

Counting

- sample space, event space
- number of possible outcome (multiplicative rule)
- 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

$$\begin{array}{l|l} (E \cap F)^c = E^c \cup F^c & (E \cup F)^c = E^c \cap F^c \\ E \cup F = E \cup (F \cap E^c) & E = (E \cap F) \cup (E \cap F^c) \end{array}$$

Multiplication Rule

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

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): $\frac{n!}{n_1!n_2!\dots n_k!} = \text{alike items: } \binom{n}{n_1, n_2, \dots, n_k}$

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

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

Probability Measure/Distribution

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

Axioms of probability

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

Inclusion-Exclusion Principle

$$\begin{array}{l} \text{non disjoint events} \rightarrow E_i \cup E_j \neq \emptyset \\ P(\cup_i^n E_i) = \sum_i^n P(E_i) - \sum_{1 \leq j < k \leq n} P(E_i \cap E_j) \\ \quad + \sum_{1 \leq i < j < k \leq n} P(E_i \cap E_j \cap E_k) + \dots \\ \quad + (-1)^{n+1} P(E_1 \cap E_2 \cap \dots \cap E_n) \end{array}$$

alternatively

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

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
 $P(A \cap B) = P(B|A)P(A)$
 $P(E_1 \cap \dots \cap E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \dots P(E_n|E_1 \cap E_2 \dots \cap E_{n-1})$

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

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

jointly independence of multiple events

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

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

Prob. Density Function (f_X) [Continuous]

absolutely continuous has a well defined f_X
 $P(X \in (a, b]) = \int_a^b f_X(x) dx = P(X \in [a, b]) = F(b) - F(a)$
 $f_X : \mathbb{R} \rightarrow [0, \infty) \mid \int_{-\infty}^{\infty} f_X(x) dx = 1$
 $F_X(x) = \int_{-\infty}^x f_X(y) dy \mid f_X(x) = F'_X(x) \text{ if } F'(x) \text{ exist}$

Expectation and Covariance

Expectation

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

$$\begin{array}{ll} \text{Discrete:} & \sum_i g(x_i) f_X(x_i) \\ \text{Continuous:} & \int_{-\infty}^{\infty} g(x) f_X(x) dx \\ \text{Multiple RV:} & E[g(X, Y)] \\ & = \sum_{(x, y)} g(x, y) f_{(X, Y)}(x, y) \\ & = \int \int g(x, y) f_{(X, Y)}(x, y) dx dy \\ E(Y) & = E(Y1_{x < 0}) + E(Y1_{x \geq 0}) \end{array}$$

Theorem

Given a real-valued RV Y (e.g. $g(\vec{X})$)
 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

$$\begin{aligned} \text{Var}(X) &= E([X - E(X)]^2) = E(X^2) - E(X)^2 \\ \text{Var}(X) &= \sum_{i=1}^n (x_i - \mu)^2 P(X = x_i) \\ \text{standard deviation } \sigma &= \sqrt{\text{Var}(X)} \end{aligned}$$

- 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

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

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

$$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} iP(X = i) = \sum_{i=1}^{\infty} iP(X = i) \\ &= \sum_{n=1}^{\infty} P(X \geq n) \end{aligned}$$

Covariance

$$\text{Cov}(X, Y) := E[(X - E(X))(Y - E(Y))]$$

Properties

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

Correlation Coefficient

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

Properties

- if $\hat{X} := \frac{X - E(X)}{\sqrt{\text{Var}(X)}}, \hat{Y} := \frac{Y - E(Y)}{\sqrt{\text{Var}(Y)}}$ then
 $\rho(X, Y) = \text{Cov}(\hat{X}, \hat{Y})$
- 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

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

$$\begin{aligned} \text{Var}(X_1 + \dots + X_n) &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \\ &= \text{Var}(X_1) + \dots + \text{Var}(X_n) \text{ for independent RV} \end{aligned}$$

Cauchy-Schwarz Inequality and Correlation

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

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

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

$$\begin{aligned} \text{Let } S &:= \sum_{i=1}^N X_i \\ E(S) &= E(\sum_{i=1}^N X_i) = \sum_{i=1}^N E(X_i) \end{aligned}$$

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

Conditional Expectation as Orthogonal Projection

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

Conditional Variance

$$\begin{aligned} \text{Var}(X|Y) &= E[(X - E(X|Y))^2|Y] = E(X^2|Y) - E(X|Y)^2 \\ \text{Var}(X) &= E(\text{Var}(X|Y)) + \text{Var}(E(X|Y)) \end{aligned}$$

Discrete Distributions

event, parameter, $P(X = x), E(X), \text{Var}(X)$

Bernoulli $Ber(p)$

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

Indicator RV (I)

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

Binomial $Bin(n, p)$

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

Poisson $Pois(\lambda)$

count n Ber with rare prob $p = \frac{\lambda}{n} \Rightarrow \lambda = np$
 $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \mid \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$
 $E(X) = \lambda \mid \text{Var}(X) = \lambda$

Poisson Limit Theorem

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

Discrete Uniform

$P(X = x_i) = \frac{1}{k} \forall 1 \leq i \leq k$
 $E(X) = \frac{1}{n} \sum X_i \mid \text{Var}(X) = \frac{1}{n} \sum X_i^2 - (\frac{1}{n} \sum X_i)^2$
 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 1st success appears
 $P(X = k) = p(1 - p)^{k-1}, k \in \mathbb{N}$
 $E(X) = \frac{1}{p} \mid \text{Var}(X) = \frac{1-p}{p^2}$

tail probability

prob that 1st success is after k^{th} count

$$P(X \geq k) = (1 - p)^{k-1}$$

memorylessness

previous k counts do not change prob of 1st success appear in i^{th} trial

$$P(X - k = i \mid X > k) = P(X = i) \quad \forall i \in \mathbb{N}$$

Negative Binomial $NB(r, p)$

count n Ber with p till r^{th} success appear

$\Rightarrow X = \sum Y_i, Y \sim \text{Geom}(p)$ 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

$$P(X = i) = \frac{\binom{m}{i} \binom{N-m}{n-i}}{\binom{N}{n}}, 0 \leq i \leq n$$

$$E(X) = n \frac{m}{N} \mid Var(X) = \left[\frac{N-n}{N-1} \right] n \left[\frac{m}{N} \right] \left(1 - \frac{m}{N} \right)$$

$$\text{if sample with replacement: } X \sim \text{Bin}(n, \frac{m}{N}) \\ \text{and } E(X) = \sum E(\xi_i) = n \frac{m}{N}$$

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

$$P(X > t) = e^{-\lambda t}$$

$$P(X > t) = e^{-\int_{\text{start}}^t \lambda(t) dt}$$

$$P(X > t \mid X > a) = e^{-\int_a^t \lambda(t) dt}$$

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)

$$X \sim \text{Geom}(\delta\lambda), Y \sim \text{exp}(\lambda), \delta \in (0, \frac{1}{\lambda}), \lambda > 0 \text{ (fixed)}$$

$$\lim_{\delta \downarrow 0} P(\delta X_\delta > t) = e^{-\lambda t} = P(Y > t)$$

Gamma distribution $\Gamma(\text{shape}:\alpha, \text{rate}:\lambda)$

waiting time for α^{th} event to occur

Gamma is the sum of independent $Exp(\lambda)$

$$f_Y(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} 1_{\{y>0\}}$$

$$F_Y(y) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \lambda y)$$

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

$$Var(Y) = \frac{\alpha}{\lambda^2}$$

$$cY \sim \Gamma(\alpha, \frac{1}{c}\lambda)$$

$$E(Y^r) = \frac{\Gamma(\alpha+r)}{\lambda^r \Gamma(\alpha)}, r > -\alpha$$

Gamma function $\Gamma(\alpha)$

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

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

$$\gamma(\alpha, x) = \int_0^x x^{\alpha-1} e^{-x} dx$$

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$$

$$\Gamma(n + 1) = n! \Gamma(1) = n!$$

$$\Gamma(1/2) = \sqrt{\pi}$$

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

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

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (Z \text{ can't be simplified})$$

$$E(X) = \mu \mid Var(X) = \sigma^2$$

Normal approximation of Binomial distribution

$$X \sim B(n, p), Z \sim N(0, 1)$$

$$P\left(\frac{X - np}{\sqrt{np(1-p)}} \in (a, b)\right) \rightarrow P(Z \in (a, b)) \text{ as } n \rightarrow \infty$$

as long as $Var(X) = np(1-p)$ is large enough

if $np \approx \lambda$ or $n(1-p) \approx \lambda \Rightarrow B(n, p) \approx \text{Pois}(\lambda)$

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

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

Double exponential Laplace(μ, λ)

$$f_X(x) = \frac{\lambda}{2} e^{-\lambda|x-\mu|}$$

$$F_X(x) = \begin{cases} \frac{1}{2} \exp(\lambda(x - \mu)) \\ 1 - \frac{1}{2} \exp(-\lambda(x - \mu)) \end{cases}$$

$$E(X) = \mu$$

$$Var(X) = 2/\lambda^2$$

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

$$F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y))$$

$$= F_X(g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X dx$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g^{-1}(y))$$

$$= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

χ^2 distribution

$$Z \sim N(0, 1), Y = Z^2$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{\frac{1}{2}-1} e^{-\frac{y}{2}} 1_{y>0} = \Gamma(\frac{1}{2}, \frac{1}{2})$$

$$F_Y(y) = \frac{1}{\Gamma(\frac{n}{2})} \gamma(\frac{n}{2}, \frac{x}{2})$$

$$\chi_n^2 = \Gamma(\frac{n}{2}, \frac{1}{2}), \text{ if } Y = Z_1^2 + \dots + Z_n^2$$

$$E(Y) = n \mid Var(Y) = 2n$$

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

$$X \sim \text{Exp}(\lambda), Y = [X] + 1$$

where $[x] = k$ if and only if $x \in [k, k+1) \forall k \in \mathbb{N}$

$$f_Y(n) = P(Y = n) = P(X \in [n-1, n)) = P(X \geq n-1) - P(X \geq n)$$

$$= e^{-\lambda(n-1)}(1 - e^{-\lambda}) \Rightarrow Y \sim \text{Geom}(p), p = 1 - e^{-\lambda}$$

Multiples of Exponentials are Exponentials

$$X \sim \text{Exp}(1), Y = \frac{X}{\lambda}, \lambda > 0$$

$$F_Y(y) = P(Y \leq y) = P\left(\frac{X}{\lambda} \leq y\right) = F_X(\lambda y)$$

$$f_Y(y) = \lambda e^{-\lambda y} \Rightarrow \frac{X}{\lambda} \sim \text{exp}(\lambda)$$

Weibull distribution

$$F_X(x) = (1 - e^{-(\frac{x-\nu}{\alpha})^\beta}) 1_{x>\nu}$$

$$f_X(x) = \frac{\beta}{\alpha} \left(\frac{x-\nu}{\alpha}\right)^{\beta-1} e^{-(\frac{x-\nu}{\alpha})^\beta}$$

$$\text{Let } Y = \left(\frac{x-\nu}{\alpha}\right)^\beta$$

$$F_Y(y) = P(X \leq \alpha y^{1/\beta} + \nu) = F_X(\alpha y^{1/\beta} + \nu) = 1 - e^{-y} \Rightarrow$$

$$Y \sim \text{exp}(1)$$

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

- $f_{(X,Y)}(x,y) \geq 0 \forall x,y \in \mathbb{R}^2$
- $\int_{\mathbb{R}} \int_{\mathbb{R}} f_{(X,Y)}(x,y) dx dy = 1$
- $\forall A \in \mathbb{R}^2, P((X,Y) \in A) = \int_A f_{(X,Y)}(x,y) dx dy$

Properties

$$\int_{\mathbb{R}} f_{(X,Y)}(x,y) dy = f_X(x)$$

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

joint cdf:

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

marginal cdf:

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

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)

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

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:

$$f_{\vec{Y}}(\vec{y}) = \frac{f_{\vec{X}}(\vec{x})}{|J_{\vec{g}}(\vec{x})|} = f_{\vec{X}}(\vec{x}) |J_{\vec{h}}(\vec{y})|, \vec{x} = \vec{h}(\vec{y})$$

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

Procedure for change of variable

1. identify the transformation
2. determine range of \vec{y} and hence that of \vec{Y}
3. identify the inverse
4. compute Jacobian
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

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

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

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

$$f_{X+Y}(n) P(X+Y=n) = \sum_{i \in \mathbb{Z}} P(X=n-i, Y=i)$$

$$= \sum_i f_X(n-i) f_Y(i) =: (f_X * f_Y)(n)$$

Two Independent $Poisson$

$$X_1 \sim Poisson(\lambda_1), X_2 \sim Poisson(\lambda_2)$$

$$\text{then } X_1 + X_2 \sim Poisson(\lambda_1 + \lambda_2)$$

Normal Distribution [Special]

Linear Transformation of Independent Z

Let X_1, \dots, X_n be i.i.d. $N(0,1)$

$$\text{pdf: } f_{\vec{X}}(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2} \vec{y}^T \Sigma^{-1} \vec{y}}$$

Covariance Matrix

Given a vector of RV $\vec{X} := (X_1, \dots, X_n)$ with mean $\vec{\mu} = (\mu_1, \dots, \mu_n)$, its covariance matrix is defined by:

$$\Sigma_{\vec{X}} := (Cov(X_i, X_j))_{1 \leq i, j \leq n} = E[(\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})^T]$$

Multivariable Normal Distribution

Since $\vec{Y} = A\vec{X}$, we have

$$E(\vec{Y}) = AE(\vec{X}) = \vec{0}$$

For $\vec{Y} = A\vec{X} + \vec{\mu}$, we have

$$E(\vec{Y}) = AE(\vec{X}) + \vec{\mu} = \vec{\mu}$$

$$\text{pdf: } f_{\vec{X}}(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2} (\vec{z} - \vec{\mu})^T \Sigma^{-1} (\vec{z} - \vec{\mu})}$$

Marginal Distribution of Multivariate Normal

Given i.i.d. normal distribution:

$$\vec{X} = (X_1, \dots, X_n) \sim (\vec{\mu}, \Sigma)$$

$$\vec{Z} = (Z_1, \dots, Z_n) \sim (\vec{0}, I)$$

Theorems:

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

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

Conditional Distribution of Multivariable Normal

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

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

Finding independent Normal

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

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

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

refer to lecture 17 for detailed working

Properties of Bivariate Normal

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

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

Law of Large Numbers

Moment Generating Functions

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

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

The joint mgf of (f_1, f_2, \dots, f_n)

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

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

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

Markov's Inequality

if X is a non-negative RV

$$P(X > a) \leq \frac{E(X)}{a}$$

Chebyshev's Inequality

Suppose $E(X) = \mu, Var(X) = \sigma^2 < \infty$

$$P(|X - \mu| > a) \leq \frac{Var(X)}{a^2}$$

Convergence in Probability

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

$$P(s : |X_n(s) - Y(s)| > \epsilon) \rightarrow 0, n \rightarrow \infty$$

Almost Sure Convergence

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

$$|X_n(s) - Y(s)| \rightarrow 0, n \rightarrow \infty$$

Convergence in Distribution

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

$$F_n(x) \rightarrow F(x), n \rightarrow \infty$$

Weak Law of Large Numbers

Let $(X_n)_{n \in \mathbb{N}}$ 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 $\epsilon > 0$

$$P(|S_n - \mu| > \epsilon) \rightarrow 0 \text{ as } n \rightarrow \infty$$

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 $(\xi_n)_{n \in \mathbb{N}}$ be i.i.d. RV with finite mean $E(\xi_1) = \mu$

Let $S_n = \frac{1}{n} \sum_{i=1}^n \xi_i$. Then almost surely,

$$|S_n - \mu| \rightarrow 0, n \rightarrow \infty$$

Central Limit Theorem

Let $(X_n)_{n \in \mathbb{N}}$ be i.i.d. RV with $E(X_1) = \mu, Var(X_1) = \sigma^2$

$$W_n := \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}} \rightarrow Z \sim N(0, \sigma^2), n \rightarrow \infty \text{ in distribution}$$