

01. COMBINATORIAL ANALYSIS

The Basic Principle of Counting

- basic principle of counting** → Suppose that two experiments are performed. If exp1 can result in any one of m possible outcomes and if, for each outcome of exp1, there are n possible outcomes of exp2, then together there are mn possible outcomes of the two experiments.
- generalized basic principle of counting** → If r experiments are performed such that the first one may result in any of n_1 possible outcomes and if for each of these n_1 possible outcomes, there are n_2 possible outcomes of the 2nd exp, and if ..., then there is a total of $n_1 \cdot n_2 \cdot \dots \cdot n_r$ possible outcomes of r experiments.

Permutations

factorials - $1! = 0! = 1$

N1 - if we know how to count the number of different ways that an event can occur, we will know the probability of the event.

N2 - there are $n!$ different arrangements for n objects.

N3 - there are $\frac{n!}{n_1! n_2! \dots n_r!}$ different arrangements of n objects, of which n_1 are alike, n_2 are alike, ..., n_r are alike.

Combinations

$$\binom{n}{r} = \frac{n!}{(n-r)! r!} = \binom{n-1}{r-1} + \binom{n-1}{r}, \quad 1 \leq r \leq n$$

N5 - The Binomial Theorem

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

Multinomial Coefficients

$$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \dots n_r!}$$

N6 - represents the number of possible divisions of n distinct objects into r distinct groups of respective sizes n_1, n_2, \dots, n_r , where $n_1 + n_2 + \dots + n_r = n$

N7 - The Multinomial Theorem: $(x_1 + x_2 + \dots + x_r)^n = \sum_{(n_1, \dots, n_r): n_1 + n_2 + \dots + n_r = n} \frac{n!}{n_1! n_2! \dots n_r!} x_1^{n_1} x_2^{n_2} \dots x_r^{n_r}$

Number of Integer Solutions of Equations

N8 - there are $\binom{n-1}{r-1}$ distinct *positive* integer-valued vectors (x_1, x_2, \dots, x_r) satisfying $x_1 + x_2 + \dots + x_r = n$, $x_i > 0$, $i = 1, 2, \dots, r$

N9 - there are $\binom{n+r-1}{r-1}$ distinct *non-negative* integer-valued vectors (x_1, x_2, \dots, x_r) satisfying $x_1 + x_2 + \dots + x_r = n$

Proof. let $y_k = x_k + 1 \Rightarrow y_1 + y_2 + \dots + y_r = n + r$

02. AXIOMS OF PROBABILITY

Sample Space and Events

- sample space** → The *set* of all outcomes of an experiment
- event** → Any *subset* of the sample space
- complement** of $E \rightarrow E^c$ is the event that contains all outcomes that are *not* in E .
- subset** → $E \subset F$ is all of the outcomes in E that are also in F .
 - $E \subset F \wedge F \subset E \Rightarrow E = F$

DeMorgan's Laws: $(\bigcup_{i=1}^n E_i)^c = \bigcap_{i=1}^n E_i^c$ and $(\bigcap_{i=1}^n E_i)^c = \bigcup_{i=1}^n E_i^c$

Axioms of Probability

definition 1: relative frequency

$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$. problems: (1) $\frac{n(E)}{n}$ may not converge when $n \rightarrow \infty$.

(2) $\frac{n(E)}{n}$ may not converge to the same value if the experiment is repeated.

Axioms (definition 2)

Consider an experiment with sample space S . For each event E of the sample space S , we assume that a number $P(E)$ is defined and satisfies the following 3 axioms:

- $0 \leq P(E) \leq 1$
- $P(S) = 1$
- For mutually exclusive events, $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$. same for finite case

mutually exclusive → events for which $E_i E_j = \emptyset$ when $i \neq j$

Simple Propositions

N1 - $P(\emptyset) = 0$

N6 - **probability function** \iff it satisfies the 3 axioms.

N8 - if $E \subset F$, then $P(E) \leq P(F)$

N10 - Inclusion-Exclusion identity where $n = 3$

$$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)$$

N11 - **Inclusion-Exclusion identity** -

$$P(E_1 \cup E_2 \cup \dots \cup E_n) = \sum_{i=1}^n P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \dots + (-1)^{r+1} \sum_{i_1 < i_2 < \dots < i_r} P(E_{i_1} E_{i_2} \dots E_{i_r}) + \dots + (-1)^{n+1} P(E_1 E_2 \dots E_n)$$

$$(i) \quad P(\bigcup_{i=1}^n E_i) \leq \sum_{i=1}^n P(E_i) \quad (\text{based on Inclusion-Exclusion identity})$$

$$(ii) \quad P(\bigcup_{i=1}^n E_i) \geq \sum_{i=1}^n P(E_i) - \sum_{j < i} P(E_i E_j)$$

$$(iii) \quad P(\bigcup_{i=1}^n E_i) \leq \sum_{i=1}^n P(E_i) - \sum_{j < i} P(E_i E_j) + \sum_{k < j < i} P(E_i E_j E_k)$$

(iv) and so on.

Sample Space having Equally Likely Outcomes

Consider an experiment with sample space $S = \{e_1, e_2, \dots, e_n\}$. Then $P(\{e_1\}) = P(\{e_2\}) = \dots = P(\{e_n\}) = \frac{1}{n}$ or $P(\{e_i\}) = \frac{1}{n}$.

N1 - for any event E , $P(E) = \frac{\# \text{ of outcomes in } E}{\# \text{ of outcomes in } S} = \frac{\# \text{ of outcomes in } E}{n}$

increasing sequence of events $\{E_n, n \geq 1\} \rightarrow E_1 \subset E_2 \subset \dots \subset E_n \subset \dots$

decreasing sequence of events $\{E_n, n \geq 1\} \rightarrow E_1 \supset E_2 \supset \dots \supset E_n \supset \dots$

increasing: $\lim_{n \rightarrow \infty} E_n = \bigcup_{i=1}^{\infty} E_i$ decreasing: $\lim_{n \rightarrow \infty} E_n = \bigcap_{i=1}^{\infty} E_i$

N2 - for both *increasing* and *decreasing* sequence, $\lim_{n \rightarrow \infty} P(E_n) = P(\lim_{n \rightarrow \infty} E_n)$

03. CONDITIONAL PROBABILITY AND INDEPENDENCE

Conditional Probability

if $P(F) > 0$, then $P(E|F) = \frac{P(E \cap F)}{P(F)}$

multiplication rule:

$$P(E_1 \dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2) \dots P(E_n|E_1E_2 \dots E_{n-1})$$

N3 - **axioms of probability** apply to conditional probability

- $0 \leq P(E|F) \leq 1$
- $P(S|F) = 1$ where S is the sample space
- If E_i ($i \in \mathbb{Z}_{\geq 1}$) are mutually exclusive, then $P(\bigcup_{i=1}^{\infty} E_i|F) = \sum_{i=1}^{\infty} P(E_i|F)$

N4 - If we define $Q(E) = P(E|F)$, then all previously proven results apply.

• $P(E_1 \cup E_2|F) = P(E_1|F) + P(E_2|F) - P(E_1 E_2|F)$

Total Probability & Bayes' Theorem

conditioning formula - $P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$

$$\begin{array}{c} P(F) \rightarrow F \begin{cases} \xrightarrow{P(E|F)} E \\ \xrightarrow{P(F^c)} E^c \end{cases} \\ P(F^c) \rightarrow F^c \begin{cases} \xrightarrow{P(E|F^c)} E \\ \xrightarrow{P(F)} E^c \end{cases} \end{array} \quad P(F|E) = \frac{P(EF)}{P(E)} = \frac{P(F) \cdot P(E|F)}{P(E)} \quad P(F^c|E) = \frac{P(EF^c)}{P(E)} = \frac{P(F^c) \cdot P(E|F^c)}{P(E)}$$

Total Probability

theorem of total probability - Suppose F_1, F_2, \dots, F_n are mutually exclusive events such that $\bigcup_{i=1}^n F_i = S$, then $P(E) = \sum_{i=1}^n P(EF_i) = \sum_{i=1}^n P(F_i)P(E|F_i)$

Bayes Theorem

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(F_j)P(E|F_j)}{\sum_{i=1}^n P(F_i)P(E|F_i)}$$

application of bayes' theorem

$$P(B_1 | A) = \frac{P(A|B_1) \cdot P(B_1)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)}$$

Let A be the event that the person test positive for a disease.

B_1 : the person has the disease. B_2 : the person does not have the disease.

true positives: $P(B_1 A)$	false negatives: $P(\bar{A} B_1)$
false positives: $P(A B_2)$	true negatives: $P(\bar{A} B_2)$

Independent Events

N1 - E and F are independent $\iff P(EF) = P(E) \cdot P(F)$

N2 - E and F are independent $\iff P(E|F) = P(E)$

N3 - E and F are independent $\iff E$ and F^c are independent.

N4 - if E, F, G are independent, then E will be independent of any event formed from F and G . (e.g. $F \cup G$)

N6 - (E and F are indep) \wedge (E and G are indep) $\nRightarrow E$ and FG are independent

N7 - For independent trials with probability p of success, probability of m successes before n failures, for $m, n \geq 1$,

method 1

method 2

$$\begin{array}{c} p \rightarrow S \begin{cases} \xrightarrow{P_{n-1}, m} \text{A win} \\ \xrightarrow{1-p} \text{B win} \end{cases} \\ 1-p \rightarrow F \begin{cases} \xrightarrow{P_{n,m-1}} \text{A win} \\ \xrightarrow{p} \text{B win} \end{cases} \end{array} \quad P_{n,m} = \sum_{k=n}^{m+n-1} \binom{m+n-1}{k} p^k (1-p)^{m+n-1-k} = P(\geq n \text{ successes in } m+n-1 \text{ trials})$$

04. RANDOM VARIABLES

- random variable** → a real-valued function defined on the sample space

Types of Random Variables

r.v.	-	$E(X)$
binomial	$X = \#$ of successes in n trials w/ replacement	np
negative binomial	$X = \#$ of trials until k successes	k/p
geometric	$X = \#$ of trials until a success	$1/p$
hypergeometric	$X = \#$ of successes in n trials, no replacement	rn/N

- X is a **Bernoulli r.v.** with parameter p if →

$$p(x) = \begin{cases} p, & x = 1, \text{ ('success')} \\ 1-p, & x = 0 \text{ ('failure')} \end{cases}$$

- Y is a **Binomial r.v.** with parameters n and $p \rightarrow Y = X_1 + X_2 + \dots + X_n$ where X_1, X_2, \dots, X_n are independent Bernoulli r.v.'s with parameter p .

- $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $P(k \text{ successes from } n \text{ independent trials each with probability } p \text{ of success})$
- $E(Y) = np, \quad \text{Var}(Y) = np(1-p)$

- Negative Binomial** → $X =$ number of trials until k successes are obtained
 - e.g. number of balls drawn (with replacement) until k red balls are obtained

- Geometric** → $X =$ number of trials until a success is obtained

- $P(X = k) = (1-p)^{k-1} \cdot p$ where k is the number of trials needed
- e.g. number of balls drawn (with replacement) until 1 red ball is obtained

- Hypergeometric** → $X =$ number of trials until success, *without replacement*

- $P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}, k = 0, 1, \dots, n$ (for m red balls of N balls)

- e.g. number of red balls out of n balls drawn without replacement

Properties

N1 - if $X \sim \text{Binomial}(n, p)$, and $Y \sim \text{Binomial}(n - 1, p)$, then $E(X^k) = np \cdot E[(Y + 1)^{k-1}]$

N2 - if $X \sim \text{Binomial}(n, p)$, then for $k \in \mathbb{Z}^+$, $P(X = k) = \frac{(n-k+1)p}{k(1-p)} \cdot P(X = k - 1)$

Coupon Collector Problem

Q. Suppose there are N distinct types of coupons. If T denotes the number of coupons needed to be collected for a complete set, what is $P(T = n)$?

A. $P(T > n - 1) = P(T \geq n) = P(T = n) + P(T > n)$
 $\Rightarrow P(T = n) = P(T > n - 1) - P(T > n)$

Let $A_j = \{\text{no type } j \text{ coupon is contained among the first } n\}$

$P(T > n) = P(\bigcup_{j=1}^N A_j)$
 $P(T > n) = \sum_j P(A_j)$ - coupon j is not among the first n collected
 $-\sum_{j_1 < j_2} P(A_{j_1} A_{j_2})$ - coupon j_1 and j_2 are not the first n
 $+\dots + (-1)^{N+1} P(A_1 A_2 \dots A_N)$ by inclusion-exclusion identity
 $P(A_{j_1} A_{j_2} \dots A_{j_k}) = (\frac{N-k}{N})^n$

Hence $P(T > n) = \sum_{i=1}^{N-1} \binom{N}{i} (\frac{N-i}{N})^n (-1)^{i+1}$

Probability Mass Function

probability mass function, pmf of X (discrete) $\rightarrow p(a) = P(X = a)$
• if X assumes one of the values x_1, x_2, \dots , then $\sum_{i=1}^{\infty} p(x_i) = 1$

Cumulative Distribution Function

• **cumulative distribution function (cdf)** of a r.v. $X \rightarrow$ the function F defined by $F(x) = P(X \leq x)$, $-\infty < x < \infty$
• $F(x)$ is defined on the entire real line. (aka *distribution function*)

pmf, $\frac{a}{p(a)} \mid \frac{1}{2} \quad \frac{2}{\frac{1}{4}} \quad \frac{4}{\frac{1}{4}}$ cdf, $F(a) = \begin{cases} 0, & a < 1 \\ 1/2, & 1 \leq a < 2 \\ 3/4, & 2 \leq a < 4 \\ 1, & a \geq 4 \end{cases}$
 $F(a) = \sum p(x)$ for all $x \leq a$

Expected Value

• aka population mean/sample mean, μ
discrete: $E(X) = \sum x \cdot p(x)$
continuous: $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$
 $E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

N1 - if a and b are constants, then $E(aX + b) = aE(X) + b$

N3 - for a non-negative r.v. Y , $E(Y) = \int_0^{\infty} P(Y > y) dy$

Proof. $\int_0^{\infty} P(Y > y) dy = \int_0^{\infty} \int_y^{\infty} f_Y(x) dx dy$ (because $f(x) = \frac{d}{dx} F(x)$)
 $= \int_0^{\infty} x f_Y(x) dx = E(Y)$

• I is an indicator variable for event A if $I = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A^c \text{ occurs} \end{cases}$. then $E(I) = P(A)$.

finding expectation of f(x)

• method 1, using pmf of Y : let $Y = f(X)$. Find corresponding X for each Y .
• method 2, using pmf of X : $E[f(x)] = \sum_x f(x)p(x)$

Variance

If X is a r.v. with mean $\mu = E[X]$, then the **variance** of X is defined by

$$\begin{aligned} Var(X) &= E[(X - \mu)^2] \\ &= E(x^2) - [E(x)]^2 \end{aligned}$$

• $Var(aX + b) = a^2 Var(x)$
• $Var(X) = \sum_{x_i} (x_i - \mu)^2 \cdot p(x_i)$ (deviation \cdot weight)

Poisson Random Variable

a r.v. X is said to be a **Poisson r.v.** with parameter λ if for some $\lambda > 0$,

$$P(X = i) = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$$
$$E(X) = \lambda, \quad Var(X) = \lambda$$

• $\sum_{i=0}^{\infty} P(X = i) = 1$
• **Poisson Approximation of Binomial** - if $X \sim \text{Binomial}(n, p)$, where n is large and p is small, then $X \sim \text{Poisson}(\lambda)$ where $\lambda = np$.
• \checkmark weak dependence is ok
• **2 ways** to look at the Poisson distribution
1. an approximation to the binomial distribution with large n and small p
2. counting the number of events that occur at *random* at certain points in time

Poisson distribution as random events

Let $N(t)$ be the number of events that occur in time interval $[0, t]$.

N1 - If the 3 assumptions are true, then $N(t) \sim \text{Poisson}(\lambda t)$.

N2 - If λ is the *rate of occurrences* of events per unit time, then the number of occurrences in an interval of length t has a Poisson distribution with mean λt .

$$P(N(t) = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!}, \text{ for } k \in \mathbb{Z}_{\geq 0}$$

o(h) notation

$o(h)$ stands for any function $f(h)$ such that $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$

• $o(h) + o(h) = o(h)$
• $\frac{\lambda t}{n} + o(\frac{t}{n}) \doteq \frac{\lambda t}{n}$ for large n

Expected Value of sum of r.v.

For a r.v. X , let $X(s)$ denote the value of X when $s \in S$

N1 - $E(x) = \sum_i x_i P(X = x_i) = \sum_{s \in S} X(s) p(s)$ where $S_i = \{s : X(s) = x_i\}$

N2 - $E(\sum_{i=1}^n) = \sum_{i=1}^n E(X_i)$ for r.v. X_1, X_2, \dots, X_n

e.g. distribution of time to next event

Q. suppose an accident happens at a rate of 5 per day. Find the distribution of time, starting from now, until the next accident.

A. Let X = time (in days) until the next accident.
Let V = be the number of accidents during time period $[0, t]$.

$V \sim \text{Poisson}(5t) \Rightarrow P(V = k) = \frac{e^{-5t} \cdot (5t)^k}{k!}$
 $P(X > t) = P(\text{no accidents happen during } [0, t]) = P(V = 0) = e^{-5t}$
 $P(X \leq t) = 1 - e^{-5t}$

example: finding pdf

Q - Find the pdf of $(b - a)X + a$ where a, b are constants, $b > a$. The pdf of X is given by $f(x) = \begin{cases} 1, & 0 \leq X \leq 1 \\ 0, & \text{otherwise} \end{cases}$.

A. Let $Y = (b - a)X + a$.
cdf, $F_Y(y) = P(Y \leq y) = P((b - a)X + a \leq y) = P(X \leq \frac{y-a}{b-a})$
 $F_Y(y) = \int_0^{\frac{y-a}{b-a}} 1 dx = \frac{y-a}{b-a}, \quad a < y < b$
 $f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{1}{b-a}, & a < y < b \\ 0, & \text{otherwise} \end{cases}$

05. CONTINUOUS RANDOM VARIABLES

X is a **continuous r.v.** \rightarrow if there exists a nonnegative function f defined for all real $x \in (-\infty, \infty)$, such that $P(X \in B) = \int_B f(x) dx$

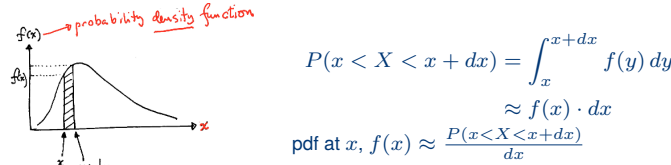
N1 - $P(X \in (-\infty, \infty)) = \int_{-\infty}^{\infty} f(x) dx = 1$

N2 - $P(a \leq X \leq b) = \int_a^b f(x) dx$

N3 - $P(X = a) = \int_a^a f(x) dx = 0$

N4 - $P(X < a) = P(X \leq a) = \int_{-\infty}^a f(x) dx$

N5 - interpretation of **probability density function**



N6 - if X is a continuous r.v. with pdf $f(x)$ and cdf $F(x)$, then $f(x) = \frac{d}{dx} F(x)$. (*Fundamental Theorem of Calculus*)

N7 - median of X , x occurs where $F(x) = \frac{1}{2}$

Generating a Uniform r.v.

if X is a continuous r.v. with cdf $F(x)$, then

• N8 - $F(X) = U \sim \text{uniform}(0, 1)$.

Proof. let $Y = F(X)$. then cdf of Y , $F_Y(y) = P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$.

• N9 - $X = F^{-1}(U) \sim$ cdf $F(x)$.
• generating a r.v. from a uniform(0, 1) r.v. and a r.v. with cdf $F(x)$.

Uniform Random Variable

X is a **uniform r.v.** on the interval (α, β) , $X \sim \text{Uniform}(\alpha, \beta)$

if its pdf is given by

$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \alpha < x < \beta \\ 0, & \text{otherwise} \end{cases}$
 $E(X) = \frac{\alpha + \beta}{2}, \quad Var(X) = \frac{(\beta - \alpha)^2}{12}$

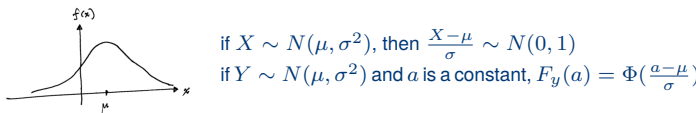
if $X \sim \text{Uniform}(\alpha, \beta)$, then $\frac{x - \alpha}{\beta - \alpha} \sim \text{Uniform}(0, 1)$

Normal Random Variable

X is a **normal r.v.** with parameters μ and σ^2 , $X \sim N(\mu, \sigma^2)$

if the pdf of X is given by

$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x - \mu}{\sigma})^2}, \quad -\infty < x < \infty$
 $E(x) = \mu, \quad Var(X) = \sigma^2$



standard normal distribution $\rightarrow X \sim N(0, 1)$

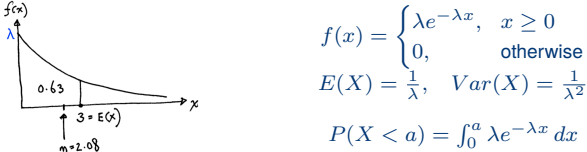
• $F(x) = P(X \leq x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy = \Phi(x)$

Normal Approximation to the Binomial Distribution

if $S_n \sim \text{Binomial}(n, p)$, then $\frac{S_n - np}{\sqrt{np(1-p)}} \sim N(0, 1)$ for large n .
 $\mu = np, \quad \sigma^2 = np(1 - p)$

Exponential Random Variable

a continuous r.v. X is a **exponential r.v.**, $X \sim Exponential(\lambda)$ or $Exp(\lambda)$ if for some $\lambda > 0$, its pdf is given by



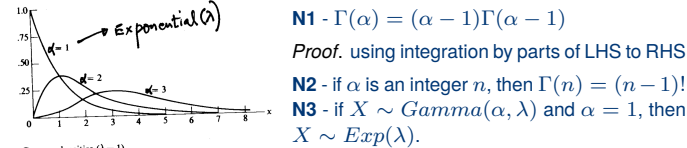
- an exponential r.v. is **memoryless**.
 - a non-negative r.v. is **memoryless** \rightarrow if $P(X > s + t | X > t) = P(X > s)$ for all $s, t > 0$.

Gamma Distribution

a r.v. X has a **gamma distribution**, $X \sim Gamma(\alpha, \lambda)$ with parameters (α, λ) , $\lambda > 0$ and $\alpha > 0$ if its pdf is given by

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$
$$E(X) = \frac{\alpha}{\lambda} \quad Var(X) = \frac{\alpha}{\lambda^2}$$

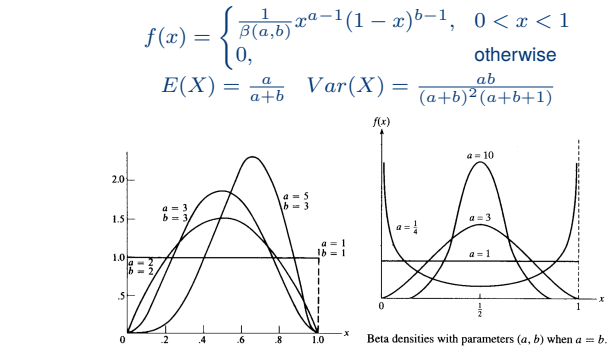
where the gamma function $\Gamma(\alpha)$ is defined as $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy$.



- N4** - for events occurring randomly in time following the 3 assumptions of poisson distribution, the **amount of time elapsed** until a total of n events has occurred is a gamma r.v. with parameters (n, λ) .
- time at which the n -th event occurs, $T_n \sim Gamma(n, \lambda)$
- number of events in time period $[0, t]$, $N(t) \sim Poisson(\lambda t)$
- N5** - $Gamma(\alpha = \frac{n}{2}, \lambda = \frac{1}{2}) = \chi_n^2$ (chi-square distribution to n degrees of freedom)

Beta Distribution

a r.v. X is said to have a **beta distribution**, $X \sim Beta(a, b)$



- N1** - $\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$
- N2** - $\beta(a=1, b=1) = Uniform(0, 1)$
- N3** - $\beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

Cauchy Distribution

a r.v. X has a cauchy distribution, $X \sim Cauchy(\theta)$ with parameter θ , $-\infty < \theta < \infty$ if its density is given by

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}, -\infty < x < \infty$$

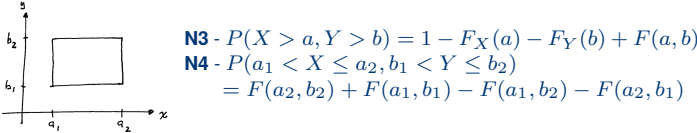
06. JOINTLY DISTRIBUTED RANDOM VARIABLES

Joint Distribution Function

the **joint cumulative distribution function** of the pair of r.v. X and Y is \rightarrow
 $F(x, y) = P(X \leq x, Y \leq y), -\infty < x < \infty, -\infty < y < \infty$

N1 - **marginal cdf of X** , $F_X(x) = \lim_{y \rightarrow \infty} F(x, y)$.

N2 - **marginal cdf of Y** , $F_Y(y) = \lim_{x \rightarrow \infty} F(x, y)$.



Joint Probability Mass Function

if X and Y are both discrete r.v., then their **joint pmf** is defined by
 $p(i, j) = P(X = i, Y = j)$

N1 - **marginal pmf of X** , $P(X = i) = \sum_j P(X = i, Y = j)$

N2 - **marginal pmf of Y** , $P(Y = i) = \sum_i P(X = i, Y = j)$

Joint Probability Density Function

the r.v. X and Y are said to be **jointly continuous** if there is a function $f(x, y)$ called the **joint pdf**, such that for any two-dimensional set C ,

$$P[(X, Y) \in C] = \iint_C f(x, y) dx dy$$

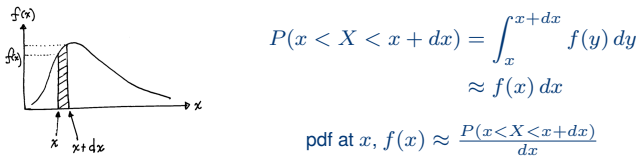
= volume under the surface over the region C .

N1 - if $C = \{(x, y) : x \in A, y \in B\}$, then
 $P(X \in A, Y \in B) = \int_A \int_B f(x, y) dx dy$

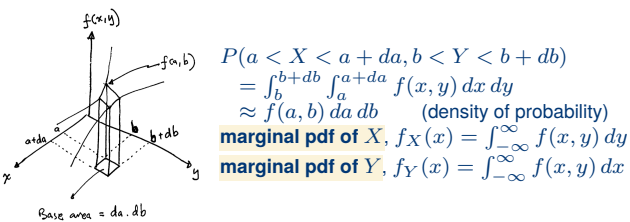
N2 - $F(a, b) = P(X \in (-\infty, a], Y \in (-\infty, b]) = \int_{-\infty}^b \int_{-\infty}^a f(x, y) dx dy$

N3 - $f(a, b) = \frac{\delta^2}{\delta a \delta b} F(a, b)$

interpretation of pdf



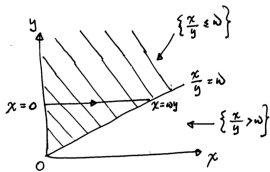
interpretation of joint pdf



how to do a double integral

example - given the joint pdf of X and Y , find the pdf of r.v. X/Y .

ans. set dummy variable $W = X/Y$, then
 $F_W(w) = P(W \leq w) = P(\frac{X}{Y} \leq w)$
 $P(\frac{X}{Y} \leq w) = \int_0^\infty \int_0^{wy} e^{-x-y} dx dy$



Independent Random Variables

N1 - X, Y are **independent** $\rightarrow P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$

N2 - X and Y are **independent** $\rightarrow P(X \leq a, Y \leq b) = P(X \leq a) \cdot P(Y \leq b)$ or $F(a, b) = F_X(a) \cdot F_Y(b) \Rightarrow$ joint cdf is the product of the marginal cdfs

N3 - **discrete case**: discrete r.v. X and Y are **independent** $\iff P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ for all x, y .

N4 - **continuous case**: jointly continuous r.v. X and Y are **independent** $\iff f(x, y) = f_X(x) \cdot f_Y(y)$ for all x, y .
 $-\infty < x < \infty, -\infty < y < \infty$

N5 - independence is a **symmetric** relation $\rightarrow X$ indep of $Y \iff Y$ indep of X
if X and Y are independent, then for any functions h and g ,
 $E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$

Sum of Independent Random Variables

N1 - for independent, continuous r.v. X and Y having pdf f_X and f_Y ,
 $F_{X+Y}(a) = \int_{-\infty}^\infty F_X(a-y) f_Y(y) dy$
 $f_{X+Y}(a) = \int_{-\infty}^\infty f_X(a-y) f_Y(y) dy$

Distribution of Sums of Independent r.v.

for $i = 1, 2, \dots, n$,

- $X_i \sim Gamma(t_i, \lambda) \Rightarrow \sum_{i=1}^n X_i \sim Gamma(\sum_{i=1}^n t_i, \lambda)$
- $X_i \sim Exp(\lambda) \Rightarrow \sum_{i=1}^n X_i \sim Gamma(n, \lambda)$
- $Z_i \sim N(0, 1) \Rightarrow \sum_{i=1}^n Z_i^2 \sim \chi_n^2 = Gamma(\frac{n}{2}, \frac{1}{2})$
- $X_i \sim N(\mu_i, \sigma_i^2) \Rightarrow \sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$
- $X \sim Poisson(\lambda_1), Y \sim Poisson(\lambda_2) \Rightarrow X + Y \sim Poisson(\lambda_1 + \lambda_2)$
- $X \sim Binom(n, p), Y \sim Binom(m, p) \Rightarrow X + Y \sim Binom(n + m, p)$

Conditional Distribution (discrete)

for discrete r.v. X and Y , the **conditional pmf** of X given that $Y = y$ is
 $P_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{p(x, y)}{p_Y(y)}$

for discrete r.v. X and Y , the **conditional pdf** of X given that $Y = y$ is
 $F_{X|Y}(x|y) = P(X \leq x|Y = y) = \sum_{a \leq x} \frac{P(X=a, Y=y)}{P(Y=y)} = \sum_{a \leq x} P_{X|Y}(a|y)$

N0 - equivalent notation:

- $P_{X|Y}(x|y) = P(X = x|Y = y)$
- $P_X(x) = P(X = x)$

N1 - if X is independent of Y , then $P_{X|Y}(x|y) = P_X(x)$

Conditional Distribution (continuous)

for X and Y with joint pdf $f(x, y)$, the **conditional pdf** of X given that $Y = y$ is
 $f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$ for all y s.t. $f_Y(y) > 0$

$$f_{X|Y}(a|y) = P(X \leq a|Y = y) = \int_{-\infty}^a f_{X|Y}(x|y) dx$$

N1 - for any set A , $P(X \in A|Y = y) = \int_A f_{X|Y}(x|y) dx$

N2 - if X is independent of Y , then $f_{X|Y}(x|y) = f_X(x)$.

! "find the marginal/conditional pdf of Y " \Rightarrow must include the **range** too!!

Joint Probability Distribution of Functions of r.v.

Let X_1 and X_2 be jointly continuous r.v. with joint pdf $f_{x_1,x_2}(x_1,x_2)$. Suppose $Y_1 = g_1(X_1,X_2)$ and $Y_2 = g_2(X_1,X_2)$ satisfy

- 1. the equations $y_1 = g_1(X_1,X_2)$ and $y_2 = g_2(X_1,X_2)$ can be *uniquely* solved for x_1,x_2 in terms of y_1 and y_2
- 2. $g_1(x_1,x_2)$ and $g_2(x_1,x_2)$ have continuous partial derivatives at all points

(x_1,x_2) such that $J(x_1,x_2) = \begin{vmatrix} \frac{\delta g_1}{\delta x_1} & \frac{\delta g_1}{\delta x_2} \\ \frac{\delta g_2}{\delta x_1} & \frac{\delta g_2}{\delta x_2} \end{vmatrix} = \frac{\delta g_1}{\delta x_1} \cdot \frac{\delta g_2}{\delta x_2} - \frac{\delta g_2}{\delta x_1} \cdot \frac{\delta g_1}{\delta x_2} \neq 0$

then

$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(x_1,x_2) \frac{1}{|J(x_1,x_2)|}$
where $x_1 = h_1(y_1,y_2), x_2 = h_2(y_1,y_2)$

07. PROPERTIES OF EXPECTATION

recap:

- for a **discrete** r.v. $X, E(X) = \sum_x x \cdot p(x) = \sum_x \cdot P(X = x)$
- for a **continuous** r.v. $X, E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx$
- for a **non-negative integer-valued** r.v. $Y, E(Y) = \sum_{i=1}^{\infty} P(Y \geq i)$
- for a **non-negative** r.v. $Y, E(Y) = \int_{-\infty}^{\infty} P(Y > y) dy$

Expectations of Sums of Random Variables

for X and Y with joint pmf $p(x,y)$ and joint pdf $f(x,y)$,

$E[g(x,y)] = \sum_y \sum_x g(x,y)p(x,y)$

$E[g(x,y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y) dx dy$

- N2** - if $P(a \leq X \leq b) = 1$, then $a \leq E(X) \leq b$
- N4** - for r.v.s X and Y , if $X \geq Y$, then $E(X) \geq E(Y)$
- N5** - let X_1, \dots, X_n be independent and identically distributed r.v.s having distribution $P(X_i \leq x) = F(x)$ and expected value $E(X_i) = \mu$.

if $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$, then $E(\bar{X}) = \mu$

N6 - \bar{X} is the **sample mean**. \Rightarrow sample mean = population mean

! trick: express a r.v. as a sum of r.v. with easier to find expectation

examples

- hypergeometric with r red balls out of N balls with n trials
 - indicator r.v. = 1 if the i th ball selected is red
 - $P(Y_i = 1) = \frac{r}{N} \Rightarrow E(Y_i) = \frac{r}{N} \Rightarrow E(X) = \sum_{i=1}^n Y_i = n \frac{r}{N}$
- coupon collector problem:
 - let X = number of coupons collected for a complete set
 - let X_i = additional number to be collected to obtain distinct type after i distinct types have been collected. $X_i \sim Geometric(p = \frac{N-i}{N})$
 - $E(X) = \sum_{i=1}^{N-1} E(X_i) = 1 + \frac{1}{\frac{N-1}{N}} + \frac{1}{\frac{N-2}{N}} + \dots + \frac{1}{\frac{1}{N}}$
 $= N(\frac{1}{N} + \frac{1}{N-1} + \dots + 1)$

Covariance, Variance of Sums and Correlations

covariance \rightarrow measure of *linear relationship*

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

N1 - X and Y are independent $\Rightarrow Cov(X,Y) = 0$

N2 - $Cov(X,Y) = 0 \nRightarrow X$ and Y are independent. *Proof.* let $E(X) = 0, E(XY) = 0 \Rightarrow Cov(X,Y) = 0$, but not independent e.g. non-linear relationship

Covariance properties

- 1. $Cov(X,Y) = Cov(Y,X)$
- 2. $Cov(X,X) = Var(X)$
- 3. $Cov(aX,Y) = aCov(X,Y)$
- 4. $Cov(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j) = \sum_{i=1}^n \sum_{j=1}^m Cov(X_i,Y_j)$

for variance:

N1 - $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + 2 \sum_{i < j} Cov(X_i,X_j)$

N2 - if X_1, \dots, X_n are *pairwise independent*, $Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i)$

N3 - for n independent and identically distributed r.v. with variance σ^2 ,

$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

$Var(\bar{X}) = \frac{\sigma^2}{n} \quad E(S^2) = \sigma^2$

$\Rightarrow S^2$ is an *unbiased estimator* for σ^2 .

Correlation

correlation of two r.v. X and $Y, \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$

N1 - $-1 \leq \rho(X,Y) \leq 1$ where -1 and 1 denote a perfect negative and positive linear relationship respectively.

N2 - $\rho(X,Y) = 0 \Rightarrow$ no *linear* relationship - uncorrelated

N3 - $\rho(X,Y) = 1 \Rightarrow Y = aX + b, a = \frac{\delta y}{\delta x} > 0$

N4 - for independent events A, B with indicator r.v. $I_A, I_B: Cov(I_A, I_B) = 0$.

N5 - deviation is not correlated with the sample mean. For independent & identically distributed r.v. X_1, X_2, \dots, X_n with variance σ^2 , then $Cov(X_i - \bar{X}, \bar{X}) = 0$.

Conditional Expectation

the **conditional expectation** of X given that $Y = y, \forall y$ s.t. $P_Y(y) > 0$, is:

$E[X|Y = y] = \sum_x x \cdot P(X = x|Y = y) = \sum_x x \cdot p_{X|Y}(x|y)$

$E(X|Y = y) = \int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) dx = \int_{-\infty}^{\infty} x \cdot \frac{f(x,y)}{f_Y(y)} dx$

! note the range for $f_{X|Y}(x|y)$

N1 - If $X, Y \sim Geometric(p)$, then $P(X = i|X + Y = n) = \frac{1}{n-1}$, a uniform distribution.

N2 - $E(X|X + Y = n) = \sum_{i=1}^{n-1} i \cdot P(X = i|X + Y = n) = \frac{n}{2}$
discrete case: $E[g(x)|Y = y] = \sum g(x)P_{X|Y}(x|y)$

continuous case: $E[g(x)|Y = y] = \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y) dx$
then $E(X) = E_{w.r.t. y}(E_{w.r.t. X|Y=y}(X|Y))$

Deriving Expectation

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

discrete case: $E(X) = \sum_y E(X|Y = y)P(Y = y)$

continuous case: $E(X) = \int_{-\infty}^{\infty} E(X|Y = y)f_Y(y) dy$

N3 - 3 methods for finding $E(X)$ given $f(x,y)$

- 1. using $E(g(x,y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y) dx dy \Rightarrow$ let $g(x,y) = x$
- 2. using $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$
- 3. using $E(X) = \int_{-\infty}^{\infty} E(X|Y = y)f_Y(y) dy$

N4 - $E(\sum_{i=1}^N X_i) = E_N(E(\sum_{i=1}^N X_i|N)) = \sum_{n=0}^{\infty} E(\sum_{i=1}^N X_i|N = n) \cdot P(N = n)$

Computing Probabilities by Conditioning

discrete: $P(E) = \sum_y P(E|Y = y)P(Y = y)$

continuous: $P(E) = \int_{-\infty}^{\infty} P(E|Y = y)f_Y(y) dy$

Proof. X is an indicator r.v.; $E(X|Y = y) = P(X = 1|Y = y) = P(E|Y = y)$

N5 - $P(X < Y) = \int P(X < Y|Y = y) \cdot f_Y(y)$

Conditional Variance

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

N6 - $Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$

N7 - $E(f(Y)) = E(f(Y)|Y = t) = E(f(t)|Y = t)$
 $= E(f(t))$ if $N(t)$ and Y are independent

Moment Generating Functions

moment generating function $M(t)$ of the r.v. $X \rightarrow$

$M(t) = E(e^{tX})$ for all real values of t

- if X is *discrete* with pmf $p(x), M(t) = \sum_x e^{tx} \cdot p(x)$
- if X is *continuous* with pdf $f(x), M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$

$M(t)$ is called the **mgf** because *all moments* of X can be obtained by successively differentiating $M(t)$ and then evaluating the result at $t = 0$.

- the n^{th} moment of of X is given as $E(X^n) = \sum_x x^n \cdot p(x)$
 - $M'(0) = E(X), M''(0) = E(X^2), M^n(0) = E(X^n), n \geq 1$
 - $M'(t) = E(X^n e^{tX}), n \geq 1$

if X and Y are independent and have mgf's $M_X(t)$ and $M_Y(t)$ respectively,

N10 - the mgf of $X + Y$ is $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$

N11 - if $M_X(t)$ exists and is finite in some region about $t = 0$, then the distribution of X is **uniquely** determined. $M_X(t) = M_Y(t) \iff X = Y$

Common mgf's

- $X \sim Normal(0, 1), M(t) = e^{e^2/2}$
- $X \sim Binomial(n, p), M(t) = (pe^t + (1 - p))^n$
- $X \sim Poisson(\lambda), M(t) \exp[\lambda(e^t - 1)]$
- $X \sim Exp(\lambda), M(t) = \frac{\lambda}{\lambda - t}$

08. LIMIT THEOREMS

Markov's Inequality \rightarrow if X is a non-negative r.v., $\forall a > 0, P(X \geq a) \leq \frac{E(x)}{a}$.

Chebyshev's inequality \rightarrow if X is an r.v. with finite mean μ and variance σ^2 , then

for any value of $k > 0, P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$.

N1 - if $Var(X) = 0$, then $P(X = E[X]) = 1$

weak law of large numbers \rightarrow let X_1, X_2, \dots be a sequence of independent and identically distributed r.v.s, each with finite mean $E[X_i] = \mu$. Then, for any $\epsilon > 0, P\{\frac{X_1 + \dots + X_n}{n} - \mu \geq \epsilon\} \rightarrow 0$ as $n \rightarrow \infty$

central limit theorem \rightarrow let X_1, X_2, \dots be a sequence of independent and identically distributed r.v.s each having mean μ and variance σ^2 . Then the distribution of $\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$ tends to the standard normal as $n \rightarrow \infty$.

- aka: $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightarrow z \sim N(0, 1)$
- for $-\infty < a < \infty, as n \rightarrow \infty,$
 $P(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq a) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx = F(a)$ - cdf of $N(0, 1)$

N2 - Let Z_1, Z_2, \dots be a sequence of r.v.s with distribution functions F_{Z_n} and moment generating functions $M_{Z_n}, n \geq 1$. Let Z be a r.v. with distribution function F_Z and mgf M_Z .

If $M_{Z_n}(t) \rightarrow M_Z(t)$ for all t , then $F_{Z_n}(t) \rightarrow F_Z(t)$ for all t at which $F_Z(t)$ is continuous.

strong law of large numbers \rightarrow let X_1, X_2, \dots be a sequence of independent and identically distribution r.v.s, each having finite mean $\mu = E[X_i]$.

Then, with probability 1, $\frac{X_1 + \dots + X_n}{n} \rightarrow \mu$ as $n \rightarrow \infty$

approximations - $\lim_{n \rightarrow \infty} (1 - \frac{\lambda}{n})^n = e^{-\lambda}$