for RV: $|E(XY)| \leq E(X^2)^{\frac{1}{2}} E(Y^2)^{\frac{1}{2}}$
$\Rightarrow \rho(X, Y) = \frac{E(XY)}{E(X^2)^{1/2}E(Y^2)^{1/2}} \in [-1, 1]$

Conditional Expectation

$E(g(X)|Y = y) = \sum_x g(x) f_{X|Y}(x|y)$

Properties

- $E(g(X)|A) = \frac{E(g(X)1_A)}{P(A)}$
- $E(g(X)) = \sum_{i=1}^n E(g(X)|A_i)P(A_i)$
- $E(g(X, Y)) = E[E(g(X, Y)|Y)]$
- $P(A) = E[P(A|Y)]$
- if X, Y are independent, $E(g(Y)|X) = E(g(Y))$
- $E(g(X)h(Y)|X) = g(X)E(h(Y)|X)$

Expectation of Random Sum

Let $S := \sum_{i=1}^N X_i$
$E(S) = E(\sum_{i=1}^N X_i) = \sum_{i=1}^N E(X_i)$

Note: first-step analysis is useful in solving recursive problems

Conditional Expectation as Orthogonal Projection

$< \overrightarrow{x}, \frac{\overrightarrow{y}}{||\overrightarrow{y}||} > \frac{\overrightarrow{y}}{||y||} =< \overrightarrow{x}, \overrightarrow{y} > \frac{\overrightarrow{y}}{||\overrightarrow{y}||^2}$

Conditional Variance

$Var(X|Y) = E[(X - E(X|Y))^2|Y] = E(X^2|Y) - E(X|Y)^2$
$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$

Discrete Distributions

event, parameter, $P(X = x)$, $E(X)$, $Var(X)$

Bernoulli $Ber(p)$

count no. of success with prob p and failure with prob $1 - p$
$P(X = k) = p^k(1 - p)^{1-k}, k \in [0, 1]$

$P(X = 1) = p$	$P(X = 0) = (1 - p)$
$E(X) = p$	$Var(X) = p(1 - p)$

Indicator RV (I)

X is an indicator random variable for event A
$X(s) = 1_A(s) \Rightarrow [1$ if $s \in A$ elif $s \in A^c$ then $0]$

Binomial $Bin(n, p)$

count n independent Ber with prob p
$\Rightarrow Y = \sum X_i, X \sim Ber(p)$
$P(Y = k) = \binom{n}{k} p^k(1 - p)^{n-k} 1_{0\leq k\leq n}$

$E(Y) = np$	$Var(Y) = np(1 - p)$

Poisson $Pois(\lambda)$

count n Ber with rare prob $p = \frac{\lambda}{n} \Rightarrow \lambda = np$

$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$	$\sum_{k=0}^\infty \frac{\lambda^k}{k!} = e^\lambda$
$E(X) = \lambda$	$Var(X) = \lambda$

Poisson Limit Theorem

$X \sim Ber(\frac{\lambda}{n})$, $Y_n = \sum X_{n,i} \sim Bin(n, \lambda/n)$, $Z \sim Pois(\lambda)$
$\lim_{n\to\infty} P(Y_n = k) = P(Z = k) \ \forall\ k \in \mathbb{N}_0$

Discrete Uniform

$P(X = x_i) = \frac{1}{k} \ \forall\ 1 \leq i \leq k$

$E(X) = \frac{1}{n}\sum X_i$	$Var(X) = \frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2$

F_X is piecewise fn with constant jumps of size $\frac{1}{k}$ at each x_i
$F_X(y) = \frac{1}{k}\sum 1_{\{x_i\leq y\}}, \ y \in \mathbb{R}$

empirical distribution

let μ be discrete prob. measures of a empirical distribution
$\mu(\{x\}) = \frac{1}{n}\sum 1_{y_i=x}, x \in \mathbb{R}$

Geometric $Geom(p)$

count k Ber with p till 1^{st} success appears
$P(X = k) = p(1 - p)^{k-1}, k \in \mathbb{N}$
$E(X) = \frac{1}{p}$ $\quad Var(X) = \frac{1-p}{p^2}$

tail probability

prob that 1^{st} success is after k^{th} count
$P(X \geq k) = (1-p)^{k-1}$

memorylessness

previous k counts do not change prob of 1^{st} success appear in i^{th} trial
$P(X-k=i \mid X>k) = P(X=i)\ \forall\ i \in \mathbb{N}$

Negative Binomial $NB(r,p)$

count n Ber with p till r^{th} success appear
$\Rightarrow X = \sum Y_i,\ Y \sim Geom(p)$ with i^{th} as success
$P(X=n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, n \geq r$
$E(X) = \frac{r}{p} \mid Var(X) = \frac{r(1-p)}{p^2}$

Hypergeometric distribution $H(n,N,m)$

sample n balls without replacement from N balls, with m white balls and $N-m$ black balls and get i white balls
$P(X=i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{m}},\ 0 \leq i \leq n$
$E(X) = n\frac{m}{N} \mid Var(X) = \left[\frac{N-n}{N-1}\right] n \left[\frac{m}{N}\right](1-\frac{m}{N})$
if sample with replacement: $X \sim Bin(n, \frac{m}{N})$
and $E(X) = \sum E(\S_i) = n\frac{m}{N}$

Continuous RV

event, parameter, $f_X, F_X, E(X), Var(X)$, properties

Uniform distribution $U(a,b)$

probability of choosing a random point in a continuous line starting from a and end at b
$f_X(x) = \frac{1_{[a,b]}(x)}{b-a} \mid F_X(x) = \frac{x-a}{b-a}$
$E(X) = \frac{a+b}{2} \mid Var(X) = \frac{(b-a)^2}{12}$
$P(x<X<y) = \frac{y-x}{b-a}$

Exponential distribution $Exp(\lambda)$

waiting time for event to happen
λ is the rate (e.g. clock ticks)
$f_X(x) = \lambda e^{-\lambda x} 1_{\{x>0\}} \mid F_X(x) = 1 - e^{-\lambda x}$
$E(X) = \frac{1}{\lambda} \mid Var(X) = \frac{1}{\lambda^2}$

tail probability

waiting time for event to happen after t
$P(X>t) = e^{-\lambda t}$
$P(X>t) = e^{-\int_{\text{start}}^{t} \lambda(t)dt}$
$P(X>t|X>a) = e^{-\int_a^t \lambda(t)dt}$

memoryless property

$P(X \geq t+s \mid X \geq t) = P(X \geq s)$

Approximation: $Geom$ to Exp

discretise time (geom) into smaller band till continuous time (exp)
$X \sim Geom(\delta\lambda), Y \sim exp(\lambda), \delta \in (0, \frac{1}{\lambda}), \lambda > 0$ (fixed)
$\lim_{\delta \downarrow 0} P(\delta X_\delta > t) = e^{-\lambda t} = P(Y>t)$

Gamma distribution Γ(shape:α, rate:λ)

waiting time for α^{th} event to occur
Gamma is the sum of independent $Exp(\lambda)$

$$\begin{aligned}
f_Y(y) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} 1_{\{y>0\}} \\
F_Y(y) &= \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \lambda y) \\
E(Y) &= \frac{\alpha}{\lambda} \\
Var(Y) &= \frac{\alpha}{\lambda^2} \\
cY &\sim \Gamma(\alpha, \tfrac{1}{c}\lambda) \\
E(Y^r) &= \frac{\Gamma(\alpha+r)}{\lambda^r \Gamma(\alpha)}, r > -\alpha
\end{aligned}$$

Gamma function $\Gamma(\alpha)$

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^\infty x^{\alpha-1} e^{-x} dx \\
\Gamma(\alpha, x) &= \int_x^\infty x^{\alpha-1} e^{-x} dx \\
\gamma(\alpha, x) &= \int_0^x x^{\alpha-1} e^{-x} dx \\
\Gamma(\alpha+1) &= \alpha\Gamma(\alpha) \\
\Gamma(n+1) &= n!\Gamma(1) = n! \\
\Gamma(1/2) &= \sqrt{\pi}
\end{aligned}$$

Normal distribution $N(\mu, \sigma^2)$

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
$F_X(x) = \int_{-\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ (Z can't be simplified)
$E(X) = \mu \mid Var(X) = \sigma^2$

Normal approximation of Binomial distribution

$X \sim B(n,p),\ Z \sim N(0,1)$
$P(\frac{X-np}{\sqrt{np(1-p)}} \in (a,b)) \to P(Z \in (a,b))$ as $n \to \infty$
as long as $Var(X) = np(1-p)$ is large enough
if $np \approx \lambda$ or $n(1-p) \approx \lambda \Rightarrow B(n,p) \approx Pois(\lambda)$

Continuity correction cc

$P(X=x) = P(X \in (x-\frac{1}{2}, x+\frac{1}{2}) \approx P(Y \in (x-\frac{1}{2}, x+\frac{1}{2}))$

Affine transformation of Normal are Normal

$X \sim N(\mu, \sigma^2),\ Y = aX+b \Rightarrow Y \sim N(\alpha\mu + b, a^2\sigma^2)$

Double exponential Laplace(μ, λ)

$$\begin{aligned}
f_X(x) &= \frac{\lambda}{2} e^{-\lambda|x-\mu|} \\
F_X(x) &= \begin{cases} \frac{1}{2} exp(\lambda(x-\mu)) \\ 1 - \frac{1}{2} exp(-\lambda(x-\mu)) \end{cases} \\
E(X) &= \mu \\
Var(X) &= 2/\lambda^2
\end{aligned}$$

Functions of RV

Remember Domain of new RV = Range of old RV
Remember to find Range of Y
X with f_X and $Y = g(x)$ with f_Y

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(X \leq g^{-1}(y)) \\
&= F_X(g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X dx \\
f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) \\
&= f_X(g^{-1}(y)) |\frac{d}{dy} g^{-1}(y)|
\end{aligned}$$

χ^2 distribution

$Z \sim N(0,1), Y = Z^2$
$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{\frac{1}{2}-1} e^{-\frac{y}{2}} 1_{y>0} = \Gamma(\frac{1}{2}, \frac{1}{2})$
$F_Y(y) = \frac{1}{\Gamma(\frac{n}{2})} \gamma(\frac{n}{2}, \frac{x}{2})$
$\chi_n^2 = \Gamma(\frac{n}{2}, \frac{1}{2})$, if $Y = Z_1^2 + \cdots + Z_n^2$
$E(Y) = n \mid Var(Y) = 2n$

Lognormal distribution

$X \sim N(\mu, \sigma^2), Y = e^X$
$F_Y(y) = F_X(\log(y)) \mid f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\log y - \mu)^2}{2\sigma^2}}$
$E(X) = e^{\mu + \frac{\sigma^2}{2}} \mid Var(X) = [e^{\sigma^2} - 1] e^{2\mu + \sigma^2}$

Discretising Exponential distribution

$X \sim Exp(\lambda), Y = [X] + 1$
where $[x] = k$ if and only if $x \in [k, k+1) \forall k \in \mathbb{N}$
$f_Y(n) = P(Y=n) = P(X \in [n-1, n)) = P(X \geq n-1) - P(X \geq n)$
$= e^{-\lambda(n-1)}(1 - e^{-\lambda}) \Rightarrow Y \sim Geom(p), p = 1 - e^{-\lambda}$

Multiples of Exponentials are Exponentials

$X \sim Exp(1), Y = \frac{X}{\lambda}, \lambda > 0$
$F_Y(y) = P(Y \leq y) = P(\frac{X}{\lambda} \leq y) = F_X(\lambda y)$
$f_Y(y) = \lambda e^{-\lambda y} \Rightarrow \frac{X}{\lambda} \sim exp(\lambda)$

Weibull distribution

$F_X(x) = (1 - e^{-(\frac{x-\nu}{\alpha})^\beta}) 1_{x>\nu}$
$f_X(x) = \frac{\beta}{\alpha} (\frac{x-\nu}{\alpha})^{\beta-1} e^{-(\frac{x-\nu}{\alpha})^\beta}$
Let $Y = (\frac{x-\nu}{\alpha})^\beta$
$F_Y(y) = P(X \leq \alpha y^{1/\beta} + \nu) = F_X(\alpha y^{1/\beta} + \nu) = 1 - e^{-y} \Rightarrow Y \sim exp(1)$

Multiple distributions

Joint Probability Mass Function

$f_{(x,y)}(x,y) = P(X=x, Y=y)$

Properties

- $f_{(x,y)}(x,y) \geq 0\ \forall x,y$
- $\sum_x \sum_y f_{(x,y)}(x,y) = 1$
- $P((X,Y) \in A) = \sum_{(x,y)\in A} f_{(X,Y)(x,y)}$
- $f_X(x) = P(X=x) = \sum_y f_{(X,Y)}(x,y)$

Joint Probability Density Function

Definition

- $f_{(X,Y)}(x,y) \geq 0 \quad \forall x,y \in R^2$
- $\int_R \int_R f_{(X,Y)}(x,y)dxdy = 1$
- $\forall A \in R^2, P((X,Y) \in A) = \int\int_A f_{(X,Y)}(x,y)dxdy$

Properties
$\int_R f_{(X,Y)}(x,y)dy = f_X(x)$

Joint c.d.f and Marginal c.d.f

joint cdf:

- $F_{(X,Y)}(x,y) := P(X \leq x, Y \leq y)$
- $\int_{-\infty}^{x} \int_{-\infty}^{y} f_{(X,Y)}(a,b)dadb$
- $f_{(X,Y)}(x,y) = \frac{\partial^2 F_{(X,Y)}}{\partial x \partial y}(x,y)$

marginal cdf:

- $F_x(x) = P(X \leq x)$
- $= \lim_{y\to\infty} P(X \leq x, Y \leq y)$
- $= \lim_{y\to\infty} F_{(X,Y)}(x,y)$

Conditional Distribution

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

Independent Random Variables

$f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$

Independent Random Variables (3 or more)

$P(X_1 \in A_1, \cdots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$

Multi-dimensional Change of Variables

(X_1, X_2) be two RV with joint pdf $f(x_1, x_2)$
Identify $(Y_1, Y_2) := (g_1(X_1,X_2), g_2(X_1,X_2))$

For 1 dimensional: $f_Y(y) = \frac{f_X(x)}{|g'(x)|} = f_X(x)h'(y), x = h(y)$

For multidimension:
$f_{\overrightarrow{Y}}(\overrightarrow{y}) = \frac{f_{\overrightarrow{X}} \overrightarrow{x}}{|J_{\overrightarrow{g}} \overrightarrow{x}|} = f_{\overrightarrow{X}} \overrightarrow{x} |J_{\overrightarrow{h}} \overrightarrow{y}|, \overrightarrow{x} = \overrightarrow{h}(\overrightarrow{y})$

Jacobian: $$J_{\overrightarrow{g}}(\overrightarrow{x}) := \begin{vmatrix} \frac{\partial g_1}{\partial x_1}(\overrightarrow{x}) & \frac{\partial g_1}{\partial x_2}(\overrightarrow{x}) \\ \frac{\partial g_2}{\partial x_1}(\overrightarrow{x}) & \frac{\partial g_2}{\partial x_2}(\overrightarrow{x}) \end{vmatrix} = \frac{\partial g_1}{\partial x_1}(\overrightarrow{x})\frac{\partial g_2}{\partial x_2}(\overrightarrow{x}) - \frac{\partial g_1}{\partial x_2}(\overrightarrow{x})\frac{\partial g_2}{\partial x_1}(\overrightarrow{x})$$

Procedure for change of variable

1. identify the transformation
2. determine range of $\overrightarrow{g}$ and hence that of $\overrightarrow{Y}$
3. identify the inverse
4. compute Jocobian
5. write down the answer

Higher Dimensional Change of Variable defined similarly

Multiple Independent RV

Sums of Independent Continuous RV

$g(X,Y) = X + Y$
$f_{X+Y}(z) = (f_X * f_Y)(z) := \int f_X(z-y) f_Y(y)dy$

Assumption: X, Y independent continuous random variable
$f * g = g * f$

Two Independent $Exp(\lambda)$

$X, Y \sim Exp(\lambda)$ and independent
$f_{X+Y}(z) = \lambda^2 z e^{-\lambda z} 1_{\{z>0\}} = \Gamma(2, \lambda)$

$$\begin{aligned} f_{X+Y}(z) &= \int f_X(z-y) f_Y(y)dy \\ &= \lambda^2 \int_R e^{-\lambda(z-y)} 1_{\{z-y>0\}} e^{-\lambda y} 1_{\{y>0\}} \\ &= \lambda^2 \int_0^z e^{-\lambda z} dy \\ &= \lambda^2 z e^{-\lambda z} \end{aligned}$$

Two Independent $\Gamma(\cdot, \lambda)$

$\alpha, \beta, \lambda > 0, X \sim \Gamma(\alpha, \lambda), Y \sim \Gamma(\beta, \lambda)$
$X + Y \sim \Gamma(\alpha + \beta, \lambda)$

Two Independent $Uniform$

$X, Y \sim uniform(0,1)$
$f_{X+Y}(z) = \int 1_{[0,1]}(z-y) 1_{[0,1]}(y)dy$

Two Independent $Normal$

$X_1 \sim N(\mu_1, \sigma_1^2, X_2 \sim N(\mu_2, \sigma_2^2$
$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Two Independent Integer Valued

X, Y be independent integer-valued r.v. with pmf f_X, f_Y
$f_{X+Y}(n) P(X+Y=n) = \sum_{i\in Z} P(X = n-i, Y = i)$
$= \sum_i f_X(n-i) f_Y(i) =: (f_X * f_Y)(n)$

Two Independent $Poisson$

$X_1 \sim Pois(\lambda_1), X_2 \sim Pois(\lambda_2)$
then $X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$

Normal Distribution [Special]

Linear Transformation of Independent Z

Let $X_1, \cdots X_n$ be i.i.d. $N(0,1)$
pdf: $f_{\overrightarrow{X}}(x_1, \cdots x_n) = \frac{1}{\sqrt{(2\pi)^n det(\sum)}} e^{-\frac{1}{2}\overrightarrow{y}^T \sum^{-1} \overrightarrow{y}}$

Covariance Matrix

Given a vector of RV $\overrightarrow{X} := (X_1, \cdots, X_n)$ with mean $\overrightarrow{\mu} = (\mu_1, \cdots, \mu_n)$, its covariance matrix is defined by:
$\sum_{\overrightarrow{X}} := (Cov(X_i, X_j))_{1\leq i,j\leq n} = E[(\overrightarrow{X} - \overrightarrow{\mu})(\overrightarrow{X} - \overrightarrow{\mu})^T]$

Multivariable Normal Distribution

Since $\overrightarrow{Y} = A\overrightarrow{X}$, we have
$E(\overrightarrow{Y}) = AE(\overrightarrow{X}) = \overrightarrow{0}$

For $\overrightarrow{Y} = A\overrightarrow{X} + \overrightarrow{\mu}$, we have
$E(\overrightarrow{Y}) = AE(\overrightarrow{X}) + \overrightarrow{\mu} = \overrightarrow{\mu}$
pdf: $f_{\overrightarrow{X}}(x_1, \cdots x_n) = \frac{1}{\sqrt{(2\pi)^n det(\sum)}} e^{-\frac{1}{2}(\overrightarrow{z} - \overrightarrow{\mu})^T \sum^{-1} (\overrightarrow{z} - \overrightarrow{\mu})}$

Marginal Distribution of Multivariate Normal

Given i.i.d. normal distribution:
$\overrightarrow{X} = (X_1, \cdots X_n) \sim (\overrightarrow{\mu}, \sum)$
$\overrightarrow{Z} = (Z_1, \cdots Z_n) \sim (\overrightarrow{0}, \overrightarrow{I})$
Theorems:

1. affine transformations of iid standard normal is multivariate normal
   if $\sum_{ii} = \sigma_i^2, \sum_{ij} = 0\ \forall\ i \neq j$
   $\Rightarrow X_1, \cdots, X_n$ independent with $X_i \sim N(\mu_i, \sigma_i^2)$
2. if $A_{n\times n}$ with $det(A) \neq 0$
   $\Rightarrow \overrightarrow{Y} := A\overrightarrow{Z} + \overrightarrow{\mu} \sim N(\overrightarrow{\mu}, \sum)$ with $\sum = AA^T$

3. affine transformation of multivariate normal are also multivariate normal
   if $\overrightarrow{Y} = A\overrightarrow{X} + \overrightarrow{v}$
   $\Rightarrow \overrightarrow{Y} \sim N(A\overrightarrow{\mu} + \overrightarrow{v}, \sum_{\overrightarrow{Y}})$
   with $\sum_{\overrightarrow{Y}} = A \sum_{\overrightarrow{X}} A^T$
4. For $\overrightarrow{X}$, there exist an $n \times n$ matrix A and iid $\overrightarrow{Z}$ such that
   $\overrightarrow{X} = A\overrightarrow{Z} + \overrightarrow{\mu}$
5. Let $\overrightarrow{Y} = (Y_1, \cdots, Y_m)$ be a subset of $\overrightarrow{X}$ with $m < n$
   then $\overrightarrow{Y} \sim (\overrightarrow{\mu}_{\overrightarrow{Y}}, \sum_{\overrightarrow{Y}})$

Conditional Distribution of Multivariable Normal

Suppose $\overrightarrow{W} = (X_1, \cdots, X_m; Y_1, \cdots, Y_n) \sim N(\overrightarrow{\mu}, \sum)$ is multivariate normal. Then given
$\overrightarrow{X} = (X_1, \cdots, X_m) = (x_1, \cdots, x_m)$, $\overrightarrow{Y} = (Y_1, \cdots, Y_n)$ is also multivariate normal

$f_{\overrightarrow{Y}|\overrightarrow{X}}(\overrightarrow{y}|\overrightarrow{x}) = \frac{f_{\overrightarrow{X};\overrightarrow{Y}}(\overrightarrow{x};\overrightarrow{y})}{f_{\overrightarrow{X}}(\overrightarrow{x})} = Ce^{-Q(\overrightarrow{y}|\overrightarrow{x})}$

Finding independent Normal

Goal: find matrix B s.t. $\overrightarrow{Y}' := \overrightarrow{Y} - B\overrightarrow{X}$ is a normal vector independent of $\overrightarrow{X}$
$\Leftrightarrow \overrightarrow{Y}'$ independent of $\overrightarrow{X} \Leftrightarrow Cov(X_i, Y_j') = 0$

$\overrightarrow{Y} = B\overrightarrow{X} + \overrightarrow{Y}' \sim N(B\overrightarrow{x} + \overrightarrow{\mu}_{\overrightarrow{Y}'}, \sum_{\overrightarrow{Y}'})$, conditioned on $\overrightarrow{X} = \overrightarrow{x}$
setting $Cov(X_i, Y_j') = 0, B = \sum_{\overrightarrow{Y},\overrightarrow{X}} \sum_{\overrightarrow{X}}^{-1}$

$\overrightarrow{\mu}_{\overrightarrow{Y}'} = \overrightarrow{\mu}_{\overrightarrow{Y}} - B\overrightarrow{\mu}_{\overrightarrow{X}}$
$\sum_{\overrightarrow{Y}'} = \sum_{\overrightarrow{Y}} - \sum_{\overrightarrow{Y},\overrightarrow{X}} \sum_{\overrightarrow{X}}^{-1} \sum_{\overrightarrow{X},\overrightarrow{Y}}$

refer to lecture 17 for detailed working

Properties of Bivariate Normal

$\overrightarrow{X} \sim N(\overrightarrow{\mu}, \sum)$

1. $f_{\overrightarrow{X}}(\overrightarrow{x}) = \frac{1}{2\pi\sqrt{det(\sum)}} e^{-\frac{1}{2}(\overrightarrow{x}-\overrightarrow{\mu})^T \sum^{-1}(\overrightarrow{x}-\overrightarrow{\mu})}$
   $E(\overrightarrow{X}) = \overrightarrow{\mu}$
   $Cov(X_i, X_j) = \sum_{ij}, 1 \le i, j \le 2$
2. X_1, X_2 are independent $\Leftrightarrow Cov(X_1, X_2) = 0$
3. Affine transformation $\overrightarrow{Y} := A\overrightarrow{X} + \overrightarrow{v}$, where $det(A) \neq 0$ preserves normality
4. We can find $\overrightarrow{Z} \sim N(0, I)$ a matrix A s.t. $\overrightarrow{X} = A\overrightarrow{Z} + \overrightarrow{\mu}$

Law of Large Numbers

Moment Generating Functions

$M_X(t) := E(e^{tX})$

$$M(t) = E[e^{tX}] = \begin{cases} \sum_i e^{tx_i} p(x_i) \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx \end{cases}$$

The joint mgf of $(f_1, f_2, \cdots, f_n)$

$$M(t_1, \cdots, t_n) = E[e^{t_1 f_1 t_2 f_2 \cdots t_n f_n}]$$

Theorem 1: $E(X^k) = M^{(k)}(0) = \frac{d^k M(t)}{d^k t}|_{t=0}$

Proposition 1 [Multiplicative Property]:
if X, Y are independent, $M_{X+Y}(t) = M_X(t) M_Y(t)$
Proposition 2 [Uniqueness Property]:
if $M_X(t) = M_Y(t)$, then X, Y have the same distribution

Common Moments

$Ber(p)$: $M_X(t) = 1 - p + pe^t$
$Bin(n,p)$: $M_X(t) = (1 - p + pe^t)^n$
$Geom(p)$: $M_X(t) = \frac{pe^t}{1-e^t(1-p)}, t < log\frac{1}{1-p}$
$Exp(\lambda)$: $M_X(t) = \frac{\lambda}{\lambda - t}, t < \lambda$
$Pois(\lambda)$: $M_X(t) = e^{-\lambda(1-e^t)}$
$N(\mu, \sigma^2)$: $M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$
$N(0, \sigma^2)$: $M_X(t) = e^{\frac{\sigma^2 t^2}{2}}$
$\Gamma(\alpha, \lambda)$: $M_X(t) = \frac{\lambda}{\lambda - t}^{\alpha}, t < \lambda$

Markov's Inequality

if X is a non-negative RV
$P(X > a) \le \frac{E(X)}{a}$

Chebyshey's Inequality

Suppose $E(X) = \mu, Var(X) = \sigma^2 < \infty$
$P(|X - \mu| > a) \le \frac{Var(X)}{a^2}$

Convergence in Probability

A sequence of RV $(X_n)_{n \in N}$ is said to converge in probability to a RV Y, if for all $\epsilon > 0$
$P(s : |X_n(s) - Y(s)| > \epsilon) \to 0, n \to \infty$

Almost Sure Convergence

A sequence of RV $(X_n)_{n \in N}$ is said to converge almost surely to a RV Y, if with probability of 1
$|X_n(s) - Y(s)| \to 0, n \to \infty$

Convergence in Distribution

A sequence of RV's $(X_n)_{n \in N}$ with cdf F_n is said to converge in distribution to a RV Y with cdf F if
$F_n(x) \to F(x), n \to \infty$

Weak Law of Large Numbers

Let $(X_n)_{n \in N}$ be i.i.d. RV with finite mean $E(X) = \mu$
Let $S_n := \frac{1}{n} \sum_{i=1}^n X_i$ Then for any $\epsilon > 0$
$P(|S_n - \mu| > \epsilon) \to 0$ as $n \to \infty$

the probability distribution of S_n concentrates more and more around its mean as n gets larger and larger

Strong Law of Large Numbers

Let $(\S_n)_{n \in N}$ be i.i.d. RV with finite mean $E(\S_1) = \mu$
Let $S_n = \frac{1}{n} \sum_{i=1}^n \S_i$. Then almost surely,
$|S_n - \mu| \to 0, n \to \infty$

Central Limit Theorem

Let $(X_n)_{n \in N}$ be i.i.d. RV with $E(X_1) = \mu, Var(X_1) = \sigma^2$
$W_n := \frac{X_1 + \cdots + X_n - n\mu}{\sqrt(n)} \to Z \sim N(0, \sigma^2), n \to \infty$ in distribution