

01. COMBINATORIAL ANALYSIS

tricky - E18, E20-22, E23, E26

The Basic Principle of Counting

- combinatorial analysis** → the mathematical theory 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, 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

N4 - $\binom{n}{r} = \frac{n!}{(n-r)! r!}$ represents the number of different groups of size r that could be selected from a set of n objects when the order of selection is not considered relevant.

N4b - $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$, $1 \leq r \leq n$

Proof. If object 1 is chosen $\Rightarrow \binom{n-1}{r-1}$ ways of choosing the remaining objects.

If object 1 is not chosen $\Rightarrow \binom{n-1}{r}$ ways of choosing the remaining objects.

N5 - The Binomial Theorem - $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$

Proof. by mathematical induction: $n = 1$ is true; expand; sub dummy variable; combine using N4b; combine back to final term

Multinomial Coefficients

N6 - $\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \dots n_r!}$ 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 $n_1 + n_2 + \dots + n_r = n$

Proof. using basic counting principle,

$$\begin{aligned} &= \binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots \binom{n-n_1-n_2-\dots-n_{r-1}}{n_r} \\ &= \frac{n!}{(n-n_1)! n_1!} \times \frac{(n-n_1)!}{(n-n_1-n_2)! n_2!} \times \dots \times \frac{(n-n_1-n_2-\dots-n_{r-1})!}{0! n_r!} \\ &= \frac{n!}{n_1! n_2! \dots n_r!} \end{aligned}$$

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 $x_1 + x_2 + \dots + x_r = n$, $x_i > 0$, $i = 1, 2, \dots, r$

! cannot be directly applied to N8 as 0 value is not included

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 (where outcomes are not predictable with certainty)
- event** → Any *subset* of the sample space
- union** of events E and $F \rightarrow E \cup F$ is the event that contains all outcomes that are either in E or F (or both).
- intersection** of events E and $F \rightarrow E \cap F$ or EF is the event that contains all outcomes that are both in E and in F .
- complement** of $E \rightarrow E^c$ is the event that contains all outcomes that are *not* in E .
- subset** → $E \subset F$ if all of the outcomes in E that are also in F .
 - $E \subset F \wedge F \subset E \Rightarrow E = F$

DeMorgan's Laws

$$\left(\bigcup_{i=1}^n E_i \right)^c = \bigcap_{i=1}^n E_i^c$$

Proof. to show $LHS \subset RHS$: let $x \in \left(\bigcup_{i=1}^n E_i \right)^c$
 $\Rightarrow x \notin \bigcup_{i=1}^n E_i \Rightarrow x \notin E_1$ and $x \notin E_2 \dots$ and $x \notin E_n$
 $\Rightarrow x \in E_1^c$ and $x \in E_2^c \dots$ and $x \in E_n^c$
 $\Rightarrow x \in \bigcap_{i=1}^n E_i^c$
 to show $RHS \subset LHS$: let $x \in \bigcap_{i=1}^n E_i^c$

$$\left(\bigcap_{i=1}^n E_i \right)^c = \bigcup_{i=1}^n E_i^c$$

Proof. using the first law of DeMorgan, negate LHS to get RHS

Axioms of Probability

definition 1: relative frequency

$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$

problems with this definition:

- $\frac{n(E)}{n}$ may not converge when $n \rightarrow \infty$
- $\frac{n(E)}{n}$ may not converge to the same value if the experiment is repeated

definition 2: Axioms

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 any sequence of mutually exclusive events E_1, E_2, \dots (i.e., events for which $E_i E_j = \emptyset$ when $i \neq j$),

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

$P(E)$ is the probability of event E .

Simple Propositions

N1 - $P(\emptyset) = 0$

N2 - $P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$ (aka axiom 3 for a finite n)

N3 - strong law of large numbers - if an experiment is repeated over and over again, then with probability 1, the proportion of time during which any specific event E occurs will be equal to $P(E)$.

N6 - the definitions of probability are mathematical definitions. They tell us which set functions can be called **probability functions**. They do not tell us what value a probability function $P(\cdot)$ assigns to a given event E .

probability function \iff it satisfies the 3 axioms.

N7 - $P(E^c) = 1 - P(E)$

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

N9 - $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

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

$$\begin{aligned} P(E \cup F \cup G) &= P(E) + P(F) + P(G) \\ &\quad - P(EF) - P(EG) - P(FG) \\ &\quad + P(EFG) \end{aligned}$$

N11 - Inclusion-Exclusion identity -

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

Proof. Suppose an outcome with probability ω is in exactly m of the events E_i , where $m > 0$. Then

LHS: the outcome is in $E_1 \cup E_2 \cup \dots \cup E_n$ and ω will be counted once in $P(E_1 \cup E_2 \cup \dots \cup E_n)$

RHS:

- the outcome is in exactly m of the events E_i and ω will be counted exactly $\binom{m}{1}$ times in $\sum_{i=1}^n P(E_i)$

- the outcome is contained in $\binom{m}{2}$ subsets of the type $E_{i_1} E_{i_2}$ and ω will be counted $\binom{m}{2}$ times in $\sum_{i_1 < i_2} P(E_{i_1} E_{i_2})$

- ... and so on

hence $RHS = \binom{m}{1} \omega - \binom{m}{2} \omega + \binom{m}{3} \omega - \dots \pm \binom{m}{m} \omega$

$$\begin{aligned} &= \omega \sum_{i=0}^m \binom{m}{i} (-1)^i = \text{binomial theorem where } x = -1, y = 1 \\ &= 0 = LHS \end{aligned}$$

e.g. For an outcome with probability ω and $n = 3$

- Case 1.** $\omega = P(E_1 E_2)$
 LHS = ω
 RHS = $(\omega + \omega + 0) - (\omega + 0 + 0) + 0 = \omega$
- Case 2.** $\omega = P(E_1 \cap E_2 \cap E_3)$
 LHS = ω
 RHS = $(\omega + \omega + \omega) - (\omega + \omega + \omega) + \omega = \omega$

N12 -

$$(i) \quad P\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n P(E_i)$$

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

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

(iv) and so on.

Proof. $\bigcup_{i=1}^n E_i = E_1 \cup E_1^c E_2 \cup E_1^c E_2^c E_3 \cup \dots \cup E_1^c E_2^c \dots E_{n-1}^c E_n$

$$P\left(\bigcup_{i=1}^n E_i\right) = P(E_1) + P(E_1^c E_2) + P(E_1^c E_2^c E_3) + \dots + P(E_1^c E_2^c \dots E_{n-1}^c E_n)$$

Sample Space having Equally Likely Outcomes

tricky - 14, 15, 16, 18, 19, 20

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 $\{E_n, n \geq 1\} \rightarrow$

$E_1 \subset E_2 \subset \dots \subset E_n \subset E_{n+1} \subset \dots$

lim_{n -> infinity} E_n = union_{i=1}^infinity E_i

decreasing sequence of events {E_n, n ≥ 1} -> E_1 supset E_2 supset ... supset E_n supset E_{n+1} supset ...

lim_{n -> infinity} E_n = intersection_{i=1}^infinity E_i

03. CONDITIONAL PROBABILITY AND INDEPENDENCE

tricky - E6, urns (p.37)

Conditional Probability

- N1 - if P(F) > 0. then P(E|F) = (P(E intersection F) / P(F))
- N2 - multiplication rule - P(E_1 E_2 ... E_n) = P(E_1) P(E_2 | E_1) P(E_3 | E_1 E_2) ... P(E_n | E_1 E_2 ... E_{n-1})
- N3 - axioms of probability apply to conditional probability

- 1. 0 ≤ P(E|F) ≤ 1
- 2. P(S|F) = 1 where S is the sample space
- 3. If E_i (i ∈ Z_{≥1}) are mutually exclusive events, then

P(intersection_{i=1}^infinity E_i | F) = intersection_{i=1}^infinity P(E_i | F)

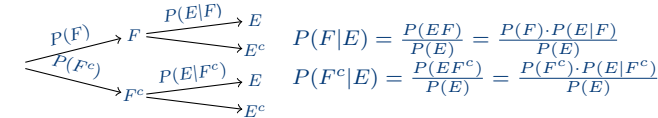
N4 - If we define Q(E) = P(E|F), then Q(E) can be regarded as a probability function on the events of S, hence all results previously proved for probabilities apply.

- Q(E_1 union E_2) = Q(E_1) + Q(E_2) - Q(E_1 E_2)
- P(E_1 union E_2 | F) = P(E_1 | F) + P(E_2 | F) - P(E_1 E_2 | F)
- theorem of total probability:
 - Q(E_1) = Q(E_1 | E_2) Q(E_2) + Q(E_1 | E_2^c) Q(E_2^c)
- P(H | F_n) = sum_{i=0}^k P(H | F_n c_i) P(c_i | F_n)

Total Probability & Bayes' Theorem

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

tree diagram -



Total Probability

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

Bayes Theorem

P(F_j | E) = (P(E F_j) / P(E)) = (P(F_j) P(E | F_j) / sum_{i=1}^n P(F_i) P(E | F_i))

application of bayes' theorem

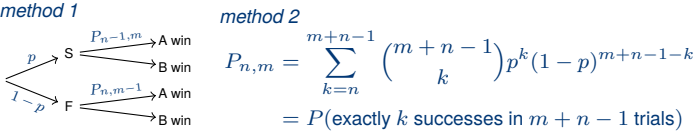
P(B_1 | A) = (P(A | B_1) · P(B_1) / (P(A | B_1) · P(B_1) + P(A | B_2) · P(B_2)))

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(A-bar B_1)
false positives: P(A B_2)	true negatives: P(A-bar B_2)

Independent Events

- N1 - E and F are independent <=> P(EF) = P(E) · P(F)
- N2 - E and F are independent <=> P(E|F) = P(E)
- N3 - if E and F are independent, then 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 union G)
- N5 - if E, F, G are independent, then P(EFG) = P(E)P(F)P(G)
- N6 - if E and F are independent and E and G are independent, \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 ≥ 1,



recursive approach to solving probabilities: see page 85 alternative approach

04. RANDOM VARIABLES

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

Types of Random Variables

- X is a Bernoulli r.v. with parameter p if -> p(x) = { p, x=1 ('success'); 1-p, x=0 ('failure') }
- Y is a Binomial r.v. with parameters n and p -> Y = X_1 + X_2 + ... + X_n where X_1, X_2, ..., X_n are independent Bernoulli r.v.'s with parameter p.
 - P(X = k) = C(n, k) p^k (1-p)^{n-k}
 - P(k successes from n independent trials each with probability p of success)
 - e.g. number of red balls out of n balls drawn with replacement
- 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} · 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
 - e.g. number of red balls out of n balls drawn without replacement

Summary

binomial	X = number of successes in n trials with replacement
negative binomial	X = number of trials until k successes
geometric	X = number of trials until a success
hypergeometric	X = number of successes in n trials without replacement

Properties

- N1 - if X ~ Binomial(n, p), and Y ~ Binomial(n - 1, p), then E(X^k) = np · E[(Y + 1)^{k-1}]
- N2 - if X ~ Binomial(n, p), then for k ∈ Z^+, P(X = k) = ((n-k+1)p / (k(1-p))) · 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 ≥ n) = P(T = n) + P(T > n) -> P(T = n) = P(T > n - 1) - P(T > n) Let A_j = {no type j coupon is contained among the first n} P(T > n) = P(union_{j=1}^N A_j)

Using the inclusion-exclusion identity,
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
+ ... + (-1)^{k+1} sum_{j_1 j_2} ... sum_{j_k} P(A_{j_1} A_{j_2} ... A_{j_n}) + ...
+ (-1)^{N+1} P(A_1 A_2 ... A_N)
P(A_{j_1} A_{j_2} ... A_{j_n}) = ((N-k) / N)^n

Hence P(T > n) = sum_{i=1}^{N-1} C(N, i) C(N-1, n) (-1)^{i+1}

Probability Mass Function

- for a discrete r.v., we define the probability mass function (pmf) of X by p(a) = P(X = a)
 - cdf, F(a) = sum p(x) for all x ≤ a
- if X assumes one of the values x_1, x_2, ..., then sum_{i=1}^infinity p(x_i) = 1
- the pmf p(a) is positive for at most a countable number of values of a

e.g. a/p(a) | 1/2, 2/4, 4/4

- discrete variable -> a random variable that can take on at most a countable number of possible values

Cumulative Distribution Function

- for a r.v. X, the function F defined by F(x) = P(X ≤ x), -infinity < x < infinity, is called the cumulative distribution function (cdf) of X.
 - aka distribution function
 - F(x) is defined on the entire real line

e.g. F(a) = { 0, a < 1; 1/2, 1 ≤ a < 2; 3/4, 2 ≤ a < 4; 1, a ≤ 4 }

Expected Value

- aka population mean/sample mean, μ
- if X is a discrete random variable having pmf p(x), the expectation or the expected value of X is defined as E(X) = sum_x x · p(x)

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

N2 - the n^{th} moment of of X is given as E(X^n) = sum_x x^n · p(x)

- I is an indicator variable for event A if I = { 1, if A occurs; 0, if A^c occurs }. then E(I) = P(A).

Proof of N1. E(aX + b) = sum_x (aX + b)p(x) = a · sum_x xp(x) + b · sum_x p(x) = a · E(X) + b

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[g(x)] = sum_i g(x_i) p(x_i)
 - where X is a discrete r.v. that takes on one of the values of x_i with the respective probabilities of p(x_i), and g is any real-valued function g

Variance

If X is a r.v. with mean μ = E[X], then the variance of X is defined by Var(X) = E[(X - μ)^2]

= sum x_i (x_i - μ)^2 · p(x_i) (deviation · weight)
= E(x^2) - [E(x)]^2

- Var(aX + b) = a^2 Var(x)

Poisson Random Variable

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

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

- notation: $X \sim \text{Poisson}(\lambda)$
- $\sum_{i=0}^\infty P(X = i) = 1$
- Poisson Approximation of Binomial** - if $X \sim \text{Binomial}(n, p)$, n is large and p is small, then $X \sim \text{Poisson}(\lambda)$ where $\lambda = np$.
 - For n independent trials with probability p of success, the number of successes is approximately a *Poisson r.v.* with parameter $\lambda = np$ if n is large & p is small.
 - Poisson approximation remains even when the trials are not independent, provided that their *dependence is weak*.
- 2 ways** to look at the Poisson distribution
 - an approximation to the binomial distribution with large n and small p
 - counting the number of events that occur at *random* at certain points in time

Mean and Variance

if $X \sim \text{Poisson}(\lambda)$, then $E(X) = \lambda, \text{Var}(X) = \lambda$

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

examples

Selecting hats problem

Let n be the number of men who select their own hats. Let I_E be an indicator r.v. for E . E_i is the event that the i -th man selects his own hat. Let X be the number of men that select their own hats.

- $X = I_{E_1} + I_{E_2} + \dots + I_{E_n}$
- $P(E_i) = \frac{1}{n}$
- $P(E_i | E_j) = \frac{1}{n-1} \neq P(E_j)$ for $j < i$ (hence E_i and E_j are not independent)
 - but dependence is weak for large n
- X satisfies the other conditions for binomial r.v., besides independence (n trials with equal probability of success)
- Poisson approximation of $X : X \sim \text{Poisson}(\lambda)$
 - $\lambda = n \cdot P(E_i) = n \cdot \frac{1}{n} = 1$
 - $P(X = i) = \frac{e^{-1} 1^i}{i!} = \frac{e^{-1}}{i!}$
 - $P(X = 0) = e^{-1} \approx 0.37$

No 2 people have the same birthday

For $\binom{n}{2}$ pairs of individuals i and j , $i \neq j$, let E_{ij} be the event where they have the same birthday. Let X be the number of pairs with the same birthday.

- $X = I_{E_1} + I_{E_2} + \dots + I_{E_n}$
- Each E_{ij} is only *pairwise independent*. $P(E_{ij}) = \frac{1}{365}$
 - i.e. E_{ij} and E_{mn} are independent

- but E_{12} and $(E_{13} \cap E_{23})$ are not independent $\Rightarrow P(E_{12} | E_{13} \cap E_{23}) = 1$
- $X \sim \text{Poisson}(\lambda), \lambda = \frac{\binom{n}{2}}{365} = \frac{n(n-1)}{730} \Rightarrow P(X = 0) = e^{-\frac{n(n-1)}{730}}$
 - for $P(X = 0) \leq \frac{1}{2}, n \geq 23$

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 $[0, t]$.

$$V \sim \text{Poisson}(5t) \Rightarrow P(V = k) = \frac{e^{-5t} \cdot (5t)^k}{k!}$$

$$P(X > t) = P(\text{no accidents happen during } [0, t]) = P(V = 0) = e^{-5t}$$

$$P(X \leq t) = 1 - e^{-5t}$$

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 $P(X \in B) = \int_B f(x) dx$

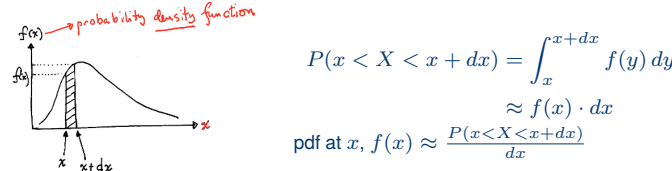
N1 - $P(X \in (-\infty, \infty)) = \int_{-\infty}^\infty f(x) dx = 1$

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 , $F_Y(y) = P(Y \leq y) = P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y$. hence Y is a uniform r.v.

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

Expectation & Variance

expectation

N1 - **expectation of X** , $E(X) = \int_{-\infty}^\infty x \cdot f(x) dx$

N2 - for a non-negative r.v. Y , $E(Y) = \int_0^\infty P(Y > y) dy$

N3 - if X is a continuous r.v. with pdf $f(x)$, then for any real-valued function g , $E[g(x)] = \int_{-\infty}^\infty g(x) f(x) dx$

• e.g. $E[aX + b] = \int_{-\infty}^\infty (aX + b) \cdot f(x) dx = a \cdot E(X) + b$

variance

N1 - variance of X , $\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$

example

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, $F_Y(y) = P(Y \leq y) = P((b - a)X + a \leq y) = P(X \leq \frac{y-a}{b-a})$

$$F_Y(y) = \int_0^{\frac{y-a}{b-a}} 1 dx = \frac{y-a}{b-a}, \quad a < y < b$$

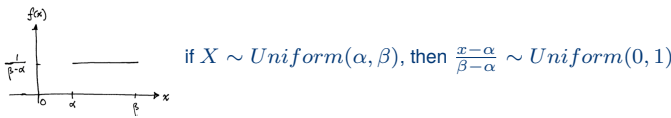
$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{1}{b-a}, & a < y < b \\ 0, & \text{otherwise} \end{cases}$$

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

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

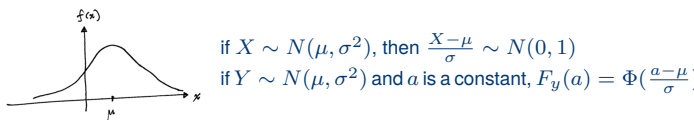


Normal Random Variable

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

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



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

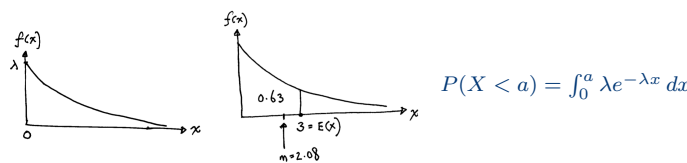
if $S_n \sim \text{Binomial}(n, p)$, then $\frac{S_n - np}{\sqrt{np(1-p)}} \sim N(0, 1)$ for large n .
 $\mu = np, \quad \sigma^2 = np(1 - p)$

Exponential Random Variable

a *continuous* r.v. X is a **exponential r.v.**, $X \sim \text{Exponential}(\lambda)$ or $\text{Exp}(\lambda)$

if for some $\lambda > 0$, its pdf is given by

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



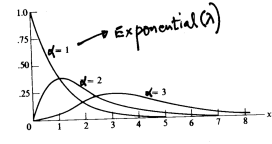
- an exponential r.v. is *memoryless*.
 - a non-negative r.v. is **memoryless** \rightarrow if $P(X > s + t | X > t) = P(X > s)$ for all $s, t > 0$.

Gamma Distribution

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

f(x) = { (lambda e^{-lambda x} (lambda x)^{alpha-1} / Gamma(alpha), x >= 0; 0, x < 0; E(X) = alpha/lambda, Var(X) = alpha/lambda^2

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

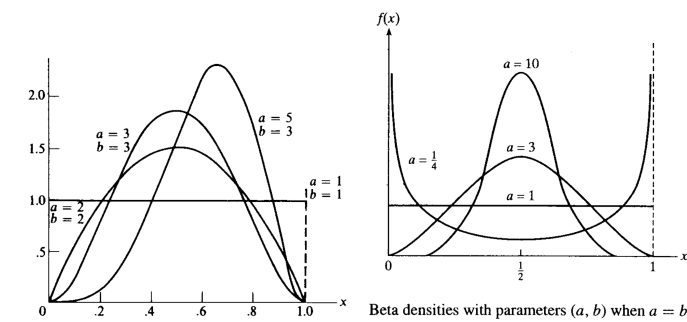


- N1 - $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- Proof. using integration by parts of LHS to RHS
- N2 - if α is an integer n , then $\Gamma(n) = (n - 1)!$
- N3 - if $X \sim \text{Gamma}(\alpha, \lambda)$ and $\alpha = 1$, then $X \sim \text{Exp}(\lambda)$.
- N4 - for events occurring randomly 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 event n occurs, $T_n \sim \text{Gamma}(n, \lambda)$
- number of events in time period $[0, t]$, $N(t) \sim \text{Poisson}(\lambda t)$
- N5 - $\text{Gamma}(\alpha = \frac{n}{2}, \lambda = \frac{1}{2}) = \chi_n^2$ (chi-square distribution to n degrees of freedom)

Beta Distribution

a r.v. X is said to have a **beta distribution**, $X \sim \text{Beta}(a, b)$ if its density is given by

f(x) = { (1/beta(a,b)) x^{a-1} (1-x)^{b-1}, 0 < x < 1; 0, otherwise; E(X) = a/(a+b), Var(X) = ab/((a+b)^2(a+b+1))



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

Cauchy Distribution

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

f(x) = 1/pi * 1/(1+(x-theta)^2), -inf < x < inf

Proof. E(X^n) does not exist for n in Z+; E(X) = integral from -inf to inf of x * f(x) dx = inf - inf (undefined)

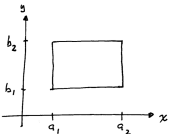
06. JOINTLY DISTRIBUTED RANDOM VARIABLES

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**, $F_Y(y) = \lim_{x \rightarrow \infty} F(x, y)$.



- N3 - $P(X > a, Y > b) = 1 - F_X(a) - F_Y(b) + F(a, b)$
- N4 - $P(a_1 < X \leq a_2, b_1 < Y \leq b_2) = F(a_2, b_2) + F(a_1, b_1) - F(a_1, b_2) - F(a_2, b_1)$

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)

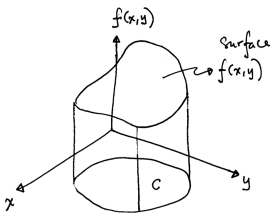
N1 - **marginal pmf of X**, $P(X = i) = \sum_j P(X = i, Y = j)$

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

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] = integral over C of f(x, y) dx dy; = volume under the surface over the region C.

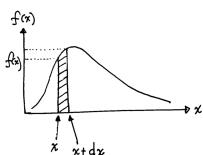


- N1 - if $C = \{(x, y) : x \in A, y \in B\}$, then $P(X \in A, Y \in B) = \int_A \int_B f(x, y) dx dy$

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

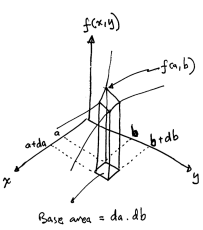
N3 - $f(a, b) = \frac{\delta^2}{\delta a \delta b} F(a, b)$

interpretation of pdf



P(x < X < x + dx) = integral from x to x+dx of f(y) dy; approx f(x) dx; pdf at x, f(x) approx P(x < X < x + dx) / dx

interpretation of joint pdf



P(a < X < a + da, b < Y < b + db) = integral from b to b+db integral from a to a+da of f(x, y) dx dy; approx f(a, b) da db; marginal pdf of X, f_X(x) = integral from -inf to inf of f(x, y) dy; marginal pdf of Y, f_Y(y) = integral from -inf to inf of f(x, y) dx

commutative	$E \cup F = F \cup E$	$E \cap F = F \cap E$
associative	$(E \cup F) \cup G = E \cup (F \cup G)$	$(E \cap F) \cap G = E \cap (F \cap G)$
distributive	$(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$	$(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$
DeMorgan's	$(\bigcup_{i=1}^n E_i)^c = \bigcap_{i=1}^n E_i^c$	$(\bigcap_{i=1}^n E_i)^c = \bigcup_{i=1}^n E_i^c$