SAMPLE SPACE & PROBABILITY.

day / date:

LECTURE 1:- AXIOMS OF PROBABILITY

→ Sample Space Ω is the set of all ~~possibilities~~ possible outcomes of an experiment.

↳ you rolled one die. What is Ω? $\{1, 2, 3, 4, 5, 6\}$

↳ the different elements of a sample space must be mutually exclusive ① and collectively exhaustive ②.

→ An event is a collection of possible outcomes

① Mutually exclusive means events are disjoint in nature, two or more events cannot occur at same time.

② collectively exhaustive means if one of events must occur.

→ Simple Event, is when only one event can occur.

→ Compound Event in probability is the chance of two or more events occuring. (they can be decomposed into simpler events)

→ A set is a collection of objects, which are called elements.

↳ the natural numbers are set, where the elements are individual numbers.

→ If an element x is in a set S, we write $\boxed{x \in S}$

→ If a set contains $\boxed{\text{no elements}}$, we call it the $\boxed{\text{empty set, } \phi}$.

→ If a set contains every possible element, we call it universal set Ω.

→ A set can be finite (e.g the set of people in this class) or infinite (e.g the set of real numbers)

→ If we can enumerate the elements of an infinite set i.e arrange the elements in a list, we say it countable

→ If we cannot enumerate the elements, we say it is uncountable.

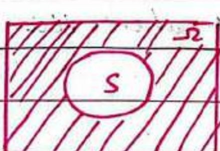
→ We can use $\boxed{\text{curly brackets}}$ to describe a set in terms of its elements.

$S = \{1, 2, 3, 4\}$ OR $S = \{x \mid x \text{ satisfies } C\}$

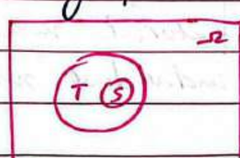
→ The complement, S^c of a set S, w.r.t Ω, is the set of all elements that are in Ω but not in S. So $\Omega^c = \phi$

→ We say $S \subseteq T$, if every element in S is also in T.

→ $S \subseteq T$ and $T \subseteq S$ if and only if $S = T$



S^c is the shaded region.



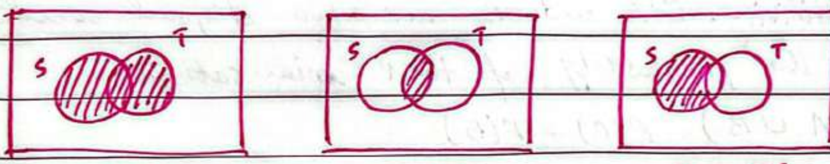
$S \subset T \subset \Omega$

day / date:

→ The union, $S \cup T$ of two sets S and T is the set of elements that are in either S or T (or both)
$S \cup T = \{x \mid x \in S \text{ or } x \in T\}$.

→ The intersection $S \cap T$ of two sets S and T is the set of elements that are in both S and T:
$S \cap T = \{x \mid x \in S \text{ and } x \in T\}$

→ The difference, $S \setminus T$ of two sets S and T is the set of elements that are in S but not in T:
$S \setminus T = \{x \mid x \in S \text{ and } x \notin T\}$



$S \cup T$ $\quad$ $S \cap T$ $\quad$ $S \cap T^c$

→ We can extend the notions of union and intersection to multiple sets:

$$\bigcup_{i=1}^{n} S_i = S_1 \cup S_2 \cup \dots S_n = \{x \mid x \in S_n \text{ for some } 1 \le i \le n\}$$

$$\bigcap_{i=1}^{n} S_i = S_1 \cap S_2 \cap \dots S_n = \{x \mid x \in S_n \text{ for all } 1 \le i \le n\}$$

* $P(\Omega) = 1$, $P(\phi) = 0$ day / date:

→ We say two sets are disjoint if their intersection is empty.

→ We say a collection of sets are disjoint if no two sets have any common elements.

→ If a collection of disjoint sets have union S, we call them a partition of S.

AXIOMS OF PROBABILITY

→ no mathematical proof

→ **Nonnegativity:** $P(A) \geq 0$ for every event A.

→ **Additivity:** If A and B are two disjoint events (→ or mutually exclusive.) then the probability of their union satisfies

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

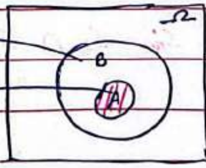
→ **Normalization:** The probability of the entire sample space Ω is equal to 1 i.e $P(\Omega) = 1$

All the following can be proven by decomposing a set into disjoint partitions and using the additivity and non negativity rules.

→ if $A \subseteq B$ then $P(A) \leq P(B)$

$A^c \cap B$

not only A, but also $A = A \cap B$



$A \subseteq B$

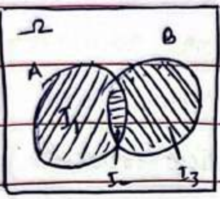
$$B = \overbrace{(A \cap B) \cup (B \cap A^c)}^{\text{disjoint}}$$

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

$$= P(A) + \underbrace{P(A^c \cap B)}_{\geq 0}$$

day / date:

→ if $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ * not dis, $A \cap B \neq \emptyset$



I_1, I_2, I_3 are disjoint

$A = I_1 \cup I_2$

$B = I_2 \cup I_3$

$P(A) = P(I_1) + P(I_2)$; $P(B) = P(I_2) + P(I_3)$

$A \cup B = I_1 \cup I_2 \cup I_3$

$P(A \cup B) = P(I_1) + P(I_2) + P(I_3)$

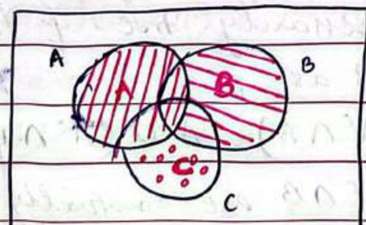
$P(A \cap B) = P(I_2)$

$P(A \cup B) = \underbrace{P(I_1) + P(I_2)}_{P(A)} + \underbrace{P(I_2) + P(I_3)}_{P(B)} - \underbrace{P(I_2)}_{P(A \cap B)}$

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

→ $P(A \cup B) \leq P(A) + P(B)$ * if disjoint then

→ $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$



AXIOM 1 $0 \leq P(E) \leq 1$

AXIOM 2 $P(\Omega) = 1$

AXIOM 3 $P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$

$\star$ P(

day / date:

axioms of probability

...om 1: $0 \le P(E) \le 1$

...xiom 2: $P(\Omega) = 1$

Axiom 3:

For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i \cap E_j = \phi$ when $i \neq j$)

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

We refer to $P(E)$ as the probability of the event E.

Proposition: If $A \subset B$, then $P(A) \le P(B)$

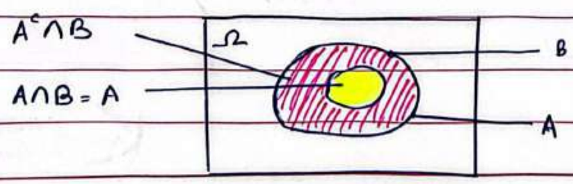
Proof: Since $A \subset B$, A is a proper subset of B, all elements of A are contained by B but the converse is not necessarily true; it follows that we can express B as

$$B = (A \cap B) \cup (A^c \cap B) = A \cup (A^c \cap B)$$

Because $A \cap B$ and $A^c \cap B$ are mutually exclusive where $A^c \cap B$ is the region (or elements) that belongs to B but is outside A; therefore we obtain from Axiom 3,

$$P(B) = P(A) + P(A^c \cap B) \ge P(A)$$

which proves the result, since $P(A^c \cap B) \ge 0$



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

Proof: Let us divide $A \cup B$ into three mutually exclusive sections as shown in Figure below. In II represents all points both in A and in B (that is $II = A \cap B$) and section III represents all points in B that are not in A (that is $III = A^c \cap B$). From figure, we see that

$$A \cup B = I \cup II \cup III$$

$$A = I \cup II$$

$$B = II \cup III$$

As I, II and III are mutually exclusive it follows from Axiom 3 that.

$$P(A \cup B) = P(I) + P(II) + P(III)$$

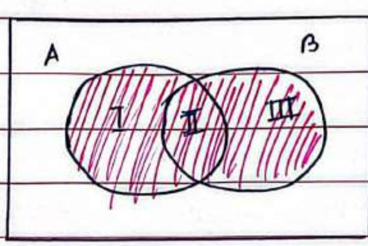
$$P(A) = P(I) + P(II)$$

$$P(B) = P(II) + P(III)$$

which shows that

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

and above Proposition is proved since region $II = A \cap B$



PROPERTIES OF PROBABILITY LAWS.

(a) if $A \subset B$, then $P(A) \leq P(B)$

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *not disjoint

(c) $P(A \cup B) \leq P(A) + P(B)$ * disjoint

(d) $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$

* when all events are equal likely we will take Area. $\frac{P(A).\text{ Area}}{\text{Total Area}}$

* sometimes however we have partial information that may affect the likelihood of given event.

*

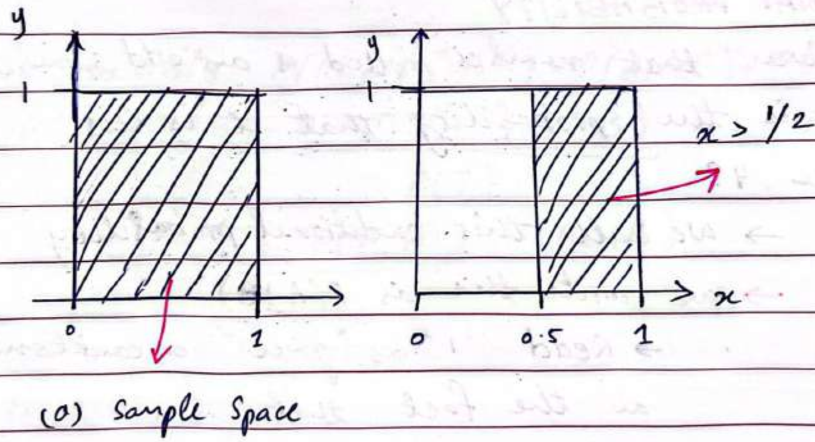
→ The conditional distribution of A given B can be written as $P(A|B)$ where ~~&~~ as you read | as "given" or "conditioned on the fact that"

day / date:

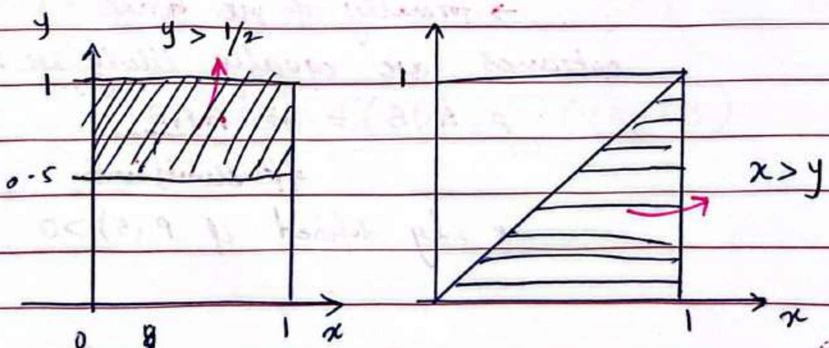
EXAMPLE:

Consider an experiment where we picked two numbers x and y at random b/w zero & one. The sample space is then the unit square. If we suppose that all pairs of numbers in the unit square are equally likely to be selected, then it is reasonable to use probability assignment in which the probability of any region R inside the unit square is equal to area of R.

Find the probability of following events: $A = \{x > 0.5\}$ and $B = \{y > 0.5\}$ $C = \{x > y\}$



(a) Sample Space



day / date:

PARTIAL INFORMATION

→ Sometimes, however we have partial information that may affect the likelihood of given event.

→ ~~You~~ the experiment involves rolling a die. You are told that the number is ODD.

→ The experiment involves the weather tomorrow. You know that the weather today is rainy.

CONDITIONAL PROBABILITY

→ Given that number rolled is an odd number, what is the probability that it is less than 4?

→ We call this conditional probability

→ we write this as $P(A|B)$

→ Read '|' as "given" or conditioned on the fact that.

→ formally if ~~we~~ ~~in~~ all outcomes are equally likely, we have

$$P(A|B) = \frac{\#\ A \cap B}{\#\ \text{elements in } B}$$

* only defined if $P(B) > 0$

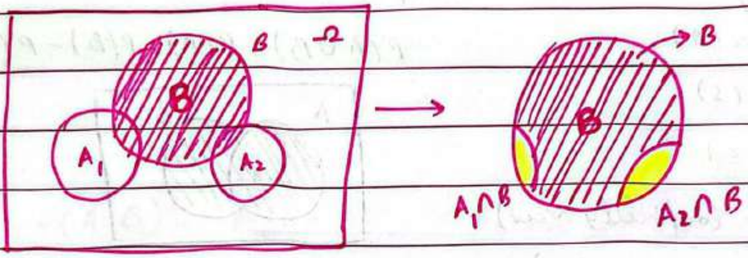
day / date:

CONDITIONAL PROBABILITY AXIOMS

→ Nonegativity: - Check

→ Normalization: - Your new universe is now B and we know that $P(B|B) = 1$

→ Additivity: - $P(A_1 \cup A_2 \mid B) = P(A_1|B) + P(A_2 \mid B)$ for two disjoint sets, A_1 and A_2.



Using additivity, $P((A_1 \cup A_2) \cap B) = P(A_1 \cap B) + P(A_2 \cap B)$, so

$$P(A_1 \cup A_2 | B) = \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)}$$

$$= P(A_1 | B) + P(A_2 | B)$$

PROPERTIES OF CONDITIONAL PROBABILITY

If $P(B) > 0$,

① → If A_i for $i \in \{1, \dots, n\}$ are all pairwise then

$$P\left(\cup_{i=1}^{n} A_i \mid B\right) = \sum_{i=1}^{n} P(A_i \mid B)$$

(2) $\rightarrow$ if $A_1 \subseteq A_2$ then $P(A_1|B) \leq P(A_2|B)$

(3) $\rightarrow$ $P(A_1 \cup A_2 | B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2 | B)$

(4) $\rightarrow$ Union bound: $P(A_1 \cup A_2 | B) \leq P(A_1|B) + P(A_2|B)$

$\hookrightarrow P\left(\bigcup_{i=1}^{n} A_i | B\right) \leq \sum_{i=1}^{n} P(A_i|B)$

FORMULAS

* Probability

$$P(A) = \frac{n(A)}{n(S)}$$

* $0 \leq P(E) \leq 1$

$P(E) = 1$ (definitely occur)

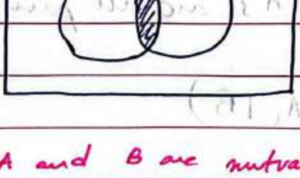
$P(E) = 0$ (impossible)

* Complement

$P(A') = 1 - P(A)$

* if not DISJOINT

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

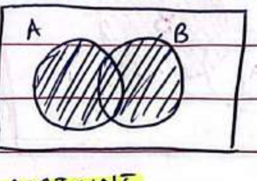


* if A and B are mutually exclusive, $A \cap B = \phi$

and $P(A|B) = 0$

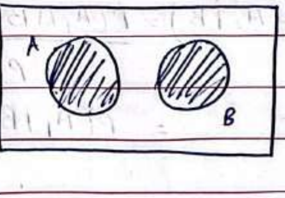
* If not DISJOINT

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



* if DISJOINT

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

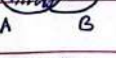


* if DISJOINT

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

* Conditional Probability

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



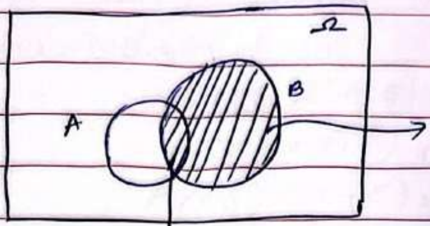
A B $A \rightarrow B$

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

A B $B \rightarrow A$

→ The second axiom is shown to be satisfied by letting $\Omega = A$

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$



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

$$= P(B|B) = 1$$

$$P(A) + P(A^c) = 1$$

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

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

Bay's Rule

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

TOTAL PROBABILITY

$A = \{A \cap B_1\} \cup \{A \cap B_2\} \cup \{A \cap B_n\} \cup \ldots$

$P(A) = P(A \cap B_1) + P(A \cap B_n)$

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

$$= \sum_{i=1}^{n} P(A|B_i)\,P(B_i)$$

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2)$$

day / date:

BAY'S RULE

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

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

① — $P(A \cap B) = P(A|B)P(B)$

② — $P(A \cap B) = P(B|A)P(A)$

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

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

INDEPENDENT EVENTS

Since $P(E|F) = \frac{P(EF)}{P(F)}$, it follows that E is independent of F if.

$$P(EF) = P(E) \times P(F)$$

→ shows that whenever E is independent of F, F is also independent of E.

→ Two events E and F are said to be independent if above equation holds

→ two events E and F that are not independent are said

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

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

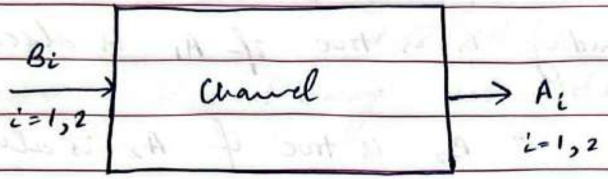
$P(A \cap B) = P(A)P(B)$

day / date:

Three events E, F and G are said to be independent if

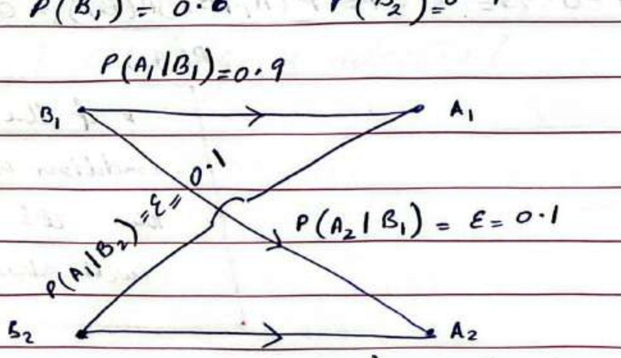
$$P(EFG) = P(E) \times P(F) \times P(G)$$
$$P(EF) = P(E) \times P(F)$$
$$P(EG) = P(E) \times P(G)$$
$$P(FG) = P(F) \times P(G)$$

—— × —— × —— ×

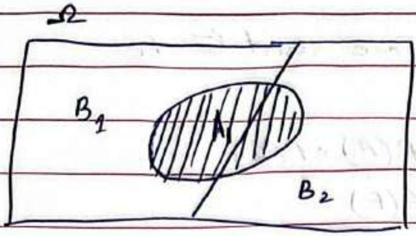


B_1 = {The symbol before the chanel is 1}
B_2 = {The symbol before the chanel is 0}
A_1 = {The symbol after the chanel is 1}
A_2 = {The symbol after the chanel is 0}

$P(B_1) = 0.6 \quad P(B_2) = 0.4$



day / date: DISJOINT



$A_1 = \{A_1 \cap B_1\} \cup \{A_1 \cap B_2\}$

$P(A_1) = P(A_1 \cap B_1) + P(A_1 \cap B_2)$

$P(A_1) = P(A_1 | B_1) P(B_1) + P(A_1 | B_2) P(B_2)$

$= (0.9)(0.6) + (0.1)(0.4)$

$= 0.58 = 58\%$ of time you'll receive 1.

$P(A_2) = 42\%$ OR 0.42

$P(B_1 | A_1)$ = Finding B_1 is true if A_1 is already true.

$P(B_2 | A_2) =$ " " B_2 is true if A_2 is already true.

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

$$P(B_1 | A_1) = \frac{P(A_1 | B_1) P(B_1)}{P(A_1)} = \frac{0.9 * 0.6}{0.58}$$

$= 0.93$

$$P(B_2 | A_1) = 1 - 0.93 = 0.07 = \frac{P(A_1 | B_2) P(B_2)}{P(A_1)} = \frac{0.1 * 0.4}{0.58}$$

* if the "given" condition is SAME then it's sum will always be 1.

KAGHAZ www.kaghaz.pk

day / date:

$$P(B_1 | A_2) = \frac{P(A_2 | B_1) P(B_1)}{P(A_2)}$$

$$= \frac{0.1 * 0.6}{0.42} = 0.143$$

Radar Detection

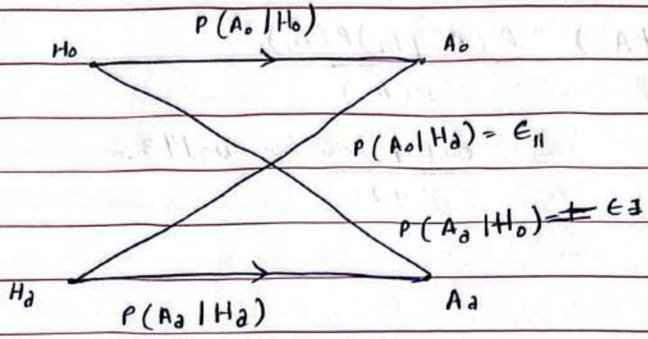
null event = H_0 = { everything is fine / no enemy detected }

alternate " " = H_A = { enemy detected }

A_0 = { nobody is detected }

A_1 = { something is detected }

day / date:



Proof of false alarm $= P(A_a | H_0) = \epsilon_I =$ Type-I Error

Proof of Miss $= P(A_0 | H_a) = \epsilon_{II} -$ Type-II Error

$$P(A_a) = P(A_a | H_0)\, P(H_0) + P(A_a | H_a)\, P(H_a)$$

INDEPENDENT EVENTS (CONTD.)

if A and B are independent events

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

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

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

- Geometric Experiment

outcomes $= \{T, F\}$

$\rightarrow$ Sample Space $\Rightarrow \{T, FT, FFT, FFFT, FFFFT, \ldots\} \cup \{FFFF\ldots\}$

day / date:

$P(T) = P$

$P(F) = 1 - P$

$P(FT) = P(F)\,P(T)$ → independent events

$= (1-P)(P)$

$P(FFT) = P(F)\,P(F)\,P(T)$

$= (1-P)^2(P)$

Let $P = 1/2$

$P(\Omega) = P(T) + P(FT) + P(FFT) + \cdots$

$= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$ (geometric series)

$= \frac{1}{1 - 1/2} - 1 = 1$

$P(FFF\cdots) = [P(F)]^{\infty}$

$= (0.5)^{\infty} = 0$ →

flip two pages

day / date:

BAYES RULE

simple rule to get conditional probability of A given B, from the conditional formula of B given A.

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

This is very useful for inferring hidden causes underlying our observations.

→ If $P(A|B) = P(A)$, we say that events A and B are independent.

→ We can rewrite our definition by writing

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

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

$$\boxed{P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + \ldots + P(E_n)P(A|E_n)}}$$

day / date:

Example

A gambler is rolling a fair dice. What is the probability that there is at least one 6 in 4 rolls.

$$P(\text{at least 1 six in 4 rolls}) = 1 - P(\text{no sixes in 4 rolls})$$
$$= 1 - P(X_1 \cap X_2 \cap X_3 \cap X_4)$$
$$= 1 - P(X_1)P(X_2)P(X_3)P(X_4)$$
$$= 1 - \left(5/6\right)^4$$
$$= 0.518$$

→ if A and B are independent then so are
  → A and B^c = $P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)P(B^c)$
  → A^c and B
  → A^c and B^c

→ To see mutually independent we first check from data and compare with $P(A) \times P(B)$
$$P(A \cap B) = P(A)P(B)$$

→ if A, B, C are mutually independent, then so are A^c, B^c, C and A, B^c, C or A^c, B, C^c etc

KAGHAZ www.kaghaz.pk

day / date:

→ Two events are conditionally independent given another event C if $P(A \cap B \mid C) = P(A \mid C)\, P(B \mid C)$

MULTIPLICATION RULE

$P(A \cap B) = P(A)\, P(B \mid A)$ — without replacement

example:

$A \to 2R \; 3G \; 4Y$

$B \to 3R \; 1G \; 2Y$

$P(A) = 1/2 \qquad P(B) = 1/2$

$P(R_1 \mid A) = 2/9 \qquad P(R_2 \mid B) = 3/6$

$P(R) = P(A)P(R_1 \mid A) + P(B)P(R_2 \mid B)$ Total Probability

$= \frac{1}{2} \times \frac{2}{9} + \frac{1}{2} \times \frac{3}{6}$

Now Bayes Rule

$$P(A \mid R) = \frac{P(A)P(R_1 \mid A)}{\underbrace{P(A)P(R_1 \mid A)}_{\text{first bag}} + \underbrace{P(B)P(R_2 \mid B)}_{\text{second bag}}}$$

day / date:

POISSON

$$P(X=r) = \frac{e^{-m}\, m^{r}}{r!}$$

→ mean

exp

random.

Mean = m.

Conditions:

→ n is very large

→ p is very small.

DISCRETE RANDOM VARIABLE

day / date:

→ A random variable is discrete if it's range is $\boxed{\text{finite}}$ or at most $\boxed{\text{countably finite}}$. COUNTABLE

→ X = sum of two rolls of die $X \in \{2, \dots, 12\}$

→ X = number of heads in 100 coin tosses
$X \in \{0, \dots, 100\}$

→ In general, the probability that a random variable X takes up a value x is written as $p_X(x)$, or $P_X(x)$ or $P_X(X = x)$ etc.

→ A random variable is always written in uppercase and numerical value we are trying to evaluate probability for is written in lower case.

PROPERTIES

→ $\sum_x P(X = x) = 1$

→ $P(X \in S) = \sum_{x \in S} P(X = x)$

→ to compute $P(X = x)$

→ collect all possible outcomes that give $\{X = x\}$ $\sum_x P(X = x)$

→ add their probabilities to get $P(X = x)$

UNIFORM RANDOM VALUE

→ Each value has equal probability mass

→ If an uniform random variable X takes on k different values, the probability mass at each of those values are $1/k$.

day / date:

BERNOULLI RANDOM VARIABLE

→ It can take up to two values : 1 if a head comes up and 0 if not.

indicator function → $X = \begin{cases} 1 & \text{if head} \\ 0 & \text{if tail} \end{cases}$

→ The pmf is given by

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

??

BIASED COIN

→ You are throwing 10 biased coins with $P(\{H\}) = p$.

· What is probability that sum X equals 5?

→ probability of length 10 binary sequence with 5 1's is $p^5(1-p)^5$

→ so $P(X=5) = \binom{10}{5} p^5 (1-p)^5$

→ What about $P(X=8)$?

$$\binom{10}{8} p^8 (1-p)^2$$

→ In general $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

day / date:

BINOMIAL RANDOM VARIABLE

→ Let X_i be Bernoulli Random Variable which is 1 if the ith toss gave a head. Then $\{X_i ; i = 1, \cdots, n\}$ are independent Bernoullis

→ Let Y be number of heads we see at end of all n tosses.

→ Y is called Binomial Random Variable. What is PMF of y?

1. $P(Y=0) = P(\{\text{no heads}\}) = (1-p)^n$ q
2. $P(Y=n) = P(\{\text{all heads}\}) = p^n$
3. $P(Y=k) = P(\{k \text{ heads}\}) = \binom{n}{k} p^k (1-p)^{n-k}$
4. $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$ (BINOMIAL EXPANSION)

$$X \sim B(n, p)$$

probability of success

no of trials.

probability of failure

$$P(X=r) = {}^nC_r \, (p)^r \, (q)^{n-r}$$

→ The bernoulli PMF describes the probability of success/failure in single trial.

→ The binomial PMF describes probability of k successes out of n trial.

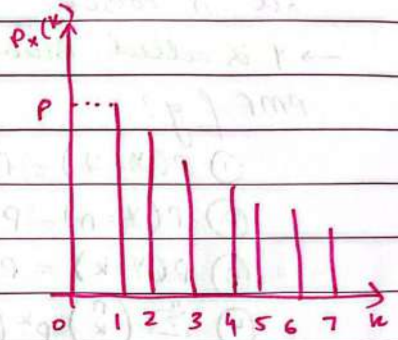
day / date:

GEOMETRIC RANDOM VARIABLE

→ we repeatedly toss a biased coin $(P(\{H\}) = p)$
The geometric random variable is number of
X of tosses to get head.

→ $P(X=k) = P(\{\underbrace{TT\ldots T}_{k-1}H\}) = (1-p)^{k-1}p$

→ $\sum_k P(X=k) = 1$



→ $P(X=n)$
$(q)^{n-1}(p)$

→ $P(X<n)$
$1-q^{n-1}$

→ $P(X \le n)$
$1-q^n$

→ $P(X>n)$
$(q)^n$

* where $q = 1-p$

→ $P(X \ge n) = (q)^{n-1}$

MEMORYLESS PROPERTY

→ What is $P(X > a+b \mid X > a)$?

$$P(X > a+b \mid X > a) = \frac{P(X > a+b)}{P(X > a)}$$

$$= \frac{(1-p)^{a+b}}{(1-p)^a} = (1-p)^b$$

$$= P(X > b)$$

day / date:

→ what is $P(X \le a+b \mid X > a)$?

$$P(X \le a+b \mid X > a) = \frac{P(a < X \le a+b)}{P(X > a)}$$

$$= \frac{P(X>a) - P(X > a+b)}{(1-p)^a}$$

$$= \frac{(1-p)^a - (1-p)^{a+b}}{(1-p)^a}$$

$$= 1 - (1-p)^b = P(X \le b)$$

POISSON RANDOM VARIABLE

→ A poisson random variable takes non negative integers as values. It has non-negative parameter

→ $P(X=k) = e^{-\lambda} \dfrac{\lambda^k}{k!}$ for $k = 0, 1, 2, \dots$

→ $\sum_{k=0}^{\infty} P(X=k) = e^{-\lambda}\left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right) = 1$

→ When n is very large and p is very small a binomial random variable can be well approximated by Poisson with $\lambda = np$.

→ More formally we see that $\binom{n}{k} p^k (1-p)^{n-k} \approx \dfrac{e^{-\lambda}\lambda^k}{k!}$

↳ when n is large, k if fixed and p is small and $\lambda = np$

↑ mean

EXAMPLE Assume that on given day 1000 cars are out in Austin. On an average (mean = λ) three out of 1000 cars run into traffic accident per day.

① What is probability that we see at least two accidents in a day?

Use poisson distribution.

$P(X \geq 2) = 1 - P(X=0) - P(X=1)$ $\qquad \lambda = 3$

$= 1 - e^{-3}(1+3)$

② If you know there is atleast one accident, what is probability that total number of accidents is atleast two?

$P(X \geq 1) = 1 - P(X=0) = 1 - e^{-3} = 0.950$

$P(X \geq 2 \mid X \geq 1) = \dfrac{P(X \geq 2)}{P(X \geq 1)} = \dfrac{0.8}{0.950} = 0.84$

CONDITIONAL PMF

→ The rules are same. $P(X=x \mid A) = \dfrac{P(\{X=x\} \cap A)}{P(A)}$

→ The conditional pmf is valid PMF.

$$\sum_{x} P(X = x \mid A) = 1$$

day / date:

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

What is different values of k for $P(X = k \mid X > 1)$?

$$P(X = k \mid X > 1) = \begin{cases} 0 & \text{if } k = 1 \\ \dfrac{P(X=k)}{P(X>1)} & \text{otherwise.} \\ = \dfrac{(1-p)^{k-1} p}{(1-p)} & \\ = (1-p)^{k-2} p = P(X = k-1) & \end{cases}$$

CUMULATIVE DISTRIBUTION FUNCTIONS

→ for any random variable, CDF is defined as

$$F_X(a) = \sum_{x \le a} p(x)$$

MEAN

EXAMPLE You want to calculate average grade points from hw1. You know that 20 students got 30/30, 30 students got 25/30 and 50 students got 20/30. What's average grade point?

$$\frac{30 \times 20 + 25 \times 30 + 20 \times 50}{100} = 30 \times 0.2 + 25 \times 0.3 + 20$$

→ how will you calculate $P(X=30)$ $X \sim$ grade points.

$P(X=30) = 20/100 = 0.2$

$P(X=25) = 30/100 = 0.3$

$P(X=20) = 50/100 = 0.5$

Average Grade $= 30 \times P(X=30) + 25 \times P(X=25) + 20 \times P(X=20)$

EXPECTATION

We define expected value of a discrete random variable X by

$$E[X] = \sum_x x P(X=x)$$

X is bernoulli random variable

$$P(X=x) = \begin{cases} p & X=1 \\ 1-p & X=0 \end{cases}$$

so $E[X] = 1 \times p + 0 \times (1-p) = p$

→ let's say you want to compute $E[g(x)]$

$$\rightarrow E[g(x)] = \sum_x g(x) P(X=x)$$

$$\rightarrow E[X^2] = \sum_x x^2 P(X=x)$$ Second Moment of X

$$\rightarrow E[X^3] = \sum_x x^3 P(X=x)$$ Third Moment of X

★ Expectation and Mean are <u>SAME</u> day / date:

★ → $E[X^k] = \sum_x x^k P(X=x)$ k^{th} moment of X

VARIANCE

$$VAR(X) = E[(X - E[X])^2] \text{ or } E[(X-\mu)^2]$$

The standard deviation of X is given by $\sigma_X = \sqrt{var_X}$

$$VAR(X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 P(X=x)$$

This can be made simpler

★ $VAR(X) = E[X^2] - (E[X])^2$

Say you are looking at linear function of your random variable X.

→ $Y = aX + b$

$= E[Y] = E[aX+b] = aE[X]+b$

$= \sum_x (ax+b) P(X=x)$

$= a\sum_x x P(X=x) + b \sum_x P(X=x)$

$= aE[X]+b$

day / date:

$\rightarrow Y = aX^2 + bX + c$

$= E[Y] = E[aX^2 + bX + C] = aE[X^2] + bE[X] + c$

$= \sum_x (ax^2 + bx + c)\, p(X = x)$

$= a\sum_x x^2 p(X = x) + b\sum_x x p(X = x) + c\sum_x p(X = x)$

$= aE[X^2] + bE[X] + c$

$\rightarrow Y = aX^3 + bX^2 + cX + d$

$= E[Y] = aE[X^3] + bE[X^2] + cE[X] + d.$

$\rightarrow Y = X + b$

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

$= E[X^2 + 2bX + b^2] - (E[X] + b)^2$

$= E[X^2] + 2bE[X] + b^2 - ((E[X])^2 + 2bE[X] + b^2)$

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

$\rightarrow Y = aX$

$= Var(aX) = E[(aX)^2] - (E[aX])^2$

$= E[a^2X^2] - (aE[X])^2$

$= a^2 E[X^2] - a^2(E[X])^2$

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

day / date:

MEAN AND VARIANCE OF BERNOULLI

X is Bernoulli Random Variable with $P(X=1)=p$.

we saw that $E[X]=p$. What is $Var(X)$?

$\hookrightarrow E[X] = 1 \times p = p$

① Let's get $E[X^2]$

$$E[X^2] = (1^2 \times P(X=1) + 0^2 \times P(X=0)) = p$$

$$E[X^2] = E[X]$$ we see they are identical.

② What is PMF of X^2?

$\rightarrow X^2$ can take two values: 0 and 1

$\rightarrow P(X^2=1) = P(X=1) = p$

$P(X^2=0) = P(X=0) = 1-p$

X and X^2 have IDENTICAL PMF!

$\rightarrow Var(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p)$

Mean $(\mu) = (0 \times (1-p) + 1 \times p) = p$

Variance $(\sigma^2) = p(1-p)$

MEAN AND VARIANCE OF BINOMIAL

Mean $(\mu) = np$

Variance $(\sigma^2) = np(1-p)$

day / date:

MEAN AND VARIANCE OF POISSON

Mean (μ) = λ

Variance (σ^2) = λ

Hint: Mean and Variance of Binomial approach that of Poisson when n is large and p is small such that $np \approx \lambda$.

MEAN AND VARIANCE OF GEOMETRIC

→ The pmf of geometric distribution is

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

Mean (μ) = $1/p$

Variance (σ^2) = $(1-p)/p^2$

CONDITIONAL EXPECTATION

→ Conditional Expectation of random variable X conditioned on event A is written as $E[X|A]$

$$E[X|A] = \sum_x x\,P(X=x|A)$$

EXAMPLE $X \sim$ geometric(p). What is $E[X|X>1]$?

$$P(X=k \mid X>1) = P(X=k-1)$$

$$E[X|X>1] = \sum_{k=2}^{\infty} k\,P(X=k|X>1)$$

$$= \sum_{k=2}^{\infty} k\,P(X=k-1)$$

day / date:

$$= \sum_{j=1}^{\infty} (j+1) P(X=j)$$

$$= \sum_{j=1}^{\infty} j\, P(X=j) + 1 = E[X]+1$$

TOTAL EXPECTATION THEOREM

→ Consider disjoint events $\{A_1, \ldots, A_n\}$ which form partition of the sample space

→ The total probability theorem says

$$P(X=k) = \sum_i P(X=k \mid A_i) P(A_i)$$

→ Similarly total expectation theorem says:

$$E[X] = \sum_i E[X \mid A_i] P(A_i)$$

DERIVATION OF MEAN OF GEOMETRIC.

→ define two disjoint events $\{X=1\}$ (first trial is success) and $\{X>1\}$

→ we have $E[X] = E[X \mid X=1] P(X=1) + E[X \mid X>1] P(X>1)$

→ $E[X \mid X=1] = 1$ $\quad$ $E[X \mid X>1] = E[X]+1$

$P(X=1) = p$ $\quad$ $P(X>1) = 1-p$

$$E[X] = p + (1+E[X])(1-p)$$

$$E[X] = 1/p$$

day / date:

DERIVATION OF VARIANCE OF GEOMETRIC

$$E[X^2] = \underbrace{E[X^2 \mid X=1]P(X=1)}_{1 \times p} + \underbrace{E[X^2 \mid X>1]P(X>1)}_{1-p}$$

Now

$$E[X^2 \mid X>1] = \sum_{k=2}^{\infty} k^2 P(X=k \mid X>1) = \sum_{k=2}^{\infty} k^2 \underbrace{P(X=k-1)}_{\text{memoryless property.}}$$

(Refer to lecture 10a slide 88)

MUlTIPLE RANDOM VARIABLE

~~So far~~

→ Consider two discrete random variables X and Y associated with same experiment.

→ The joint PMF of X and Y are defined as $P_{X,Y}(x,y) = P(X=x, Y=y)$ for all pairs of values x, y X and Y can take.

→ This is none other than $P(\{X=x\} \cap \{Y=y\})$

PROPERTIES OF JOINT PMF

→ $A_1, A_2, \ldots, A_k$ is partition of Ω

$$P(B) = P\left(\bigcup_k (B \cap A_k)\right) = \sum_k P(B \cap A_k)$$

day / date:

→ $\{X=x\}$ is disjoint union of $\{X=x\} \cap \{Y=y\}$ for all y values Y can take.

→ $\{X=x\} \cap \{Y=y\}$ is none other than $\{X=x, Y=y\}$

→ we can extend this to PMFs:

$$\sum_y P(X=x, Y=y) = P(X=x)$$

$$\sum_x P(X=x, Y=y) = P(Y=y)$$

→ Marginal PMFs

so, $\sum_x \sum_y P(X=x, Y=y) = \sum_x P(X=x) = 1$

FUNCTIONS OF MULTI RANDOM VARIABLES

→ $E(g(X,Y)) = \sum_{x,y} g(x,y) P(X=x, Y=y)$

→ let $g(X,Y) = aX+bY$

→ $E(g(X,Y)) = \sum_{x,y} (ax+by) P(X=x, Y=y) = aE[X] + bE[Y]$

→ What if $g(X,Y) = aX^2 + bY^2 + c$?

$E[g(X,Y)] = aE[X^2] + bE[Y^2] + c$

MULTIPLE RANDOM VARIABLE

How about three variables?

→ we will write $p_{X,Y,Z} = P(X=x, Y=y, Z=z)$

day / date:

The rules are same

① $P(X=x, Y=y) = \sum_z P(X=x, Y=y, Z=z)$

② $P(X=x) = \sum_{y,z} P(X=x, Y=y, Z=z)$

③ $P(Y=y) = \sum_{x,z} P(X=x, Y=y, Z=z)$

④ $P(Z=z) = \sum_{x,y} P(X=x, Y=y, Z=z)$

⑤ $\sum_{x,y,z} P(X=x, Y=y, Z=z) = 1$

LINEARITY OF EXPECTATION.

$$E[aX + bY + cZ + d] = aE[X] + bE[Y] + cE[Z] + d$$

more generally

$$E[a_1X_1 + a_2X_2 + \ldots + a_nX_n] = a_1E[X_1] + \ldots + a_nE[X_n]$$

This is extremely general ! $X_1, \ldots X_n$ do not have to be mutually independent for this to hold!

This generalizes to $E\left[\sum_i a_i \cdot f(X_i)\right] = \sum_i a_i E[f(X_i)]$

day / date:

→ The conditional PMF of random variable X, conditioned on a particular event A.

$$P_{X|A}(x) = P(X=x|A) = \frac{P(\{X=x\} \cap A)}{P(A)}$$

$$\sum_x P(X=x|A) = \sum_x \frac{P(\{X=x\} \cap A)}{P(A)}$$

$$= \frac{\sum_x P(\{X=x\} \cap A)}{P(A)}$$

but A can be written as disjoint union of events $\{X=x\} \cap A$

So $\sum_x P(X=x|A) = 1$.

CONDITIONING ONE RANDOM VARIABLE ON ANOTHER.

→ The conditional PMF of X given Y is given by

$P_{X|Y}(x,y) = P(X=x|\{Y=y\})$

→ Using same set of rules as before we can write:

$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{P_{X,Y}(x,y)}{P_Y(y)}$$

→ for any fixed y such that $P(Y=y) > 0$, we also have: $\sum_x P(X=x|Y=y) = 1$

→ So a conditional pmf satisfies properties of PMF.

* Remember conditional PMF is valid PMF.
* Since $P(X=x \mid Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}$, we also have multiplication rule

→ $P(X=x, Y=y) = P(X=x \mid Y=y)\, P(Y=y)$

→ But $P(X=x, Y=y) = P(Y=y, X=x)$ and so we have:

$$P(X=x, Y=y) = P(Y=y \mid X=x)\, P(X=x)$$

INDEPENDENCE OF RANDOM VARIABLE

→ Consider two events $\{X=x\}$ and A. We know that these two events are independent if $P(\{X=x\}, A) = P(\{X=x\})P(A)$

→ In other words if $P(A)>0$ then $P(X=x \mid A) = P(X=x)$ i.e knowing occurence of A does not change our belief about $\{X=x\}$.

→ We will call the random variable X and event A to be independent if

$$P(X=x, A) = P(X=x)P(A) \text{ for all } x.$$

→ Two random variables are said to be independent if $P(X=x, Y=y) = P(X=x)P(Y=y)$ for all x and y.

→ $P(X=x \mid Y=y) = P(X=x)$ for all x and y such that $P(Y=y)>0$

day / date:

→ we saw that $E[X+Y] = E[X]+E[Y]$ no matter whether X and Y are independent or not.

→ If X and Y are independent $E[XY] = E[X]E[Y]$

→ $E[XY] = \sum_{x,y} xy\, P(X=x, Y=y) = \sum_{x,y} P(X=x)P(Y=y)$

$= \left(\sum_x x P(X=x)\right)\left(\sum_y y P(Y=y)\right) = E[X]E[Y]$

→ let X and Y be independent

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

$E[(X+Y)^2] = E[X^2+Y^2+2XY] = E[X^2] + E[Y^2] + 2E[XY$

$= E[X^2] + E[Y^2] + 2E[X]E[Y]$

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

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

$= \underbrace{E[X^2]-E[X]^2}_{var(X)} + \underbrace{E[Y^2]-E[Y]^2}_{var(Y)} = var(X) + var(Y$

→ variance of sum of independent random variables equals the sum of variances!

→ if three random variables X, Y and Z are said to be independent if $P(X=x, Y=y, Z=z) = P(X=x)P(Y=y)P(Z=z)$

→ if X, Y and Z are independent, then so are $f(X)$, $g(Y)$ and $h(Z)$.

→ for n independent variables we have

$Var(X_1+X_2+X_3+\dots+X_n) = Var(X_1) + Var(X_2) + \dots + Var(X_n)$

day / date:

CONDITIONAL INPENDENCE

→ X and Y are conditionally independent given positive probability event A if.

$$P(X=x, Y=y|A) = P(X=x|A)P(Y=y|A)$$

→ same as saying $P(X=x|Y=y, A) = P(X=x|A)$

MARGINAL PMF

suppose that discrete random variables X and Y have joint pmf $p(x,y)$. Let $x_1, x_2, \ldots, x_i$ and let $y_1, y_2, \ldots, y_j$ denote possible values of Y. The marginal pmf of X and Y are respectively given by following:

$$p_X(x) = \sum_j p(x, y_j)$$

$$p_Y(y) = \sum_i p(x_i, y)$$

CHAPTER 3 :-

day / date:

Continuos Random Variable and PDFs.

→ A random variable is called CONTINUOS if there is nonnegative function f_x called **PROBABILITY DENSITY FUNCTION** of X. or PDF for short,

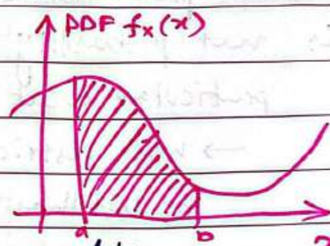
$$P(X \in B) = \int_B f_X(x)\,dx$$

OR

$$f_X(x) = \frac{d F_X(x)}{dx} \quad \text{(Garcia)}$$

→ The probability that X is between a and b is the area under the curve/is in a small interval.

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx$$



PDF Properties

1. $f_X(x) \ge 0$, PDF can never be negative.
2. Total Area under curve must be 1 $\int_{-\infty}^{\infty} f_X(x)\,dx =$
3. If δ is very small then
   $P([x, x+\delta]) \approx f_X(x) \cdot \delta$
4. for any subset B of real line
   $P(X \in B) = \int_B f_X(x)\,dx.$

day / date:

→ for any single value a, we have $P(X=a)=\int_a^a f_X(x)\,dx=0$
So for this reason including/excluding the endpoints of an interval has NO EFFECT on probability.

$$P(a\leq x\leq b)=P(a<X<b)=P(a\leq X<b)=P(a<X\leq b)$$

$$P(a\leq x\leq b)=P(X=a)+P(X=b)+P(a<x<b)$$

→ To interpret the PDF, note that for interval $[x,\ x+\delta]$ with very small length δ, we have

$$P([x,x+\delta])=\int_x^{x+\delta} f_X(t)\,dt \approx f_X(x)\cdot\delta$$

→ we can view $f_X(x)$ as "probability mass per unit length" near x.

* PDF is used to calculate event probabilities, $f_X(x)$ is not probability of any particular event.
  → not restricted to be less than or equal to one.

* PDF's are not probabilities they are densities
  → their units are probability per unit length.

day / date:

Example:- Alvin's driving time to work is b/w 15 and 20 minutes if the day is sunny and b/w 20 and 25 minutes if the day is rainy with all times being equally likely in each case. Assume that a day is sunny with probability 2/3 and rainy with probability 1/3. What is PDF of driving time, viewed as a random variable X?

→ We interpret statement that "all times are equally likely" in sunny & rainy cases to mean that PDF of X is constant in each of interval [15, 20] and [20, 25]. Furthermore, since these two intervals contain all possible driving times, the PDF should be zero everywhere else

$$f_X(x) = \begin{cases} c_1 & \text{if } 15 \leq x < 20 \\ c_2 & \text{if } 20 \leq x \leq 25 \\ 0 & \text{otherwise.} \end{cases}$$

We can determine these constants by using the given probabilities of sunny and rainy day

$$\frac{2}{3} = P(\text{sunny day}) = \int_{15}^{20} f_X(x)\,dx = \int_{15}^{20} c_1\,dx = 5c_1$$

$$\frac{1}{3} = P(\text{rainy day}) = \int_{20}^{25} f_X(x)\,dx = \int_{20}^{25} c_2\,dx = 5c_2$$

so that $c_1 = \frac{2}{15}$, $c_2 = \frac{1}{15}$

day / date:

EXPECTATION / MEAN OF CONTINUOUS RANDOM VARIABLE

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

→ PMF replaced by PDF

→ summation replaced by integral

Properties

① $E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$

② $E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

③ $\mathrm{var}(X) = E[(X-E[X])^2] = \int_{-\infty}^{\infty} (x-E[X])^2 f_X(x)\,dx$

④ $0 \leq \mathrm{var}(X) = E[X^2] - (E[X])^2$

⑤ $Y = aX + b$, where a and b are given scalars then

$$E[Y] = aE[X] + b, \quad \mathrm{var}(Y) = a^2 \mathrm{Var}(X)$$

EXAMPLE

Let X be a uniform random variable over $[a,b]$. What is its expected value?

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

$$f_X(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x > b \end{cases}$$

day / date:

So, $E[X] = \int_{-\infty}^{a} x(0)\,dx + \int_{a}^{b} \frac{x}{b-a}\,dx + \int_{b}^{\infty} x(0)\,dx$

$= \int_{a}^{b} \frac{x}{b-a}\,dx$

$= \left[\frac{x^2}{2(b-a)}\right]_a^b$

$= \frac{1}{2(b-a)}(b^2-a^2) = \frac{(a+b)(b-a)}{2(b-a)} = \frac{a+b}{2}$.

VARIANCE OF CONTINUOS RANDOM VARIABLE

$var[X] = E[X^2] - E[X]^2$

$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - E[X]^2$

→ $var(aX+b) = a^2 var(X)$

Standard Deviation

$\sigma_X = \sqrt{var(X)}$

To obtain variance, we first calculate second moment.

$E[X^2] = \int_a^b \frac{x^2}{b-a}\,dx = \frac{1}{b-a}\int_a^b x^2\,dx$

$= \frac{1}{b-a} \cdot \frac{1}{3} x^3 \Big|_a^b = \frac{b^3-a^3}{3(b-a)}$

$= \frac{a^2+ab+b^2}{3}$

Thus variance obtained by

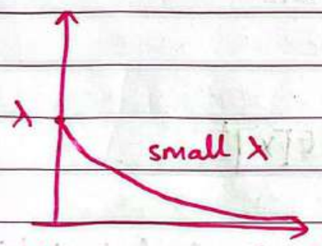
$$\mathrm{var}(X) = E[X^2] - (E[X])^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4}$$

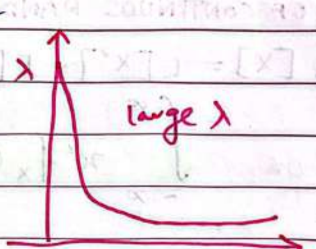
$$= \frac{(b-a)^2}{12}$$

Exponential Random Variable. (parameter $\lambda > 0$)

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

→ PDF form





$$F_X(x) = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big|_0^{\infty} = 1$$

↗ CDF

$$E[X] = \int_0^{\infty} x \cdot \lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}\int_0^{\infty} (\lambda x)\,e^{-\lambda x}\,d(\lambda x)$$

$$= \frac{1}{\lambda}\int_0^{\infty} u e^{-u}\,du = \frac{1}{\lambda}$$

day / date:

$$E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx$$

$$= 0 + \frac{2}{\lambda} E[X] = \frac{2}{\lambda^2}$$

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

$$E[X] = \frac{1}{\lambda} \quad Var(X) = \frac{1}{\lambda^2}$$

$$P(X \geq a) = \int_a^\infty \lambda e^{-\lambda x} dx = e^{-\lambda x} \Big|_a^\infty = e^{-\lambda a}.$$

Example

The time until a small meteorite first lands anywhere in the Sahara desert is modeled as an exponential random variable with mean of 10 days. The time is currently midnight. What is the probability that a meteorite first lands some time between 6 a.m and 6 p.m of first day?

Let X be time elapsed until the event of interest, measured in days. Then X is exponential with mean $1/\lambda = 10$ which yields $\lambda = 1/10$

day / date:

$$P\left(\tfrac{1}{4} \leq X \leq \tfrac{3}{4}\right) = P\left(X \geq \tfrac{1}{4}\right) - P\left(X > \tfrac{3}{4}\right)$$
$$= e^{-1/40} - e^{-3/40} = 0.0476.$$

Cumulative Distribution Frequency (CDF).

$$F_X(x) = P(X \leq x) = \begin{cases} \sum_{k \leq x} P_X(k) & \text{if } X \text{ is dicrete} \\ \int_{-\infty}^{x} f_X(t)\,dt & \text{if } X \text{ continuous} \end{cases}$$

CDF definition:- $F_X(x) = P(X \leq x)$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

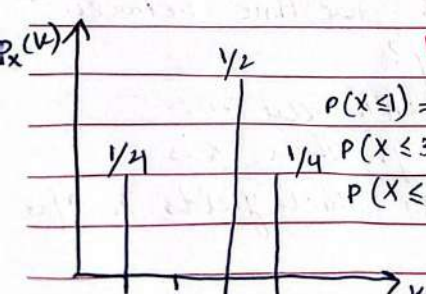
$$\boxed{\frac{d}{dx} F_X(x) = f_X(x)}$$

→ taking derivative of CDF we get PDF

→ taking integration of PDF we get CDF

→ Discrete Random Variables

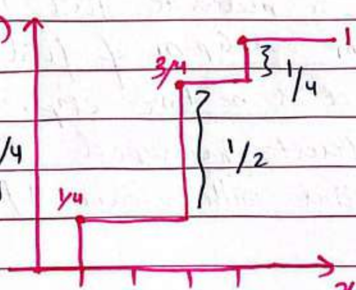
$$F_X(x) = P(X \leq x) = \sum_{k \leq x} P_X(k)$$



$P(X \leq 1) = 1/4$

$P(X \leq 3) = 3/4$

$P(X \leq 4) = 1$



day / date:

General CDF Properties

① $F_X(x) = P(X \le x)$

② F_X is monotonically nondecreasing function
if $x \le y$, then $F_X(x) \le F_X(y)$

③ $F_X(x)$ tends to 0 as $x \to -\infty$ and to 1 as $x \to \infty$

④ If X is discrete, then $F_X(x)$ is piecewise constant function of x.

⑤ If X is continuous, then $F_X(x)$ is continuous function of x.

⑥ If X is discrete and takes integer values the PMF and CDF can be obtained from each other by summing or differencing

$$F_X(k) = \sum_{i=-\infty}^{k} P_X(i) \longrightarrow \text{CDF}$$

$$P_X(k) = P(X \le k) - P(X \le k-1)$$
$$= F_X(k) - F_X(k-1) \longrightarrow \text{PMF.}$$

⑦ If X is continuous, the PDF and CDF can be obtained by

$$\text{CDF} \leftarrow F_X(k) = \int_{-\infty}^{x} f_X(t)\,dt \qquad f_X(x) = \frac{dF_X}{dx}(x) \rightarrow \text{PDF.}$$

day / date:

Geometric & Exponential CDFS

Let X be a geometric random variable with parameter p, that is X is number of trials until first success in sequence of independent. Bernoulli trials, where probability of success at each trial is p. Thus, for $k = 1, 2, \ldots$ we have $P(X=k) = p(1-p)^{k-1}$ and CDF is given by

$$F_{geo}(n) = \sum_{k=1}^{n} p(1-p)^{k-1} = p\frac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n$$

Suppose now ~~for~~ that X is an exponential random variable with parameter $\lambda > 0$. It's CDF is given by

$$F_{exp}(x) = P(X \le x) = 0$$

and

$$F_{exp}(x) = \int_0^x \lambda e^{-\lambda t}\,dt = -e^{-\lambda t}\Big|_0^x = 1-e^{-\lambda x}$$

The values of CDF are equal whenever $x = n\delta$ with $n = 1, 2, \ldots$

mean day / date: variance.

NORMAL RANDOM VARIABLE $X \sim N(\mu, \sigma^2)$

A continuos random variable X is said to be normal or Gaussian if it has a PDF of form

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

* Standard Normal $N(0,1)$: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

$\rightarrow E[X] = 0$

$\rightarrow Var(X) = 1$

* General Normal $N(\mu, \sigma^2)$

$\rightarrow E[X] = \mu$

$\rightarrow Var(X) = \sigma^2$

$$Var(X) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} d$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-y^2/2}\, dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\left(-y e^{-y^2/2}\right)\Big|_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/}$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}\, dy$$

$Var(X) = \sigma^2$

day / date:

* Linear function of Normal Random Variable

→ Let $Y = aX + b$ $\quad X \sim N(\mu, \sigma^2)$

→ What is mean & variance?

→ $E[Y] = a\mu + b$

→ $Var[Y] = a^2 \sigma^2$

STANDARD NORMAL RANDOM VARIABLE

A normal random variable Y with zero mean and unit variance is said to be a standard normal. It's CDF is denoted by Φ

$$\Phi(y) = P(Y \le y) = P(Y < y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2} \, dt$$

Note the table only provides values of $\Phi(y)$ for $y \geq 0$.

e.g. $\Phi(-0.5) = P(Y \le -0.5) = P(Y \geq 0.5)$

$= 1 - P(Y < 0.5)$

$= 1 - \Phi(0.5) = 1 - 0.6915$

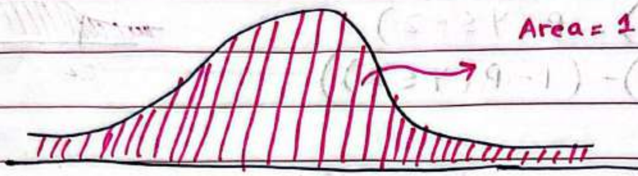
$= 0.3085$

$\Phi(-y) = 1 - \Phi(y)$ for all y.

→ If $X \sim N(\mu, \sigma^2)$ then $Z = \dfrac{x - \mu}{\sigma}$ $\quad (\because Z$ is RV$)$

day / date:

$$\Phi(y) = F_Y(y) = P(Y \le y)$$



Area = 1

→ let $Y = \dfrac{X-\mu}{\sigma}$ (if Y is 3 that means X is 3 standard deviation away from μ)

$$E[Y] = \frac{E[X]-\mu}{\sigma} = 0$$

$$E[Y] = 0$$

$$\mathrm{Var}(Y) = \frac{\mathrm{Var}(X)}{\sigma^2} = 1$$

$$\mathrm{Var}(Y) = 1$$

$$E[Y] = 0, \quad \mathrm{Var}(Y) = 1$$

CALCULATING NORMAL PROBABILITIES

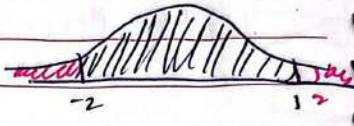
e.g. $X \sim N(6,4)$ $\quad \sigma = 2$

$$\frac{2-6}{2} \le \frac{X-6}{2} \le \frac{8-6}{2}$$

$\frac{2-6}{2}=-2$ $\qquad$ $\frac{8-6}{2}=1$

day / date:

$$P(2 \le X \le 8) = P(-2 \le Y \le 1)$$



$$= P(Y \le 1) - P(Y \le -2)$$

$$= P(Y \le 1) - (1 - P(Y \le 2))$$

Example: The anual snowfall at a particular geographic location is modeled as normal random variable with mean of $\mu = 60$ inches and standard deviation of $\sigma = 20$. What is probability that this years snowfall will be at least 80 inches?

$$Y = \frac{X-\mu}{\sigma} = \frac{X-60}{20}$$

$$P(X \ge 80) = P\left(\frac{X-60}{20} \ge \frac{80-60}{20}\right)$$

$$= P\left(Y \ge \frac{80-60}{20}\right)$$

$$= P(Y \ge 1)$$

$$= 1 - \Phi(1)$$

$$= 1 - 0.8413 = 0.1587$$

day / date:

$$\boxed{P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Y \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)}$$

JOINT PDF

We say that continuous random variables associated with same experiment are jointly continuous and can be described in terms of joint PDF

$$f_{X,Y}(x,y) \ge 0$$

$$P((X,Y) \in B) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$$

→ integrating all over all values of x and y $(x,y) \in B$.

→ If B is of rectangle form $B = \{(x,y) \mid a \le x \le b, c \le y \le d\}$, we have:

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

→ By letting B be two dimensional plane,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

day / date:

→ To interpret joint PDF we let δ be small positive number and consider probability of small rectangle.

$$P(a \leq X \leq a+\delta, c \leq Y \leq c+\delta) = \int_{c}^{c+\delta} \int_{a}^{a+\delta} f_{X,Y}(x,y)\, dx\, dy$$

$$\approx f_{X,Y}(a,c) \cdot \delta^2$$

→ We can get from JOINT PMF of X and Y to MARGINAL PMF of X by summing over Y.

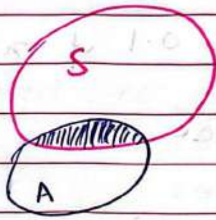
$$P_X(x) = \sum_{y} P_{X,Y}(x,y)$$

→ We can get from JOINT PDF of X and Y to the MARGINAL PDF of X by integrating over Y:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy.$$

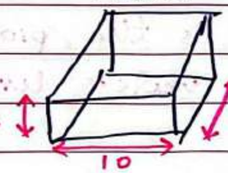
Uniform Joint PDF on set S

$$f_{X,Y} = \begin{cases} \dfrac{1}{\text{area of } S} & \text{if } (x,y) \in S \\ 0 & \text{otherwise.} \end{cases}$$



$$\frac{\text{area}(A \cap S)}{\text{area}(S)} = P(A)$$

Example Anita and Benjamin (Y) both pick a number b/w 0 and 10, according to continuous ~~random variable~~ uniform distribution. What is $f_{X,Y}(x,y)$?



→ we know all pairs (x,y) are equally likely, so we know $f_{X,Y} = c$. It must satisfy $\int_0^{10}\int_0^{10} f_{X,Y}(x,y)\,dx\,dy = 1$

→ So $c \underbrace{\int_0^{10}\int_0^{10} dx\,dy}_{100} = 1$

→ So $c = \dfrac{1}{100} = 0.01$

Example

$\rightarrow f_{x,y}(x,y) = \begin{cases} 0.01 & \text{if } x,y \in [0,10] \\ 0 & \text{otherwise} \end{cases}$

$\rightarrow$ What is $f_x(x)$?

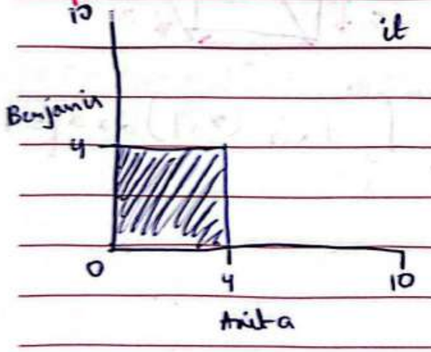
$\rightarrow f_x(x) = \begin{cases} \int_{y=0}^{10} 0.01\,dy = 0.1 & \text{if } x \in [0,10] \\ 0 & \text{otherwise} \end{cases}$

$\rightarrow$ in general we have

$$f_x(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y)\,dy$$

$\rightarrow$ we call $f_x(x)$ the marginal PDF of x.

what is the probability that they both pick numbers less than 4?



it will be $0.01 \int_0^4 \int_0^4 dx\,dy$

$= 0.01 \times 16 = 0.16$

or $\frac{16}{100}$

day / date:

EXPECTED VALUE RULE :-

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y)$$

$$\Downarrow$$

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

JOINT CDF

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

$$= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(s,t)\, dt\, ds$$

Conversely, the PDF can be recovered from CDF by differentiating

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y)$$

Example: Let X and Y be described by uniform PDF on unit square. The joint CDF is given by

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

We then verify that

$$\frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y) = \frac{\partial^2 (xy)}{\partial x \partial y}(x,y) = 1 = f_{X,Y}(x,y)$$

day / date:

More than two Random Variable.

1. $P((X,Y,Z) \in B) = \iiint\limits_{(x,y,z)\in B} f_{X,Y,Z}(x,y,z)\, dx\, dy\, dz$

2. $f_{X,Y}(x,y) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dz$

and

3. $f_X(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dy\, dz$

$$E[g(X,Y,Z)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y,z)\, f_{X,Y,Z}(x,y,z)\, dx\, dy\, dz$$

and if g is linear of form $aX+bY+cZ$ then

$$E[aX+bY+cZ] = aE[X] + bE[Y] + cE[Z]$$

$$E[a_1X_1 + a_2X_2 + \dots + a_nX_n] = a_1E[X_1] + a_2E[X_2] + \dots + a_nE[X_n]$$

day / date:

SUMMARY

→ The JOINT PDF is used to calculate probabilities.

$$P((X,Y) \in B) = \iint_{(x,y)\in B} f_{X,Y}(x,y)\,dx\,dy$$

→ The MARGINAL PDF of X and Y can be obtained from joint PDF using formulas

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

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

→ The JOINT CDF is defined by $F_{X,Y}(x,y) = P(X \le x, Y \le y)$ and determines joint PDF through formula

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y)$$

→ $$E[g(x,y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy.$$

CONDITIONAL PDF, GIVEN AN EVENT

$P_X(x) = P(X=x) \implies f_X(x)\cdot\delta \approx P(\cancel{x} \le X \le x+\delta)$

$P_{X|A}(x) = P(X=x|A) \implies f_{X|A}(x)\cdot\delta \approx P(x \le X \le x + \delta|A)$

$P(X\in B) = \sum_{x\in B} P_X(x) \implies P(X\in B) = \int_B f_X(x)\,dx$

$P(X\in B|A) = \sum_{x\in B} P_{X|A}(x) \implies P(X\in B|A) = \int_B f_{X|A}(x)\,dx$

$\sum_x P_{X|A}(x) = 1 \implies \int f_{X|A}(x)\,dx = 1$

$$P(X\in B|A) = \int_B f_{X|A}(x)\,dx$$

→ By letting B be entire real line, we obtain

$$\int_{-\infty}^{\infty} f_{X|A}(x)\,dx = 1$$

day / date:

$$P(X\in B \mid X\in A) = \frac{P(X\in B, X\in A)}{P(X\in A)} = \frac{\int_{A\cap B} f_X(x)\,dx}{P(X\in A)}$$

$$f_{X\mid\{X\in A\}}(x) = \begin{cases} \dfrac{f_X(x)}{P(X\in A)} & \text{if } x\in A \\ 0 & \text{otherwise.} \end{cases}$$

Example The time T until a new light bulb burns out is an exponential random variable with parameter λ. Ariadne turns the light on, leaves the room and when she returns, t time units later, finds that the light bulb is still on. which corresponds to event $A = \{T > t\}$.
Let X be additional time until the light bulb burns out.
What is conditional CDF of X, given event A?

Solution

$$P(X > x \mid A) = P(T > t+x \mid T > t) = \frac{P(T > t+x \text{ and } T > t)}{P(T > t)}$$

$$= \frac{P(T > t+x)}{P(T > t)} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}$$

day / date:

MEMORYLESS PROPERTY OF EXPONENTIAL.

$X \sim Exp(\lambda)$

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

$\hookrightarrow P(X > s+t \mid X > s) = \dfrac{P(X > s+t, X > s)}{P(X > s)}$

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

$$\Rightarrow e^{-\lambda t}$$

OR

$$P(X > s+t \mid X > s) = \int_{s+t}^{\infty} f_{X \mid X > s}(x)\, dx$$

$$= \lambda \int_{s+t}^{\infty} e^{-\lambda(x-s)}\, dx$$

$$= \lambda \int_{t}^{\infty} e^{-\lambda u}\, du = e^{-\lambda t}$$