Experiments, Models and Probabilities (Chapter 1 RY)

The fun begins!

Many slides courtesy RY (2nd Ed) instructor material

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Defining Our Universe of Interest

- Our universe of interest has to do with our experiment

- What was it for the coin tossing experiment we performed?

- Set theory provides the formalism we need to capture our universe of interest and its inhabitants

Set Theory

- A set is a collection of objects
- A = {Sanjit, Vinayak, Pushpendra, Amarjeet}
  - A is the set of faculty members in the MUC group at IIIT-Delhi
- We will use capital letters to denote sets.

$$A = \{x | x \text{ is an Integer}\}$$
$$B = \{y | y\%2 = 0, \text{ y is an Integer}\}$$
$$B = \{y \text{ is an Integer} | y\%2 = 0\}$$
$$C = \{x^2 | x = 1, 2, 3\}$$
$$D = \{x^2 | x = 1, 2, 3, 4, \ldots\}$$

Set theory

- Sets can be finite or infinite
- Sets can be finite (and hence also countable), countably infinite, or a continuum
  - Set of days in a week
  - Set of students in MTH 201
  - Set of integers
  - Set of rational numbers
  - Set of all real numbers in the interval (0,1)

Set Theory: Definitions and Notation

$A \subset B$ says that A is a subset of B.
If $A \subset B$, then by definition all members of A are also members of B.

$A \supset B$ implies that A is a superset of B.

Two sets A and B are equal, that is $A = B$, *if and only if* $A \subset B$ and $B \subset A$

Examples?

Set Theory

- Universal Set S: It contains *everything*
  - All sets are a subset of S
- Null Set ϕ
  - It contains nothing
  - It is a subset of every set

$$\text{For any set } A,\ \phi \subset A$$

- To say that an element *a* belongs to (is a member of) the set A we write

$$a \in A$$

- We can also say $\{a\} \subset A$

Set Operations

Union: $A \cup B$, $A \cup B \cup C$, $\cup_{i=1}^{\infty} A_i$

$\rightarrow$ $x \in A \cup B$ iff $x \in A$ or $x \in B$

Intersection: $\cap_{i=1}^{\infty} A_i$

$\rightarrow$ $x \in A \cap B$ iff $x \in A$ and $x \in B$

Complement of A is denoted as A^c

$\rightarrow$ $x \in A^c$ iff $x \notin A$

Difference:

$\rightarrow$ $x \in A - B$ iff $x \in A$ and $x \notin B$

Venn Diagrams

Set Theory

- **Mutually Exclusive Sets**

  Sets A_i, $i = 1, 2, \ldots, n$ are mutually exclusive iff $A_i \cap A_j = \phi$ for $i \neq j$

- **Collectively Exhaustive Sets**

  Sets A_i, $i = 1, 2, \ldots, n$ are collectively exhaustive iff $\cup_{i=1}^{n} A_i = S$

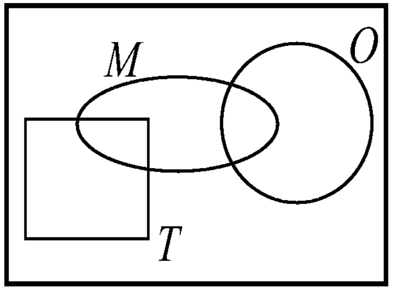
Theorem 1.1

De Morgan's law relates all three basic operations:

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

Quiz 1.1

A pizza at Gerlanda's is either regular (R) or Tuscan (T). In addition, each slice may have mushrooms (M) or onions (O) as described by the Venn diagram at right. For the sets specified below, shade the corresponding region of the Venn diagram.



(1) R

(2) $M \cup O$

(3) $M \cap O$

(4) $R \cup M$

(5) $R \cap M$

(6) $T^c - M$

In-Class Submission

The Story We Will Often Skip But Must Not Be Forgotten

Experiment

Call it the Metro Experiment

Procedure

Stand at a Metro Station and note the time of arrival of two trains that arrive one after the other

Observations

Time difference between two consecutive train arrivals

Model

Probability that the next train arrives within *t* minutes

Examples?

Examples Of Experiments

- Experiment consists of a repeatable procedure
  - Repeats may lead to different observations
- Example Procedures
  - Toss a coin
  - Toss a coin 10 times
- Example Observations
  - Number of heads
  - Sequence of heads and tails
  - Largest number of consecutive heads in the sequence

Outcome

- **Def 1.1:** Outcome
  - An outcome of an experiment is any observation of that experiment
- Outcomes by definition are mutually exclusive
- Heads, Tails are the possible outcomes of a coin tossing experiment
  - They are mutually exclusive, cannot happen at the same time
- Outcomes of the example experiments?

Sample Space

- **Def 1.2:** The sample space S of an experiment is the
  - finest-grain,
  - mutually exclusive,
  - collectively exhaustive set of all possible outcomes.
- Every element of the sample space is an outcome
  - Each of the outcomes are atomic in nature
    - Cannot be split into more than one outcome
  - The outcomes are also mutually exclusive
  - No outcomes are left out of the sample space
- By default, S is your universe of interest. It is a given. *Everything is conditioned on S (super set) having happened*
- In the ten coin toss experiment S = {0,1,2,…,10}

Sample Space

- You want to observe the gender sequence obtained after two people have walked past you
- What is the sample space?
- Possible observations are MM, FF, FM, and MF
- Each one of the above is also an outcome
  - Note MM for your stated observation cannot be split into smaller observations
  - Also, the outcomes are mutually exclusive. Only one can happen at any given time
- S = {MM, FF, FM, MF}

Event and Event Space

- What if I am interested in whether a man was spotted among the first two people?
- No individual outcome is sufficient
- **Def 1.3 Event:** An Event is a set of outcomes
- For our example one event of interest, call it E, is
  - E = {MM, MF, FM}
  - Note that we say that event E occurred if any of the outcomes MM, MF and FM occur.
  - Also, no two outcomes can occur at once
- **Def 1.4 Event Space:** A set of mutually exclusive and collectively exhaustive events is an event space

Example – Experiment Coin Flips (In Class Exercise Submission)

- Procedure: Flip four coins
- Observation: Sequence of heads (h)/tails (t) that is obtained
- Give an example outcome
- What is the Sample space of the experiment?
- How many elements does it contain?
- Let B_i be the event when the sequence contains *i* heads
  - What range of values can *i* take?
  - Are the B_i mutually exclusive?
- Is $B = \{B_0, B_1\}$ an event space?

Why Event Space?

- They maybe easier to handle
- For the above example, sample space contains 16 outcomes, while the event space contains just 5 events
- If the sequence length is increased to 50
  - Sample space has 2^{50} outcomes
  - Event Space has just 51 events, and is a much smaller set
- Clearly the event space does not contain all the information of the sample space
  - However, we may not always be interested in all the information

Partitioning An Event into Mutually Exclusive Events

- **Theorem 1.2**

For an event space $B = \{B_1, B_2, \ldots\}$ and any event A in the sample space, let $C_i = A \cap B_i$. For $i \neq j$, the events C_i and C_j are ME and $A = C_1 \cup C_2 \cup \ldots$.

- A very useful theorem!

Experiment Tweet

- Procedure: You send a tweet over amateur radio using Morse code. Your tweet is 140 characters long.
- A character is received incorrectly with probability *p*.
- Your observation is the number of characters that were received correctly.
  - What is your sample space?
- Let A_k be the event that at least k characters are received correctly
  - Do the A_k together form an event space?
- Can you think of another set of events that does?
- Express one in terms of the other…

Quiz 1.2 In-class exercise

Monitor three consecutive phone calls going through a telephone switching office. Classify each one as a voice call (v) if someone is speaking, or a data call (d) if the call is carrying a modem or fax signal. Your observation is a sequence of three letters (each letter is either v or d). For example, two voice calls followed by one data call corresponds to vvd. Write the elements of the following sets:

(1) $A_1 = \{\text{first call is a voice call}\}$
(2) $B_1 = \{\text{first call is a data call}\}$
(3) $A_2 = \{\text{second call is a voice call}\}$
(4) $B_2 = \{\text{second call is a data call}\}$
(5) $A_3 = \{\text{all calls are the same}\}$
(6) $B_3 = \{\text{voice and data alternate}\}$
(7) $A_4 = \{\text{one or more voice calls}\}$
(8) $B_4 = \{\text{two or more data calls}\}$

For each pair of events A_1 and B_1, A_2 and B_2, and so on, identify whether the pair of events is either mutually exclusive or collectively exhaustive or both.

Problem 1.2.2

An integrated circuit factory has three machines X, Y, and Z. Test one integrated circuit produced by each machine. Either a circuit is acceptable (a) or it fails (f). An observation is a sequence of three test results corresponding to the circuits from machines X, Y, and Z, respectively. For example, aaf is the observation that the circuits from X and Y pass the test and the circuit from Z fails the test.

(a) What are the elements of the sample space of this experiment?

(b) What are the elements of the sets

$$Z_F = \{\text{circuit from } Z \text{ fails}\},$$
$$X_A = \{\text{circuit from } X \text{ is acceptable}\}.$$

(c) Are Z_F and X_A mutually exclusive?

(d) Are Z_F and X_A collectively exhaustive?

(e) What are the elements of the sets

$$C = \{\text{more than one circuit acceptable}\},$$
$$D = \{\text{at least two circuits fail}\}.$$

(f) Are C and D mutually exclusive?

(g) Are C and D collectively exhaustive?

Definitions of Probability

- Classical Definition

  Probability of an event A is given by

  $P[A] = \frac{N_A}{N}$

  - Basically, number of favorable outcomes divided by the total number of possible outcomes
    - Probability that roll of a die leads to an even number…?
  - How about an improvement to the above?
    - The above holds provided that all outcomes are equally likely
- Relative Frequency Definition

$$P[A] = \lim_{n \to \infty} \frac{n_A}{n}$$

  - You perform *n* experiments and observe the event A n_A times to come up with the probability above.

Definitions of Probability

- Axiomatic Definition
  - A few axioms and a few definitions is all you need ☺
  - In this course we will use the Axiomatic definition
  - Once in a while we will go back to the Relative Frequency Definition

Axioms of Probability

- We started with an Experiment
- Experiments involved a procedure, observations and a model
- We started with a model by mapping observations to outcomes, sample space, events, and event space
- Now we want to associate a probability (a number, a measure) with each event in S

Let $P[A]$ denote the probability of event A

Three Axioms of Probability

A1: For any event A, $P[A] \geq 0$

A2: $P[S] = 1$

A3: For any countable collection $A_1, A_2, \ldots$ of ME events $P[A_1 \cup A_2 \cup \ldots] = P[A_1] + P[A_2] + \ldots$

- We assume that the above are true.
- All of the rest follows from these axioms
- This is also called the Axiomatic approach and is a fairly recent approach (starting early 1900(s))

Theorem 1.4

If the events A_i, $i = 1, 2, \ldots, m$ are ME, then $P[A_1 \cup A_2 \cup \ldots \cup A_m] = \sum_{i=1}^{m} P[A_i]$

We will use just the three axioms!

Let $B_1, B_2, \ldots$ be mutually exclusive sets. Axiom $A3$ applies to the sets B_i.

Let $B_i = A_i$ for $1 \leq i \leq m$ and for $i > m$, let $B_i = \phi$

Proof of Theorem 1.4

Therefore, we have $\cup_{i=1}^{m} A_i = \cup_{i=1}^{\infty} B_i$

Also $P[\cup_{i=1}^{m} A_i] = P[\cup_{i=1}^{\infty} B_i]$

$$= \sum_{i=1}^{m} P[A_i] + \sum_{i=m+1}^{\infty} P[\phi]$$

Show that $P[\phi] = 0$ and we are done!

Using Theorem 1.2 from basic set theory,

write $\phi = (\phi \cap B_1) \cup (\phi \cap B_2) \cup \ldots$

and... we are done (How?)

Theorem 1.5

The probability of an event $B = \{s_1, s_2, \ldots, s_m\}$ is the sum of the probabilities of the outcomes contained in the event:

$$P[B] = \sum_{i=1}^{m} P[\{s_i\}].$$

Theorem 1.6

For an experiment with $S = \{s_1, s_2, \ldots, s_n\}$ in which each outcome s_i is equally likely $P[s_i] = 1/n,\ 1 \leq i \leq n$

Proof: Simple. How?

Example – In Class Exercise

Score T is an integer between 0 and 100 corresponding to the outcomes $s_0, \ldots, s_{100}$. A score of 90 to 100 is an A and below 60 is a failing grade of F. Given that all scores between 51 and 100 are equally likely and a score of 50 or less never occurs,
find $P[\{s_{79}\}]$, $P[T \geq 90]$, $P[student\ passes]$.

Theorem 1.8

For any event A, and event space $\{B_1, B_2, \ldots, B_m\}$,

- Think Event Space…
- How are the B(s) related?
- Can I express A in terms of the event space?
- How about the P[A]?

$$P[A] = \sum_{i=1}^{m} P[A \cap B_i]$$

Quiz 1.4

Monitor a phone call. Classify the call as a voice call (V) if someone is speaking, or a data call (D) if the call is carrying a modem or fax signal. Classify the call as long (L) if the call lasts for more than three minutes; otherwise classify the call as brief (B). Based on data collected by the telephone company, we use the following probability model: $P[V] = 0.7$, $P[L] = 0.6$, $P[VL] = 0.35$. Find the following probabilities:

(1) $P[DL]$

(2) $P[D \cup L]$

(3) $P[VB]$

(4) $P[V \cup L]$

(5) $P[V \cup D]$

(6) $P[LB]$

Problem 1.4.6

Suppose a cellular telephone is equally likely to make zero handoffs (H_0), one handoff (H_1), or more than one handoff (H_2). Also, a caller is either on foot (F) with probability $5/12$ or in a vehicle (V).

(a) Given the preceding information, find three ways to fill in the following probability table:

	H_0	H_1	H_2
F			
V			

(b) Suppose we also learn that $1/4$ of all callers are on foot making calls with no handoffs and that $1/6$ of all callers are vehicle users making calls with a single handoff. Given these additional facts, find all possible ways to fill in the table of probabilities.

Conditional Probability

- What is the probability that it is raining right now?
- The number you state is an expression of your belief
- What if I tell you that it is very cloudy outside?
  - Does this change your belief about rain?
- If A is the event {rain} and B is the event {cloudy sky}, then P[A] is your belief in absence of any other information
- P[A|B] is your belief after you are told that B has occurred

Conditional Probability

- $P[A]$ is called the a priori probability of event A
- It is the prior probability of A, prior to you knowing other facts
- $P[A|B]$ is the probability of A given that B has occurred.
  - It is the conditional probability of A given B
- What is $P[A|A]$?
- Roll of a die: What is the probability of the outcome $\{2\}$?
- Let E be the event that an even number was obtained. What is $P[\{2\}|E]$?

Conditional Probability

- **Definition 1.6**

$$P[A|B] = \frac{P[AB]}{P[B]}$$

If A and B are mutually exclusive, then P[AB] = ?

P[A|B] = ?

Theorem 1.9

A conditional probability measure $P[A|B]$ has the following properties that correspond to the axioms of probability.

Axiom 1: $P[A|B] \geq 0$.

Axiom 2: $P[B|B] = 1$.

Axiom 3: If $A = A_1 \cup A_2 \cup \cdots$ with $A_i \cap A_j = \phi$ for $i \neq j$, then

$$P[A|B] = P[A_1|B] + P[A_2|B] + \cdots$$

Example 1.18 Problem

Roll two fair four-sided dice. Let X_1 and X_2 denote the number of dots that appear on die 1 and die 2, respectively. Let A be the event $X_1 \geq 2$. What is $P[A]$? Let B denote the event $X_2 > X_1$. What is $P[B]$? What is $P[A|B]$?

Law of Total Probability (Theorem 1.10)

For an event space $\{B_1, \ldots, B_m\}$ with $P[B_i] > 0$ for all i, $P[A] = \sum_{i=1}^{i=m} P[A|B_i]P[B_i]$

Why this?

Proof...?

Example 1.19 Problem

A company has three machines B_1, B_2, and B_3 for making 1 kΩ resistors. It has been observed that 80% of resistors produced by B_1 are within 50 Ω of the nominal value. Machine B_2 produces 90% of resistors within 50 Ω of the nominal value. The percentage for machine B_3 is 60%. Each hour, machine B_1 produces 3000 resistors, B_2 produces 4000 resistors, and B_3 produces 3000 resistors. All of the resistors are mixed together at random in one bin and packed for shipment. What is the probability that the company ships a resistor that is within 50 Ω of the nominal value?

Bayes' Theorem – The Most Important Theorem Ever!

$$P[B|A] = \frac{P[A|B]P[B]}{P[A]}$$

Proof follows from the definition of conditional probability

Quiz 1.5

Monitor three consecutive phone calls going through a telephone switching office. Classify each one as a voice call (v) if someone is speaking, or a data call (d) if the call is carrying a modem or fax signal. Your observation is a sequence of three letters (each one is either v or d). For example, three voice calls corresponds to vvv. The outcomes vvv and ddd have probability 0.2 whereas each of the other outcomes vvd, vdv, vdd, dvv, dvd, and ddv has probability 0.1. Count the number of voice calls N_V in the three calls you have observed. Consider the four events $N_V = 0$, $N_V = 1$, $N_V = 2$, $N_V = 3$. Describe in words and also calculate the following probabilities:

(1) $P[N_V = 2]$

(2) $P[N_V \geq 1]$

(3) $P[\{vvd\}|N_V = 2]$

(4) $P[\{ddv\}|N_V = 2]$

(5) $P[N_V = 2|N_V \geq 1]$

(6) $P[N_V \geq 1|N_V = 2]$

Problem 1.5.6

Deer ticks can carry both Lyme disease and human granulocytic ehrlichiosis (HGE). In a study of ticks in the Midwest, it was found that 16% carried Lyme disease, 10% had HGE, and that 10% of the ticks that had either Lyme disease or HGE carried both diseases.

(a) What is the probability $P[LH]$ that a tick carries both Lyme disease (L) and HGE (H)?

(b) What is the conditional probability that a tick has HGE given that it has Lyme disease?

Bayes' and Law of Total Probability

For an event space $\{B_1, \ldots, B_m\}$ with $P[B_i] > 0$ for all i,

$$P[B_i|A] = \frac{P[AB_i]}{P[A]} = \frac{P[A|B_i]P[B_i]}{P[A]}$$

$$= \frac{P[A|B_i]P[B_i]}{\sum_{i=1}^{m} P[A|B_i]P[B_i]}$$

We know the a priori probabilities of the B_i(s) and also the probability of an observable event A given B_i. Having seen A (the effect) we want to know the chance that a certain B_i is the cause (that caused A)

Bayes' and Law of Total Probability

- Let A be the event {A world class athlete is seen}
- Consider the event space (World) contains countries of origins. It is W = {China, Jamaica, India, Malaysia, Pakistan, Sri Lanka, UK, US,…}. Let B_1 = China and so on…
- $P[B_i]$ is the probability of the event that the country of origin of a human is B_i
- A priori we could calculate $P[B_i]$ as the ratio of the population of B_i and the population of the world.
- We could also arrive at $P[A|B_i]$, say based on past statistics
- Now say I tell you that I saw an athlete, that is observed A, and I want to know what is probability that the athlete is from India
  - I want to know P[India|A] or $P[B_3|A]$

Independence

- **Two Events** A and B are independent *iff* P[AB] = P[A]P[B]
- The occurrence of event B does not change your belief about whether event A has occurred or not
  - Implies P[A|B] = P[?]
- **Definition 1.9**
  - If n>3, the sets A_1, A_2,…, A_n are independent *if*
    - *Every set of n-1 sets taken from* A_1, A_2,…, A_n is independent,
    - $P[A_1 \cap A_2 \cap \ldots \cap A_n] = P[A_1]\, P[A_2] \ldots P[A_n]$

Quiz 1.6

Monitor two consecutive phone calls going through a telephone switching office. Classify each one as a voice call (v) if someone is speaking, or a data call (d) if the call is carrying a modem or fax signal. Your observation is a sequence of two letters (either v or d). For example, two voice calls corresponds to vv. The two calls are independent and the probability that any one of them is a voice call is 0.8. Denote the identity of call i by C_i. If call i is a voice call, then $C_i = v$; otherwise, $C_i = d$. Count the number of voice calls in the two calls you have observed. N_V is the number of voice calls. Consider the three events $N_V = 0$, $N_V = 1$, $N_V = 2$. Determine whether the following pairs of events are independent:

(1) $\{N_V = 2\}$ and $\{N_V \geq 1\}$

(2) $\{N_V \geq 1\}$ and $\{C_1 = v\}$

(3) $\{C_2 = v\}$ and $\{C_1 = d\}$

(4) $\{C_2 = v\}$ and $\{N_V$ is even$\}$

Problem 1.6.7

For independent events A and B, prove that

(a) A and B^c are independent.

(b) A^c and B are independent.

(c) A^c and B^c are independent.

Sequential Diagrams

Example 1.25 Problem

Suppose traffic engineers have coordinated the timing of two traffic lights to encourage a run of green lights. In particular, the timing was designed so that with probability 0.8 a driver will find the second light to have the same color as the first. Assuming the first light is equally likely to be red or green, what is the probability $P[G_2]$ that the second light is green? Also, what is $P[W]$, the probability that you wait for at least one light? Lastly, what is $P[G_1|R_2]$, the conditional probability of a green first light given a red second light?

Quiz 1.7

In a cellular phone system, a mobile phone must be paged to receive a phone call. However, paging attempts don't always succeed because the mobile phone may not receive the paging signal clearly. Consequently, the system will page a phone up to three times before giving up. If a single paging attempt succeeds with probability 0.8, sketch a probability tree for this experiment and find the probability $P[F]$ that the phone is found.

Problem 1.7.6

A machine produces photo detectors in pairs. Tests show that the first photo detector is acceptable with probability $3/5$. When the first photo detector is acceptable, the second photo detector is acceptable with probability $4/5$. If the first photo detector is defective, the second photo detector is acceptable with probability $2/5$.

(a) What is the probability that exactly one photo detector of a pair is acceptable?

(b) What is the probability that both photo detectors in a pair are defective?

Problem 1.7.9

The quality of each pair of photodiodes produced by the machine in Problem 1.7.6 is independent of the quality of every other pair of diodes.

(a) What is the probability of finding no good diodes in a collection of n pairs produced by the machine?

(b) How many pairs of diodes must the machine produce to reach a probability of 0.99 that there will be at least one acceptable diode?

Section 1.8

Counting Methods

Definition 1.10 Fundamental Principle of Counting

If subexperiment A has n possible outcomes, and subexperiment B has k possible outcomes, then there are nk possible outcomes when you perform both subexperiments.

Example 1.29 Problem

Shuffle a deck and observe each card starting from the top. The outcome of the experiment is an ordered sequence of the 52 cards of the deck. How many possible outcomes are there?

Example 1.30 Problem

Shuffle the deck and choose three cards in order. How many outcomes are there?

Theorem 1.12

The number of k-permutations of n distinguishable objects is

$$(n)_k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

Theorem 1.13

The number of ways to choose k objects out of n distinguishable objects is

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!}.$$

Definition 1.11 n choose k

For an integer $n \geq 0$, we define

$$\binom{n}{k} = \begin{cases} \dfrac{n!}{k!(n-k)!} & k = 0, 1, \ldots, n, \\ 0 & \textit{otherwise.} \end{cases}$$

Example 1.33 Problem

A laptop computer has PCMCIA expansion card slots A and B. Each slot can be filled with either a modem card (m), a SCSI interface (i), or a GPS card (g). From the set $\{m, i, g\}$ of possible cards, what is the set of possible ways to fill the two slots when we sample with replacement? In other words, how many ways can we fill the two card slots when we allow both slots to hold the same type of card?

Theorem 1.14

Given m distinguishable objects, there are m^n ways to choose with replacement an ordered sample of n objects.

Theorem 1.15

For n repetitions of a subexperiment with sample space $S = \{s_0, \ldots, s_{m-1}\}$, there are m^n possible observation sequences.

Example 1.38 Problem

For five subexperiments with sample space $S = \{0, 1\}$, how many observation sequences are there in which 0 appears $n_0 = 2$ times and 1 appears $n_1 = 3$ times?

Theorem 1.16

The number of observation sequences for n subexperiments with sample space $S = \{0, 1\}$ with 0 appearing n_0 times and 1 appearing $n_1 = n - n_0$ times is $\binom{n}{n_1}$.

Theorem 1.17

For n repetitions of a subexperiment with sample space $S = \{s_0, \ldots, s_{m-1}\}$, the number of length $n = n_0 + \cdots + n_{m-1}$ observation sequences with s_i appearing n_i times is

$$\binom{n}{n_0, \ldots, n_{m-1}} = \frac{n!}{n_0!n_1! \cdots n_{m-1}!}.$$

Definition 1.12 Multinomial Coefficient

For an integer $n \geq 0$*, we define*

$$\binom{n}{n_0, \ldots, n_{m-1}} = \begin{cases} \dfrac{n!}{n_0! n_1! \cdots n_{m-1}!} & n_0 + \cdots + n_{m-1} = n; \\ & n_i \in \{0, 1, \ldots, n\}, i = 0, 1, \ldots, m-1, \\ 0 & \text{otherwise.} \end{cases}$$

Quiz 1.8

Consider a binary code with 4 bits (0 or 1) in each code word. An example of a code word is 0110.

(1) How many different code words are there?

(2) How many code words have exactly two zeroes?

(3) How many code words begin with a zero?

(4) In a constant-ratio binary code, each code word has N bits. In every word, M of the N bits are 1 and the other $N - M$ bits are 0. How many different code words are in the code with $N = 8$ and $M = 3$?

Problem 1.8.7

An instant lottery ticket consists of a collection of boxes covered with gray wax. For a subset of the boxes, the gray wax hides a special mark. If a player scratches off the correct number of the marked boxes (and no boxes without the mark), then that ticket is a winner. Design an instant lottery game in which a player scratches five boxes and the probability that a ticket is a winner is approximately 0.01.

Section 1.9

Independent Trials

Example 1.39 Problem

What is the probability $P[S_{2,3}]$ of two failures and three successes in five independent trials with success probability p.

Theorem 1.18

The probability of n_0 failures and n_1 successes in $n = n_0 + n_1$ independent trials is

$$P\left[S_{n_0,n_1}\right] = \binom{n}{n_1}(1-p)^{n-n_1}p^{n_1} = \binom{n}{n_0}(1-p)^{n_0}p^{n-n_0}.$$

Example 1.41 Problem

To communicate one bit of information reliably, cellular phones transmit the same binary symbol five times. Thus the information "zero" is transmitted as 00000 and "one" is 11111. The receiver detects the correct information if three or more binary symbols are received correctly. What is the information error probability $P[E]$, if the binary symbol error probability is $q = 0.1$?

Theorem 1.19

A subexperiment has sample space $S = \{s_0, \ldots, s_{m-1}\}$ with $P[s_i] = p_i$. For $n = n_0 + \cdots + n_{m-1}$ independent trials, the probability of n_i occurences of s_i, $i = 0, 1, \ldots, m-1$, is

$$P\left[S_{n_0,\ldots,n_{m-1}}\right] = \binom{n}{n_0, \ldots, n_{m-1}} p_0^{n_0} \cdots p_{m-1}^{n_{m-1}}.$$

Quiz 1.9

Data packets containing 100 bits are transmitted over a communication link. A transmitted bit is received in error (either a 0 sent is mistaken for a 1, or a 1 sent is mistaken for a 0) with probability $\epsilon = 0.01$, independent of the correctness of any other bit. The packet has been coded in such a way that if three or fewer bits are received in error, then those bits can be corrected. If more than three bits are received in error, then the packet is decoded with errors.

(1) Let $S_{k,100-k}$ denote the event that a received packet has k bits in error and $100-k$ correctly decoded bits. What is $P[S_{k,100-k}]$ for $k = 0, 1, 2, 3$?

(2) Let C denote the event that a packet is decoded correctly. What is $P[C]$?

Problem 1.9.5

There is a collection of field goal kickers, which can be divided into two groups 1 and 2. Group i has $3i$ kickers. On any kick, a kicker from group i will kick a field goal with probability $1/(i+1)$, independent of the outcome of any other kicks by that kicker or any other kicker.

(a) A kicker is selected at random from among all the kickers and attempts one field goal. Let K be the event that a field goal is kicked. Find $P[K]$.

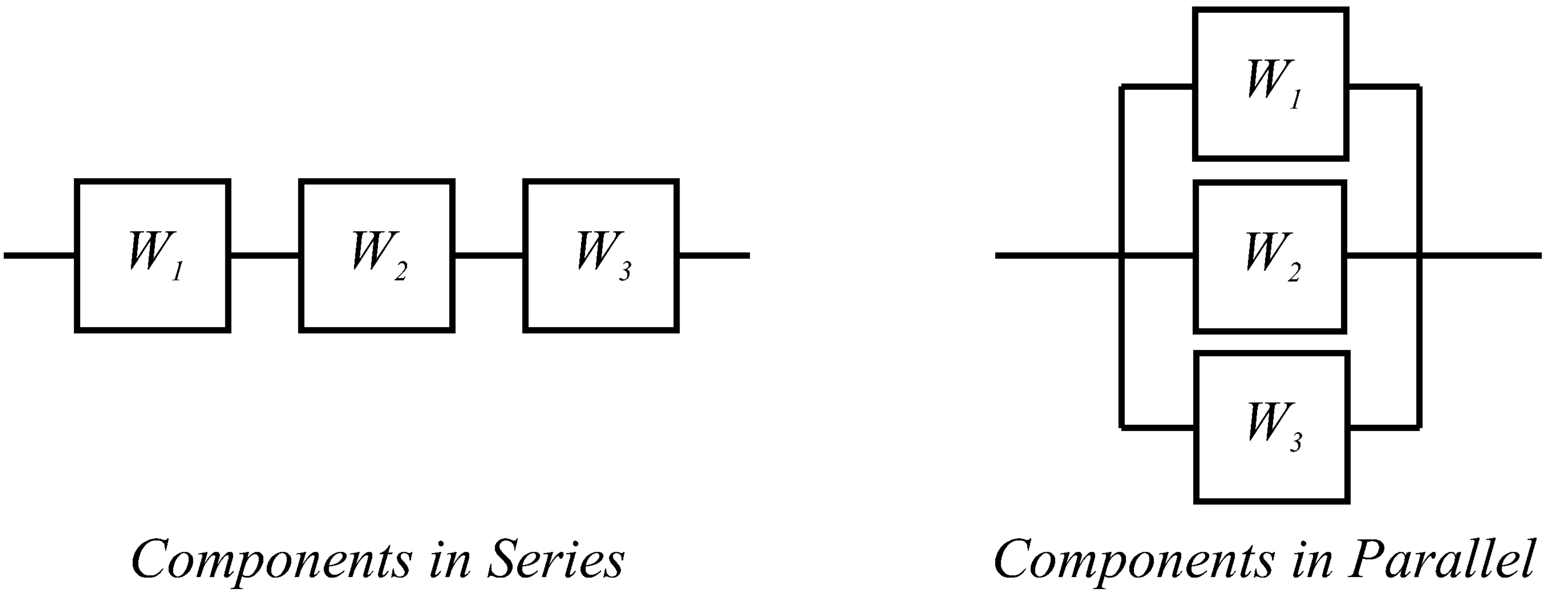
(b) Two kickers are selected at random. For $j = 1, 2$, let K_j be the event that kicker j kicks a field goal. Find $P[K_1 \cap K_2]$. Are K_1 and K_2 independent events?

(c) A kicker is selected at random and attempts 10 field goals. Let M be the number of misses. Find $P[M = 5]$.

Section 1.10

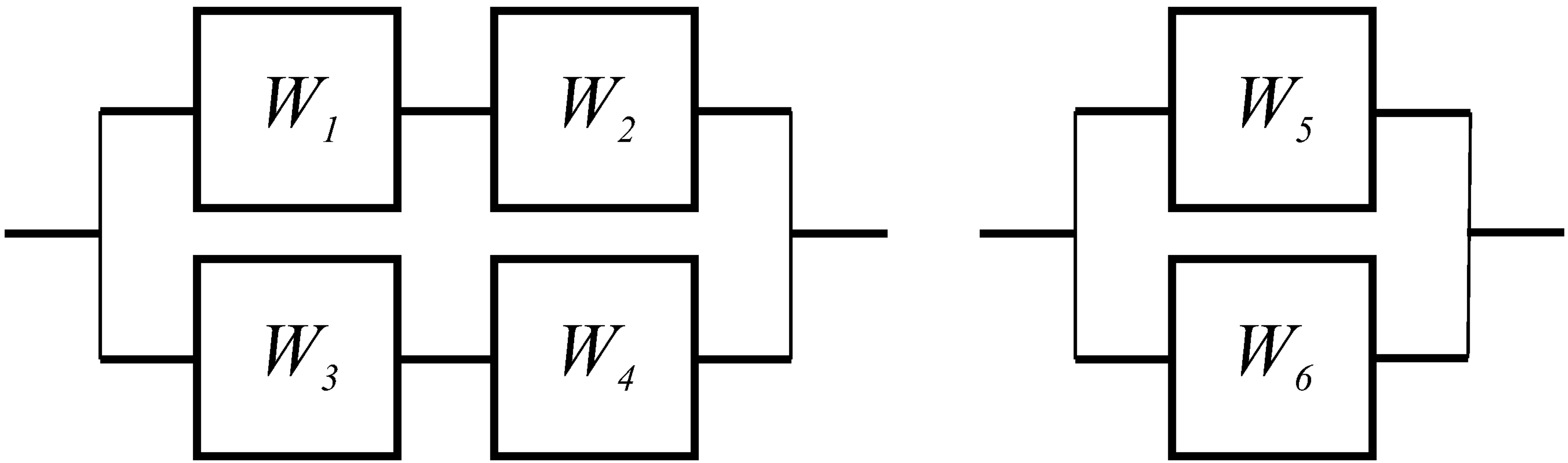
Reliability Problems

Figure 1.3



Serial and parallel devices.

Figure 1.4



The operation described in Example 1.44. On the left is the original operation. On the right is the equivalent operation with each pair of series components replaced with an equivalent component.

Example 1.44 Problem

An operation consists of two redundant parts. The first part has two components in series (W_1 and W_2) and the second part has two components in series (W_3 and W_4). All components succeed with probability $p = 0.9$. Draw a diagram of the operation and calculate the probability that the operation succeeds.

Quiz 1.10

A memory module consists of nine chips. The device is designed with redundancy so that it works even if one of its chips is defective. Each chip contains n transistors and functions properly if all of its transistors work. A transistor works with probability p independent of any other transistor. What is the probability $P[C]$ that a chip works? What is the probability $P[M]$ that the memory module works?

Problem 1.10.4

Consider the device described in Problem 1.10.1. Suppose we can replace any one of the components with an ultrareliable component that has a failure probability of $q/2$. Which component should we replace?

Assignment

- 1.1.1, 1.1.2
- 1.2.1
- 1.3.1 – 1.3.3
- 1.4.3, 1.6.6, 1.5.4
- 1.7.5
- 1.7.10
- 1.9.4
- 1.10.1, 1.10.4
- Always specify the events of interest. Express in terms of union, intersection and etc. of what is given.
- MATLAB: 1.11.3, 1.11.4, 1.11.6 (Online Submission)

MATLAB Example

```
X = rand(75,100) < 0.5;
Y = sum(X,2);
figure; plot(Y/100)
X = rand(75,10) < 0.5;
Y = sum(X,2);
hold on; plot(Y/10)
```

Problem 1.11.6

In this problem, we use a MATLAB simulation to "solve" Problem 1.10.4. Recall that a particular operation has six components. Each component has a failure probability q, independent of any other component. The operation is successful if both

- Components 1, 2, and 3 all work, or component 4 works.
- Either component 5 or component 6 works.

With $q = 0.2$, simulate the replacement of a component with an ultrareliable component. For each replacement of a regular component, perform 100 trials. Are 100 trials sufficient to conclude which component should be replaced?

Discrete Random Variables

Chapter 2 in the book by RY

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

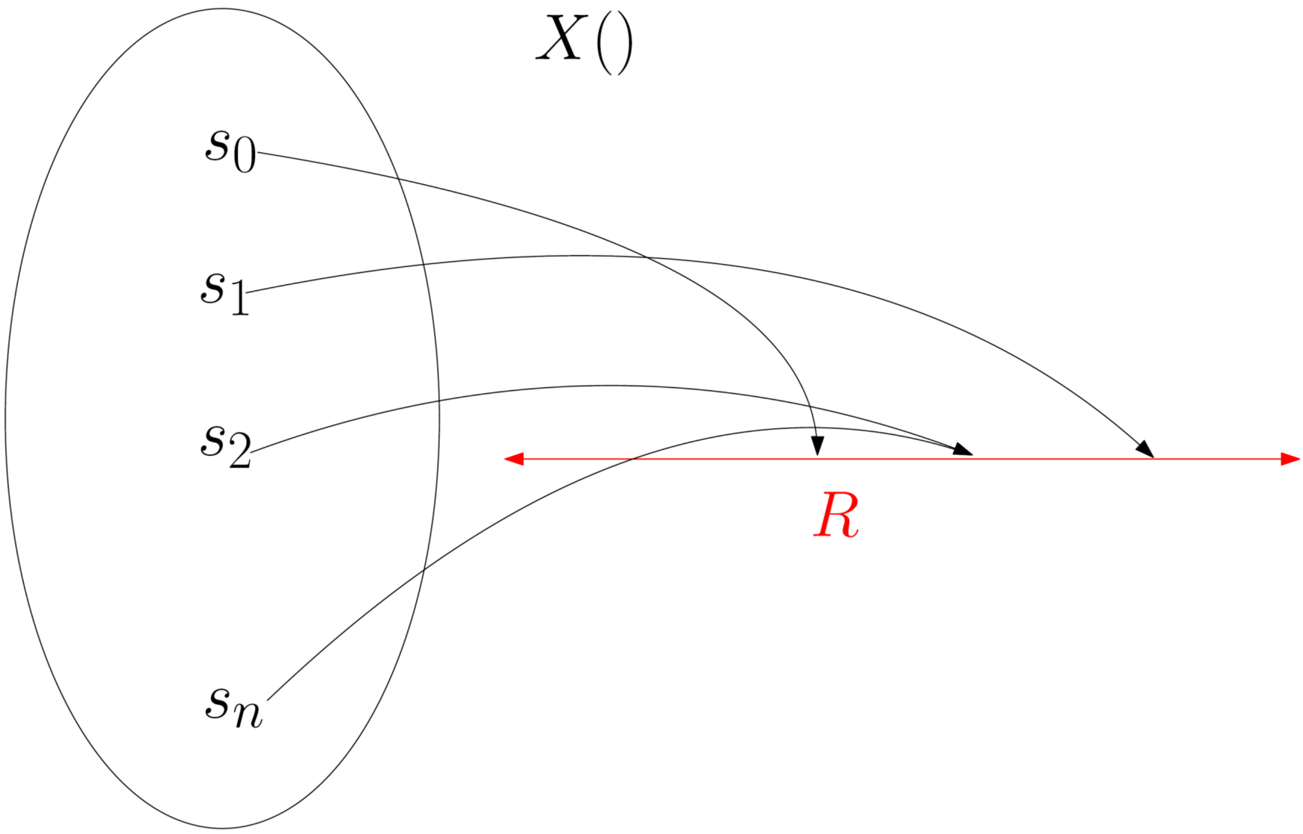
An Example Motivation

- Your book got 50 ratings of Excellent, 30 ratings of Good and 20 ratings of Bad on Flipkart. What is the average rating of your book?
- Can you propose a way out?
  - Assign numbers to Excellent, Good and Bad!
  - Excellent = 10, Good = 5, and Bad = 0 => Average Rating is (500 + 150 + 0)/100 = 6.5, which is better than Good but far from Excellent.
- Exactly what Random Variables do. They map outcomes/events to numbers

Model

- Our Experiment contained
  - Procedure
  - Observation
  - Model
- Under the model we assigned probabilities to outcomes, and by extension events, in the sample space S
- **Now we will assign numbers to outcomes in S**
  - **Each number will occur with probability of the corresponding outcome (or set of outcomes)**
- **A RV is a function that maps outcomes to numbers**

Random Variable



Random Variable

- Coin Tossing Experiment
  - Procedure: Toss a coin once
  - Observation: Heads (1) or Tails (0)
    - Absence/Presence; Pass/Fail; Success/Failure; Accept/Reject; Rich/Poor; Male/Female; Infected/Healthy
    - You are happy to make Boolean observations! That is good enough for your experiment
  - Model: P[0] = 0.5, P[1] = 0.5
  - Our Random Variable, say X(.), takes values of 0 and 1
- **Some Notation**

Random variable X has a range $S_X = \{0, 1\}$.

Range of X is the set of possible values of X

Random variable Y has a range $S_Y \ldots$

Experiment and the Random Variable

- Random variable may be the observation itself
  - You are asked to count the number of buses that pass in 10 minutes.
  - The number of buses belongs to the set of non-negative integers and is a random variable
- Random variable maybe a function of the observation
  - Toss a coin five times
  - Number of heads belongs to the set {0,1,2,3,4,5} and is a random variable.

Number is X and $S_X = \{0, 1, 2, 3, 4, 5\}$

Experiment and the Random Variable

- The RV is a function of another RV
  - Suppose you earn Rs 100 if the number of heads is even and have to pay Rs 100 if the number of heads is odd
  - Let RV G denote your gain at the end of the experiment of 5 coin tosses

$$G = f(X) = \begin{cases} 100 & \text{if } X = 0, 2, 4, \\ -100 & \text{if } X = 1, 3, 5. \end{cases}$$

Pop Q: What is the probability of the event that 0 heads are observed?

Experiment and the Random Variable

- The RV is a function of another RV
  - Suppose you earn Rs 100 for every head and lose Rs 100 for every tail. Your gain G is

$$G = f(X) = 100X - 100(5 - X)$$

Def 2.1: Random Variable

- **Random Variable**: A RV **consists of** an experiment with a probability measure P[.] defined on a sample space S and a **function** that assigns a real number to each outcome in the sample space of the experiment
  - There is an underlying experiment
  - The function defined is just a mapping from outcomes to numbers and is in no way related to a specific experiment
- Notation
  - RV is denoted by a capital letter. The value it assumes is denoted by the corresponding small letter
  - X = x, X is the RV and x is the value
  - Writing X = z is correct too. Here z is the value.
  - X(.) denotes the function corresponding to the RV X

Never Forget the Experiment!

$$\{X = x\} = \{s \in S | X(s) = x\}$$

Reminder that a RV includes a function that maps outcomes to numbers

<u>The event that the RV took the value x in reality is:</u>
The event defined by the set of outcomes in the sample space S that the function X(.) maps to the number x

Discrete Random Variable

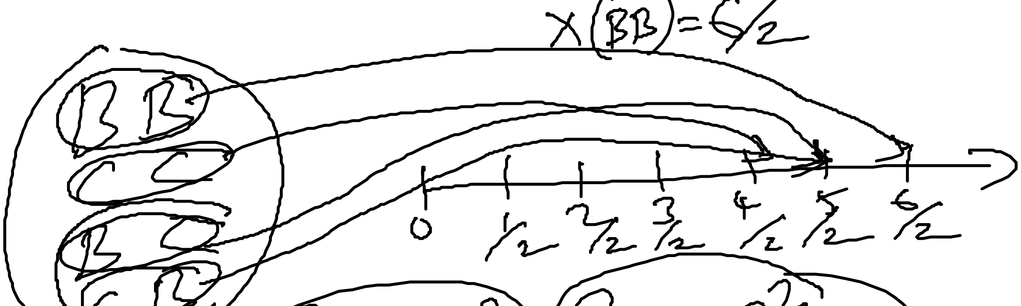
- X is a **discrete** random variable if its range S_X is a countable set

$$S_X = \{x_1, x_2, \ldots\}$$

- $\{X = x_i\}$ and $\{X = x_j\}$, where $x_i \neq x_j$ are disjoint!
- Note that there is no constraint on the values x_1, x_2,…
- The range must be a countable set
  - Finite, Countably infinite
  - Range cannot be the interval (0,1)
  - It can be {0.01, 0.05, 0.08, 0.1,…}
  - It can be {0.001, 0.5, 0.08, 0.9999}
  - It can be the set of integers
- X(.) maps outcomes in sample space S to values in the range S_X

Quiz 2.1

A student takes two courses. In each course, the student will earn a B with probability 0.6 or a C with probability 0.4, independent of the other course. To calculate a grade point average (GPA), a B is worth 3 points and a C is worth 2 points. The student's GPA is the sum of the GPA for each course divided by 2. Make a table of the sample space of the experiment and the corresponding values of the student's GPA, G.



$\{X=2.5\}=\{BC, CB\}$

$\{s \in S \mid X(s)=2.5\}$

Probability Mass Function

- **Def 2.4** The probability mass function (PMF) of a discrete random variable X is

$$P_X(x) = P[X = x]$$

Notation!

Probability of the set of outcomes X=x corresponds to

- Note that the PMF is defined for all *x*, not just the *x* that have a mapping to one or more outcomes
  - **The domain of P_X is the set of real numbers**

- There is nothing sacrosanct about x. Just a convention to use a capital letter for a RV and corresponding small for the value

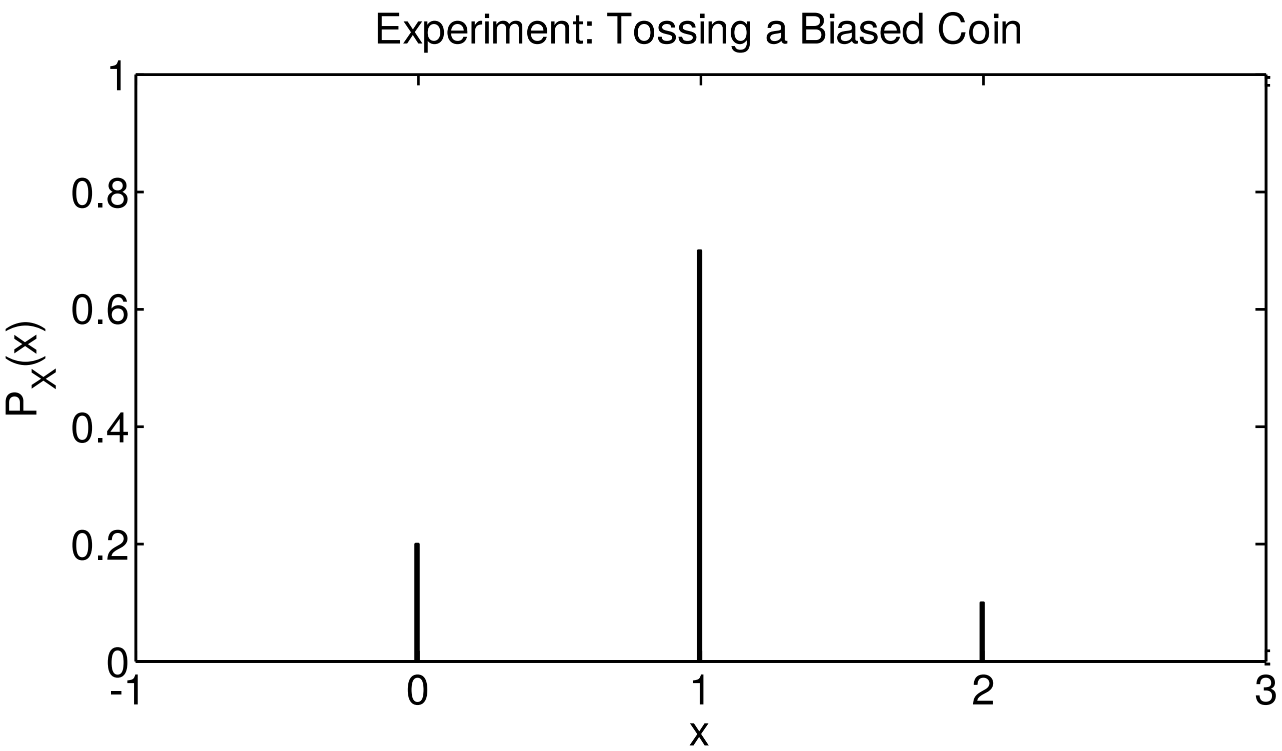
$$P_X(u) = P[X = u]$$

Fickle Coin

- You toss a fickle coin once. The sample space S={heads, tails, standing}. The outcomes heads, tails and standing occur with probabilities 0.7, 0.2 and 0.1 respectively.
- In our model, if tails is observed, then the RV X = 0. For the outcome heads, X = 1. For the outcome standing X=2.
- P[X=0] = 0.7, P[X=1] = 0.2, P[X=2] = 0.1
- The PMF is

$$P_X(x) = \begin{cases} 0.2 & x = 0, \\ 0.7 & x = 1, \\ 0.1 & x = 2, \\ 0 & otherwise. \end{cases}$$

PMF: Tossing a Biased Coin...



PMF

- Is $P_X(\pi)$ defined for the experiment? What is its value?
- **The PMF contains all the information about the RV X**
- Remember that $P_X(x)$ is a probability. Therefore:

$$\text{For any } x,\ P_X(x) \geq 0.$$

- Also

$$\sum_{x \in S_x} P_X(x) = 1$$

- This is because for every outcome s in the sample space S, there is a number x in the range S_X

PMF Properties

For any event $B \subset S_X$, the probability that X is in set B is $P[B] = \sum_{x \in B} P_X(x)$

This is rather straightforward!

Problem 2.2.1

The random variable N has PMF

$$P_N(n) = \begin{cases} c(1/2)^n & n = 0, 1, 2, \\ 0 & \text{otherwise.} \end{cases}$$

(a) What is the value of the constant c?

(b) What is $P[N \leq 1]$?

Problem 2.2.9

When someone presses “SEND” on a cellular phone, the phone attempts to set up a call by transmitting a “SETUP” message to a nearby base station. The phone waits for a response and if none arrives within 0.5 seconds it tries again. If it doesn’t get a response after $n = 6$ tries the phone stops transmitting messages and generates a busy signal.

(a) Draw a tree diagram that describes the call setup procedure.

(b) If all transmissions are independent and the probability is p that a “SETUP” message will get through, what is the PMF of K, the number of messages transmitted in a call attempt? $P[K=6] = (1-p)^6 + (1-p)^5 p$

(c) What is the probability that the phone will generate a busy signal?

(d) As manager of a cellular phone system, you want the probability of a busy signal to be less than 0.02 If $p = 0.9$, what is the minimum value of n necessary to achieve your goal?

$P[K=6 \wedge \text{Busy Failure}] = P[\text{Six Failures}] = (1-p)^n < 0.02$

$(0.1)^n < 0.02$

Families of Random Variables

- We added the RV and its PMF to the model of an experiment
- In the real world many different experiments may be modeled by the same family of random variables!
- Experiment 1: Select a random human being and note the gender. Your observation is either male/female
- Experiment 2: Toss a fair coin. Your observation is either heads/tails

Experiments and Models

- Both experiments can be modeled by the RV X with PMF

$$P_X(x) = \begin{cases} 0.5 & x = 0 \\ 0.5 & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

- The values x=0, x=1 could also correspond to an odd number and even number respectively, on the roll of a die
- The fact that the same models may work for very different experiments is also a good reason to map outcomes to numbers

Bernoulli(p) RV

- The simplest of the many we will see
- Very very useful to model reality
- The parameter p is a probability and can be
  - the probability that a circuit fails a test,
  - probability that a bit is received in error,
  - probability that a coin toss leads to heads and so on…(add your examples)
- **Def 2.5:** X is a Bernoulli(p) RV if the PMF of X has the form

$$P_X(x) = \begin{cases} 1-p & x = 0, \\ p & x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

where $0 < p < 1$.

Geometric RV

- Your experiment involves looking at a sequence of received bits (1 or 0), starting with the first bit of the sequence
- Your observation is the number x, where the x^{th} bit is the very first bit that was received in error
  - Bit 1, Bit 2, …, Bit x-1 are all received correctly
  - Bit x is in error
- Each bit is in error with probability p independent of any other bit that was transmitted

Geometric RV

- Let X be the RV corresponding to outcomes of the experiment
- P[X = x] = P[Event that Bit 1 is correct ∩ Event that Bit 2 is correct ∩ …. ∩ Event that Bit x-1 is correct ∩ Event that Bit x is incorrect]
- $P[X = x] = (1-p)^{(x-1)}\, p$, for $x = 1,2,\ldots$
- It is 0 for all other x.
- Note that **each subexperiment is a Bernoulli trial**

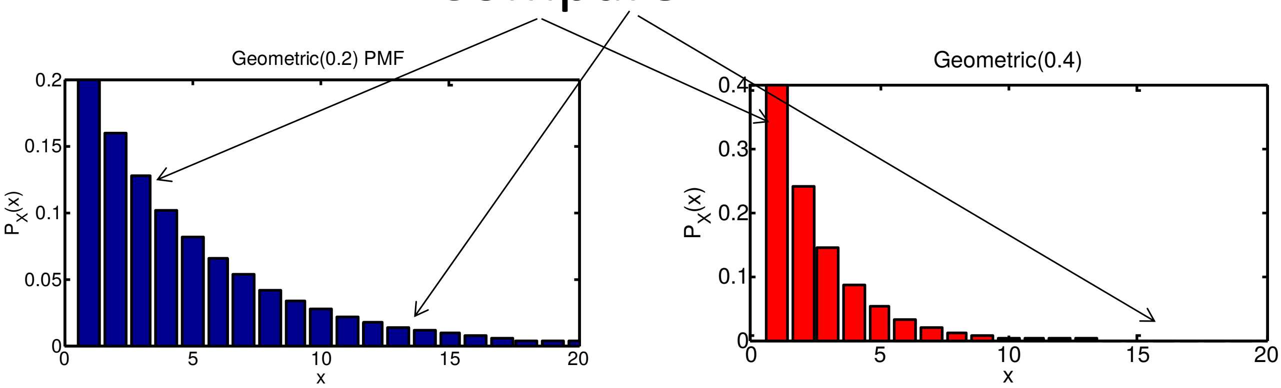
Geometric RV

- **Def 2.6** X is a geometric(p) random variable if the PMF of X has the form

$$P_X(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

Plot the PMF(s)! It will take a minute.
Play around with the parameter *p*

Compare



Binomial(n,p) RV

- You are not just interested in the first bit error
- You are interested in the number of bit errors in a packet of size n bits
- What is the probability that x out of n bits are in error, when each bit is in error with probability p, independent of other bits?
- **Def 2.7:** X is a binomial(n,p) random variable if the PMF of X has the form

$$P_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in \{0, 1, \ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $n \geq 1$ and $0 < p < 1$.

Pascal(k,p) RV

- Example 2.15 and Definition 2.8
- Straightforward extension of Binomial
- Read from book!

Discrete Uniform(k,l) RV

- A fair coin has an equal probability of landing heads up and tails up
- The roll of a fair die has an equal probability of giving each of 1,2,3,4,5 and 6
- If 10 outcomes are possible and you have no reason to believe that any one outcome is more likely than another, you model the outcomes to be equi-probable

Discrete uniform*(k,l)* RV

- **Def 2.9** X is a discrete uniform*(k,l)* RV if the PMF of X has the form

$$P_X(x) = \begin{cases} 1/(l-k+1) & x = k, k+1, \ldots, l, \\ 0 & \text{otherwise.} \end{cases}$$

- Note that bit *k* and *l, k < l,* are integers by the definition of a **discrete** uniform RV.
- Tossing a fair coin can be modeled using a discrete uniform*(100,101)* RV
  - Of course, typically we use discrete uniform*(0,1)*!
- For an experiment that requires rolling a fair die and noting the number
  - Use the discrete uniform(1,6) RV

Poisson (α) Random Variable

- You stand at a railway ticket counter and note the number of customers that arrive at the counter every minute
- Turns out that in many such real world situations
  - Customers arriving at a restaurant
  - Packets arriving at a router
  - Students arriving to a class?

the **number of arrivals in a fixed time interval** can be modeled as a Poisson RV

Poisson (α) Random Variable

- Note that the observation is a count and hence the range S_X is the set of non-negative integers
  - S_X is countably infinite
  - Poisson is indeed a discrete RV
- **Def 2.10** X is a Poisson(α) RV if the PMF of X has the form

$$P_X(x) = \begin{cases} \alpha^x e^{-\alpha}/x! & x = 0, 1, 2, \ldots, \\ 0 & otherwise. \end{cases}$$

where the parameter $\alpha > 0$.

- If the average rate of arrivals is λ/second and time interval is of T seconds, $\alpha = \lambda T$.

Word Problem (Example 2.20)

- The number of database queries processed by a computer in any 10-second interval is a Poisson RV K with α=5 queries.
  - What is the probability that there will be no queries processed in a 10-second interval?
  - What is the probability that at least two queries will be processed in a 2-second interval?

$$P_X(x) = \begin{cases} \alpha^x e^{-\alpha}/x! & x = 0, 1, 2, \ldots, \\ 0 & otherwise. \end{cases}$$

where the parameter $\alpha > 0$. $P[0 \text{ queries in } 10\text{sec}]$

$P[X \geq 2] = 1 - P[X=0 \; X=1]$

$= 1 - \{P[X=0] + P[X=1]\}$ =

Quiz 2.3

Each time a modem transmits one bit, the receiving modem analyzes the signal that arrives and decides whether the transmitted bit is 0 or 1. It makes an error with probability p, independent of whether any other bit is received correctly.

(1) If the transmission continues until the receiving modem makes its first error, what is the PMF of X, the number of bits transmitted?

$P[X=1] = p \qquad P[X=2] = (1-p)p$

$\cancel{P[x=1]}$

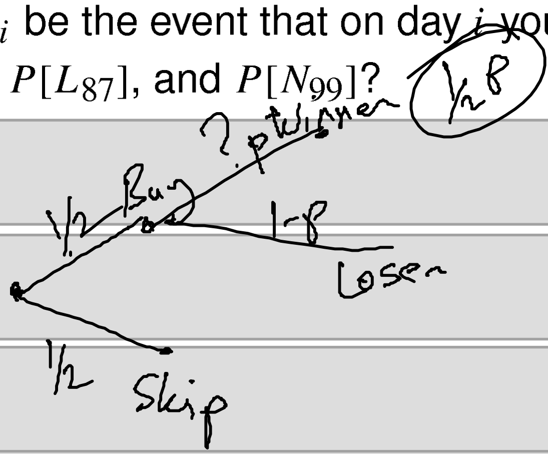
$\vdots$

$$P[X=k] = \begin{cases} (1-p)^{k-1}p & k \in \{1,2,\dots\} \\ 0 & \text{otherwise.} \end{cases}$$

Problem 2.3.12

Suppose each day (starting on day 1) you buy one lottery ticket with probability $1/2$; otherwise, you buy no tickets. A ticket is a winner with probability p independent of the outcome of all other tickets. Let N_i be the event that on day i you do *not* buy a ticket. Let W_i be the event that on day i, you buy a winning ticket. Let L_i be the event that on day i you buy a losing ticket.

(a) What are $P[W_{33}]$, $P[L_{87}]$, and $P[N_{99}]$?



A student attends a lecture with probability 0.3. Every lecture attendance is taken with probability 0.7. A student's attendance is recorded when the student attends the lecture and the attendance is taken. Suppose there are a total of 10 lectures.

P[Attends 7 out of 10, Has attendance of 4 out of 10]

Answer the following questions:

P[You attend & got attended] = (0.3)(0.7)   (0.3)(0.3)

(a) What is the probability that the student attends 7 out of 10 lectures?

$^{10}C_4 \; ^{7}C_3 \; ^{3}C_3 \left((0.3)(0.7)\right)^4 \left((0.3)(0.3)\right)^3 (0.7)^3$

(b) What is the probability that the student attends 7 out of 10 lectures and has an attendance of 4 out of 10?

10th lec was attended & attendance was taken

(c) Suppose you are told that the student attended the 10th lecture and attendance was taken in the lecture. What is your updated belief about the occurrence of the event in part (b)?

(d) Suppose you are told that the student attended the 10th lecture. What is your updated belief about the occurrence of the event in part (b)?

Cumulative Distribution Function (CDF)

- **Def 2.11** The cumulative distribution function of a RV *X* is

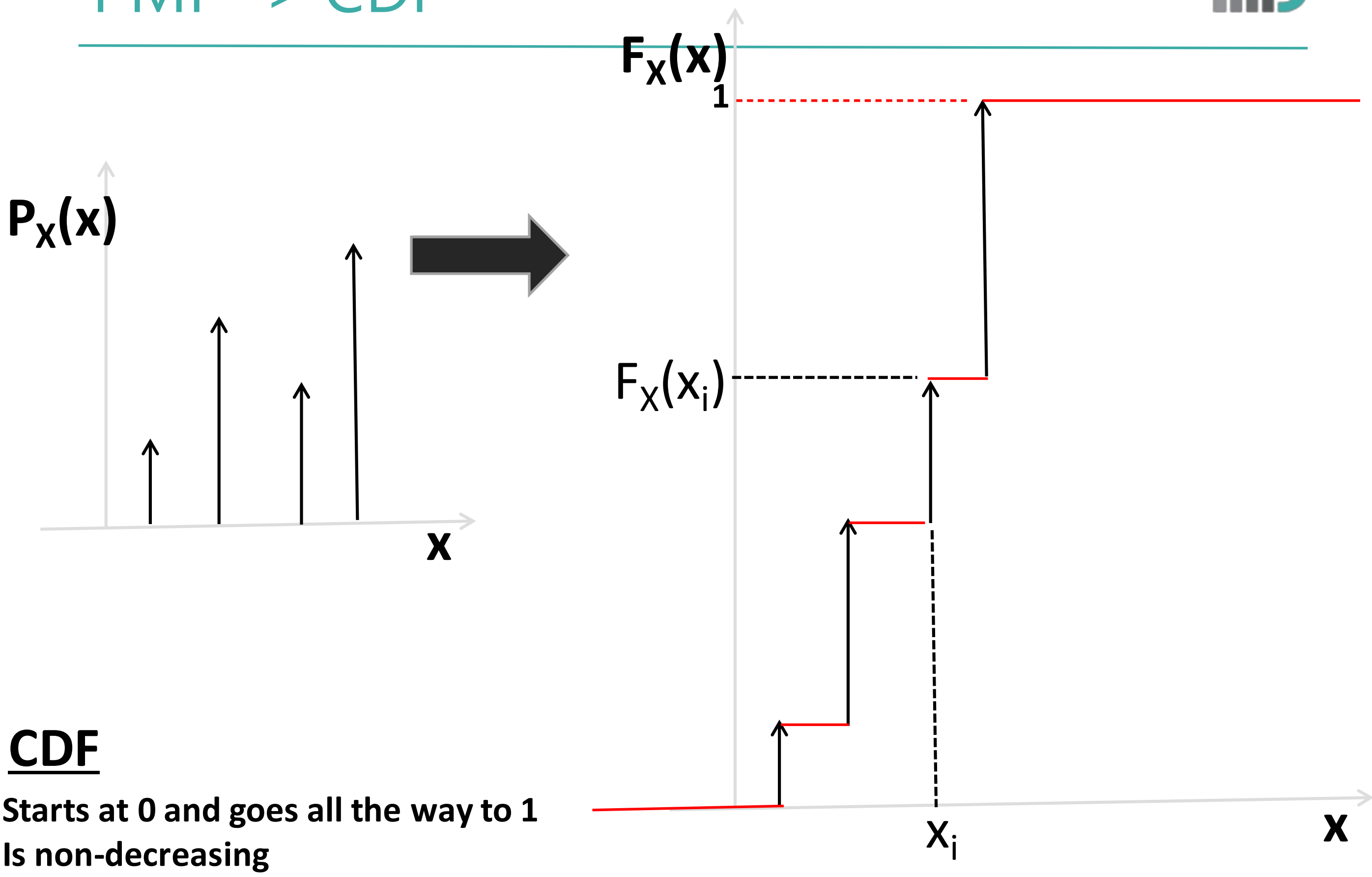
$$F_X(x) = P[X \leq x]$$

- **It is defined for all real *x***
- **Theorem 2.2 (part)**

For any DRV X with range $S_X = \{x_1, x_2, \ldots\}$ satisfying $x_1 \leq x_2 \leq x_3 \ldots$,

$$F_X(-\infty) = 0$$
$$F_X(\infty) = 1$$
$$\text{For all } x' \geq x, F_X(x') \geq F_X(x)$$

PMF -> CDF

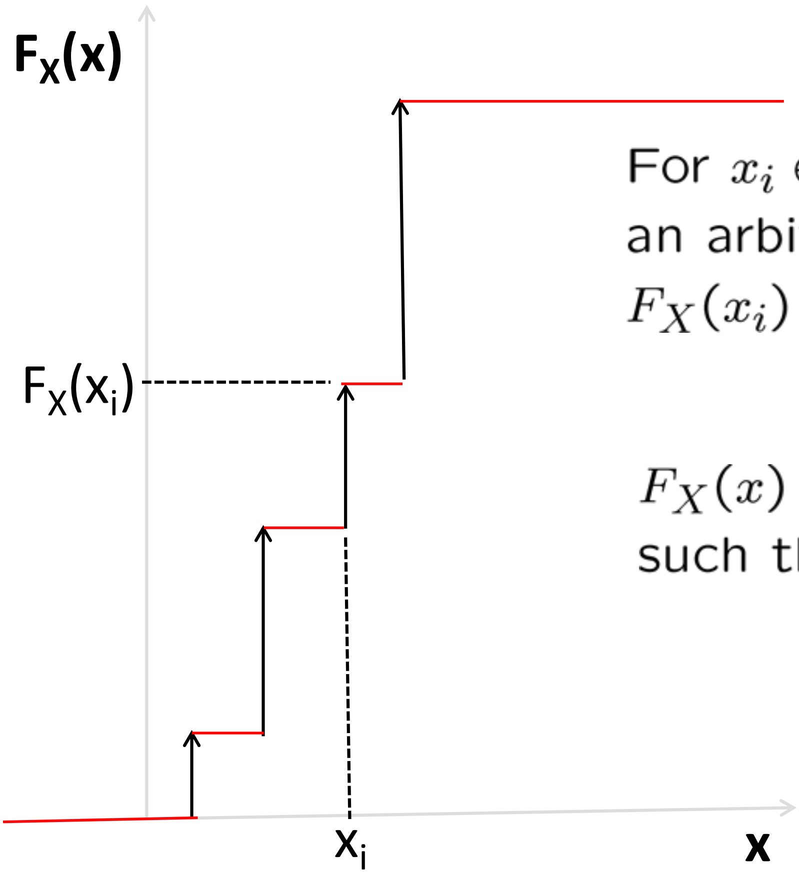


CDF

Starts at 0 and goes all the way to 1
Is non-decreasing
Jumps at points in the range of the RV
Jump at a point x equal to P[X=x]

Accumulate the probabilities in a PMF to get a CDF

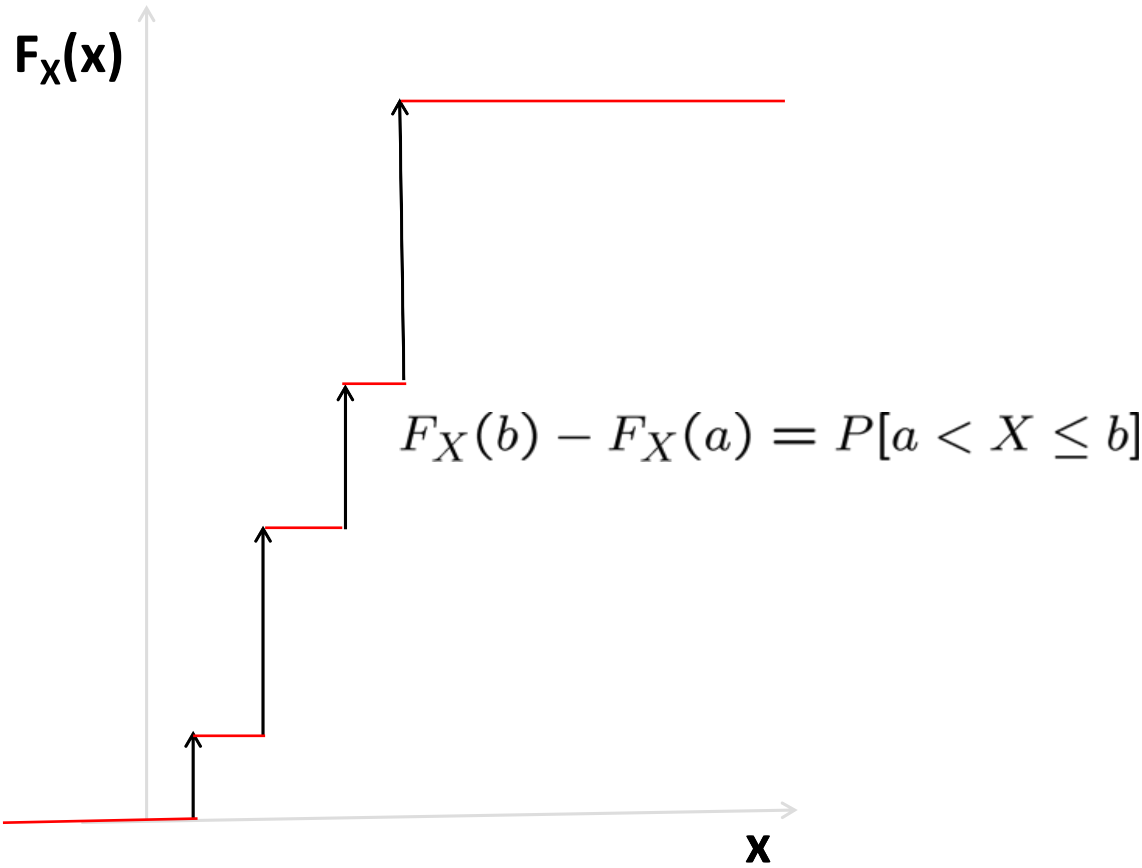
CDF



For $x_i \in S_X$ and ϵ,
an arbitrary small +ve no.,
$F_X(x_i) - F_X(x_i - \epsilon) = P_X(x_i)$

$F_X(x) = F_X(x_i)$ for all x
such that $x_i \leq x < x_{i+1}$.

CDF: Theorem 2.3



CDF

Let $S_X = \{x_{min}, \ldots, x_{max}\}$.

$F_X(x_{max}) = ?$

$F_X(x_{min}) = ?$

For $x > x_{max}$, $F_X(x) = ?$

1.0246

Example 2.24 Problem

In Example 2.11, let the probability that a circuit is rejected equal $p = 1/4$. The PMF of Y, the number of tests up to and including the first reject, is the geometric $(1/4)$ random variable with PMF

$$P_Y(y) = \begin{cases} (1/4)(3/4)^{y-1} & y = 1, 2, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{2.40}$$

What is the CDF of Y?

$$P[Y \le y] = \begin{cases} & y = \{1, 2, \ldots, \} \quad y \ge 1 \\ 0 & y < 1 \end{cases}$$

$$P[Y \le y] = P(Y=1) + P[Y=2] + \cdots + P[Y=y]$$

$$= \sum_{k=1}^{y} P[Y=k]$$

Average of a Discrete Random Variable

- Mode (Def 2.12 – read from book)
- Median (Def 2.13 – read from book)
- Expected Value
  - When someone says average of a random variable, they usually mean the expected value of the RV!
- **Def 2.14** The expected value of X is

$Y = E[X]$

$P[Y = y]$

$$E[X] = \mu_X = \sum_{x \in S_X} x P_X(x).$$

Note the Notation!

Is E[X] a random variable?

Expected Value

- Consider n trials of an experiment.
- We obtain n values x(1), x(2), …, x(n) that the RV X (part of the model) takes during the trials
- The empirical average is

$$m_n = \frac{1}{n} \sum_{i=1}^{n} x(i)$$

- For a range S_X, we can rewrite the above as

$$m_n = \frac{1}{n} \sum_{x \in S_X} N_x x$$

Where N_x is the number of times the value x in S_X occurs during the n trials

Expected Value

$$m_n = \frac{1}{n} \sum_{x \in S_X} N_x x$$

As $n \to \infty$, $N_x/n \to P_X(x)$,
and we say that $m_n \to E[X]$.

We used the relative frequency definition of probability to link empirical averages obtained from multiple experiment trials with the expected value of the RV X.

Some Examples of E[X]

- **Def 2.5:** X is a Bernoulli(p) RV if the PMF of X has the form

$$P_X(x) = \begin{cases} 1-p & x=0, \\ p & x=1, \\ 0 & \text{otherwise.} \end{cases}$$

where $0 < p < 1$.

- E[X] = ?
  - p

Some Examples of E[X]

- Geometric(p)

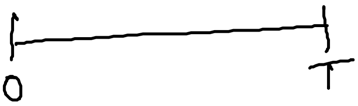
$$P_X(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases}$$

$$\begin{aligned} E[X] &= \sum_{x=1}^{\infty} x P_X(x) \\ &= \sum_{x=1}^{\infty} x p (1-p)^{x-1} \\ &= p \sum_{x=1}^{\infty} x q^{x-1} \\ &= \frac{1}{p}. \end{aligned}$$

We define q=1-p

E[X] Examples

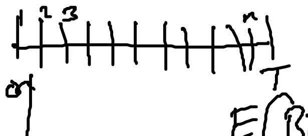
- When X is Geometric(p)
  - E[X] = 1/p
- A 0.1 probability of a faulty device =>
  - On an average the 10^{th} device will be the first to be found faulty
  - This does not mean that every 10^{th} device will be faulty. In fact P[X=10] = 0.03!
- Please see Theorem 2.6 and 2.7 for E[.] of other RVs



λT

$\boxed{np = \lambda T}$

$p = \left(\frac{\lambda T}{n}\right)$



Bernoulli(p)

$E[\text{Binom}(n,p)] = np$

Theorem 2.8

Perform n Bernoulli trials. In each trial, let the probability of success be α/n, where $\alpha > 0$ is a constant and $n > \alpha$. Let the random variable K_n be the number of successes in the n trials. As $n \to \infty$, $P_{K_n}(k)$ converges to the PMF of a Poisson (α) random variable.

$$P[K=k] = {}^nC_k\, p^k (1-p)^{n-k} = {}^nC_k \left(\frac{\lambda T}{n}\right)^k \left(1-\frac{\lambda T}{n}\right)^{n-k}$$

$$\operatorname{Lt}_{n\to\infty} \frac{\lfloor n}{\lfloor k \,\lfloor n-k} \frac{(\lambda T)^k}{n^k} \left(1-\frac{\lambda T}{n}\right)^{n-k}$$

$$\operatorname{Lt}_{n\to\infty} \frac{1}{\lfloor k} \frac{n(n-1)\cdots(n-k+1)}{n^k} (\lambda T)^k \left(1-\frac{\lambda T}{n}\right)^{n-k}$$

$$\to 1 \qquad \frac{\left(1-\frac{\lambda T}{n}\right)^n}{\cancel{\left(1-\frac{\lambda T}{n}\right)^k}}$$

$$\frac{(\lambda T)^k e^{-\lambda T}}{\lfloor k}$$

Problem 2.5.10

Let binomial random variable X_n denote the number of successes in n Bernoulli trials with success probability p. Prove that $E[X_n] = np$. Hint: Use the fact that $\boxed{\sum_{x=0}^{n-1} P_{X_{n-1}}(x) = 1.}$

$$E[X_n] = \sum_{x=0}^{n} P[X_n = x]$$

Functions of a Random Variable

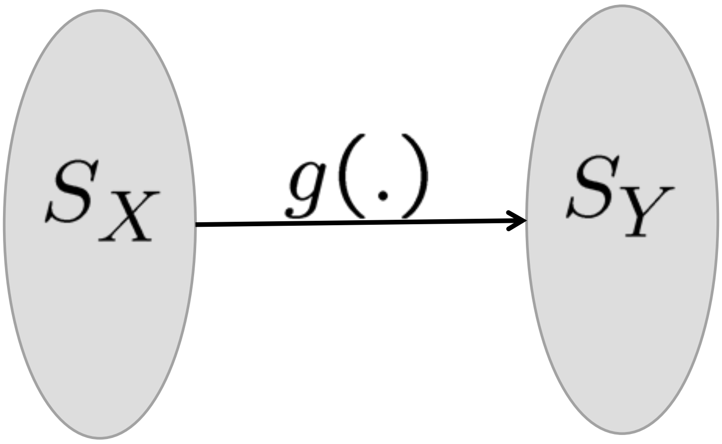
- You are measuring the power consumed by your appliance
- The power observed is a random variable, say W and is in watts
  - The randomness in the observed power can be due to measurement errors introduced by your equipment
- The power in dB, Y = 10 * log10(W) is also a random variable and is a function of the RV W.
  - We say that Y is derived from W

$Y = f(X)$ $\quad P_X(x)$

$E[Y]$ $\quad F_X(x)$

Derived Random Variable

- **Def 2.15** Each sample value *y* of a derived random variable Y is a **mathematical function** g(x) of a sample value *x* of another random variable X.
  - The notation Y=g(X) is used to describe the relationship of the two random variables.



<u>Mathematical function:</u>
One-to-one or a many-to-one relation

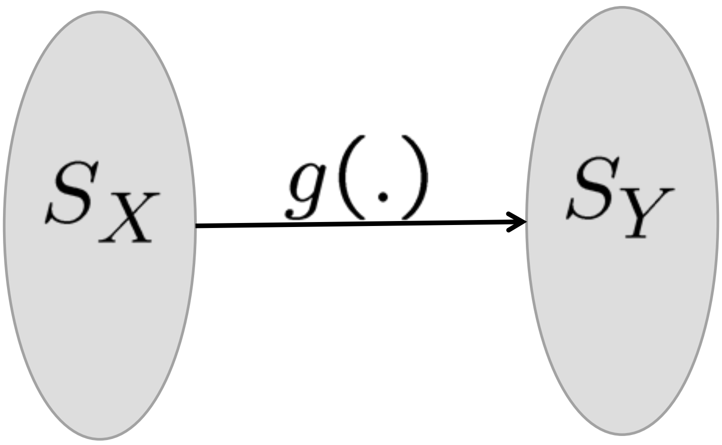
g(.) maps the range of X into the range of Y

$P_X(x)$ -> $P_Y(y)$

- **Theorem 2.9** For a discrete random variable X, the PMF of Y=g(X) is

$$P_Y(y) = \sum_{x:g(x)=y} P_X(x)$$

Remember this is same as P[X=x]



Derived Random Variable

$\sum_{y \in S_Y} \left[y \sum_{\{x : g(x) = y\}} P[X=x] \right]$

- We can get $P_Y(y)$ from $P_X(x)$ and can then get the $E[Y]$ using $P_Y(y)$
  - Note that $P_Y(y)$ is a complete description of Y $\sum_{\{x : g(x) = y\}}$
- What if we do not want to calculate $P_Y(y)$? What can we do with $P_X(x)$? $\sum_{x \in S_X} g(x)$
- **Theorem 2.10** Given a random variable X with PMF $P_X(x)$ and the derived random variable Y=g(X), the expected value of Y is

$$E[Y] = \mu_Y = \sum_{x \in S_X} g(x) P_X(x).$$

$$E[Y] = \sum_{y \in S_Y} y P[Y=y]$$

Derived Random Variable

$$E[Y] = \mu_Y = \sum_{y \in S_Y} y P_Y(y).$$

$$E[Y] = \sum_{y \in S_Y} y \sum_{x : g(x) = y} P_X(x)$$

$$E[Y] = \sum_{y \in S_Y} \sum_{x : g(x) = y} g(x) P_X(x)$$

$$E[Y] = \sum_{x \in S_X} g(x) P_X(x)$$

What we want!

$$Y = X - E[X] \qquad Y = g(X) = X - E[X]$$

$$E[Y] = \sum_{y \in S_Y} y\, P[Y=y] = \sum_{x \in S_X} g(x)\, P[X=x]$$

$$= \sum_{x \in S_X} (x - \mu_X)\, P[X=x]$$

$$= \sum_{x \in S_X} x\, P[X=x] - \sum_{x \in S_X} \mu_X\, \boxed{P[X=x]}$$

$$\underbrace{E[X]}_{\mu_X} - \mu_X = 0$$

More on the E[.] Operator

- Consider the RV $Y = g(X) = X - E[X] = X - \mu_X$
- The expected value of X is subtracted from each number in S_X. The resulting range is the range S_Y of the random variable Y
- What is E[Y]?
- **Theorem 2.11** For any RV X,

$$E[X - \mu_X] = 0.$$

$$\begin{aligned} E[X - \mu_X] &= \sum_{x \in S_X} (x - \mu_X) P_X(x) \\ &= \sum_{x \in S_X} x P_X(x) - \mu_X \sum_{x \in S_X} P_X(x) \\ &= \mu_X - \mu_X = 0. \end{aligned}$$

The E[.] Operator

- **Theorem 2.12** For any random variable X,

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

- Proof is similar to the previous theorem

$g(X) = aX + b.$ $\quad E[g(X)] = \sum_{x \in S_X} g(x)\, P[X=x]$

$= \sum_{x \in S_X} (ax+b)\, P[X=x]$

The Properties of the E[.] Operator

- E[.] is a linear operator
- Every linear operator has two properties:

$$\hat{A}(f+g) = \hat{A}f + \hat{A}g \qquad (1)$$

$$\hat{A}(cf) = c\hat{A}f \qquad (2)$$

where $\hat{A}$ is the operator, c is a scalar, f and g are functions.

See http://vergil.chemistry.gatech.edu/notes/quantrev/node14.html

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

$b = g(X)$

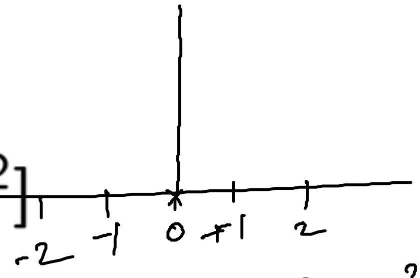
$\sum_{x \in S_X} b\, P(X=x)$

- Scalar Multiplication of a Random Variable
  - $E[aX] = a\,E[X]$
  - $E[aX^2] = a\,E[X^2]$ (Note that X^2 is a RV)
- $E[X + X^2] = E[X] + E[X^2]$
- $E[(X + c)^2] = E[X^2 + c^2 + 2cX] = E[X^2] + E[c^2] + 2E[cX]$
  $= E[X^2] + c^2 + 2cE[X]$

Variance

- **Def 2.16** Variance of X is

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



Think of the relative frequency interpretation of variance (which is also the empirical estimate)

$S_X = \{-1, 0, 1\}$

$S_Y = \{-2, -1, 0, 1, 2\}$

- Standard deviation is just the square-root of Variance
  - Values of the RV within the SD of the Expected Value are referred to as its *typical values*

Properties of Variance

- **Theorem 2.13**

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

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

$$Var[aX+b] = ?$$

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

$E[g(X)]$

- **Theorem 2.14**

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

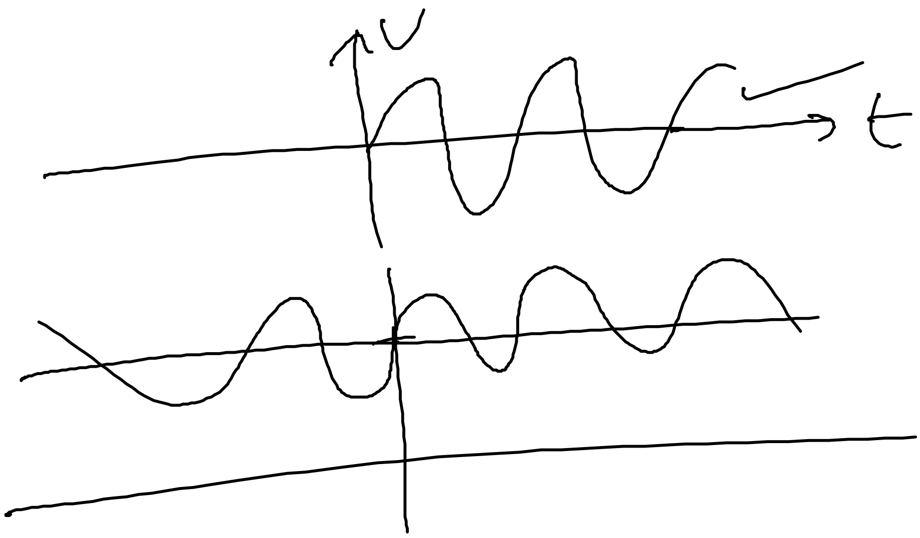
$$Var[aX+b] = a^2 Var[X]$$

  - Offsetting a RV by a constant does not change its variance!

•Show both results using the properties of the E[.] operator
•Also, show using first principles, that is by expanding E[g(x)]

Moments

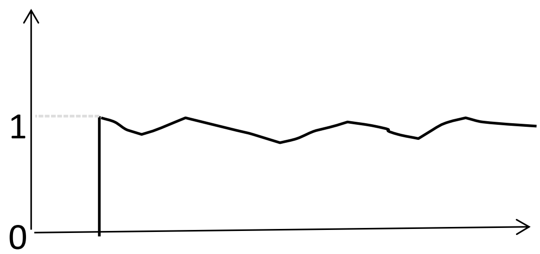
- For random variable X
  - The n^{th} moment is $E[X^n]$
  - The n^{th} central moment is $E[(X - E[X])^n]$

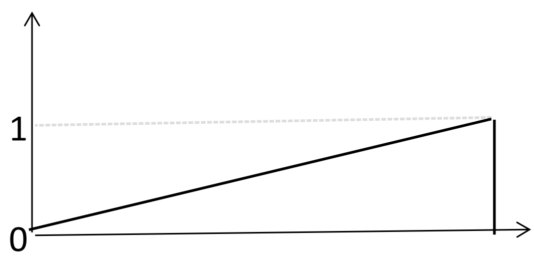


VVS Word Problems

- X is a discrete uniform RV with range S_X = {1,2,3,4,5,…,100}. Find E[X] and E[X^2].

- Which are valid CDFs?





VVS Word Problems

- Y = X + 10
- Express Var[Y] in terms of Var[X]

- Y = 10X + 10
- Express Var[Y] in terms of Var[X]

Quiz 2.6

Monitor three phone calls and observe whether each one is a voice call or a data call. The random variable N is the number of voice calls. Assume N has PMF

$$P_N(n) = \begin{cases} 0.1 & n = 0, \\ 0.3 & n = 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases} \tag{2.75}$$

Voice calls cost 25 cents each and data calls cost 40 cents each. T cents is the cost of the three telephone calls monitored in the experiment.

(1) Express T as a function of N.

(2) Find $P_T(t)$ and $E[T]$.

Quiz 2.6

Monitor three phone calls and observe whether each one is a voice call or a data call. The random variable N is the number of voice calls. Assume N has PMF

$S_N \rightarrow S_T$

$$P_N(n) = \begin{cases} 0.1 & n = 0, \\ 0.3 & n = 1, 2, 3, \\ 0 & \text{otherwise.} \end{cases} \tag{2.75}$$

Voice calls cost 25 cents each and data calls cost 40 cents each. T cents is the cost of the three telephone calls monitored in the experiment.

N voice and 3-N data calls.

(1) Express T as a function of N. $T = 25N + 40(3-N)$

$S_T = \{120, 105, 90, 75\}$

(2) Find $P_T(t)$ and $E[T]$. $P[T = 75] = P[N = 3]$

S_X $\quad P_X(x)$

$$P[X=x \mid B] = P_{X|B}(x) = \frac{P[(X=x, B)]}{P[B]}$$

μ

$\{X=x\} \subset B \qquad x \in B$

$x \notin B$

$$P[X=x \mid B] = \begin{cases} \dfrac{P_X(x)}{P[B]} & x \in B \\ 0 & x \notin B \end{cases}$$

Conditional PMF of a Discrete RV

- The PMF of a RV X gives the probability of the event {X=x}
- PMF $P_X(x) = P[X=x]$
- The PMF of a RV conditioned on an event B is $P[X=x|B]$
- $P[X=x|B]$ = ?
- **Def 2.19** Given the event B, with $P[B] > 0$, the conditional probability mass function of X is $P_{X|B}(x) = P[X=x|B]$.

Notation!

Theorem 2.16

- A RV X resulting from an experiment with event space $B_1, B_2, \ldots, B_m$ has PMF __________?
- What properties do $B_1, B_2, \ldots, B_m$ satisfy?

$$\underline{P_X(x)} = \sum_{i=1}^{m} P_{X|B_i}(x) P[B_i].$$

$X = x$

$$\{X=x\} = \bigcup_{i=1}^{n} \{X = x \cap B_i\}$$

$$P[X=x] = P_X(x) = \sum_{i=1}^{n} \underbrace{P[X=x, B_i]}$$

VS Word Problem (RY)

$P[X=x|H]=$

$P[X=x|R]$

Example 2.38 Problem

$X|H \sim Geom(0.1)$

Let X denote the number of additional years that a randomly chosen 70 year old person will live. If the person has high blood pressure, denoted as event H, then X is a geometric $(p = 0.1)$ random variable. Otherwise, if the person's blood pressure is regular, event R, then X has a geometric $(p = 0.05)$ PMF with parameter. Find the conditional PMFs $P_{X|H}(x)$ and $P_{X|R}(x)$.

$X|R \sim Geom(0.05)$

$$P[X = x|H] = \begin{cases} (0.9)^{x-1}(0.1) & x = 1, 2, \ldots \\ 0 & otherwise \end{cases}$$

- Similarly you can find P[X = x|R]

$X|H \sim Geom(0.1)$

$X|R \sim Geom(0.05)$

VS Word Problem (RY)

$P[H] = 0.4$

Example 2.38 Problem

$P[\{X=x\}] = P[\{X=x \cap H\} \cup \{X=x \cap R\}]$

$= P[X=x, H] + P[X=x, R]$

Let X denote the number of additional years that a randomly chosen 70 year old person will live. If the person has high blood pressure, denoted as event H, then X is a geometric $(p = 0.1)$ random variable. Otherwise, if the person's blood pressure is regular, event R, then X has a geometric $(p = 0.05)$ PMF with parameter. Find the conditional PMFs $P_{X|H}(x)$ and $P_{X|R}(x)$. If 40 percent of all seventy year olds have high blood pressure, what is the PMF of X?

- The last sentence implies that P[H] = 0.4.
- Note that {H,R} form an event space
- Therefore P[R] = 0.6
- The previous theorem (Thm 2.16) gives us $P_X(x)$!

Theorem 2.17

- Consider an Event B. Let B be a subset of the range S_X of RV X.
- What is $P_{X|B}(x)$?
- By definition $P_{X|B}(x) = P[X=x|B]$
- $P[X=x|B] = P[X=x,B]/ P[B]$ (By definition of conditional probability)
- $P[X=x, B] = ?$
  - Two possibilities exist. $\{X=x\}$ may or may not a subset of B
  - If $\{X=x\}$ is a subset then $\{X=x\} \cap B = \{X=x\}$!
  - Else $\{X=x\} \cap B = \{X=x\}$!

Theorem 2.17

- We have

$$P_{X|B}(x) = \begin{cases} \dfrac{P_X(x)}{P[B]} & x \in B \\ 0 & \text{otherwise.} \end{cases}$$

Note that this is > $P_X(x)$. Knowledge that B occurred and that {X=x} is in B increases our belief that {X=x} occurred

$$\boxed{P[X=x \mid X \geq 9]} = \begin{cases} \dfrac{P[X=x] \leftarrow (1/20)}{P[X \geq 9] \rightarrow 12/20} & x \in \{9, 10, \ldots, 20\} \\ 0 & \cancel{x < 9} \text{ otherwise} \end{cases}$$

VVS Word Problem

Example 2.40 Problem

Suppose X, the time in integer minutes you must wait for a bus, has the uniform PMF

$$P_X(x) = \begin{cases} 1/20 & x = 1, 2, \ldots, 20, \\ 0 & \text{otherwise.} \end{cases} \tag{2.119}$$

Suppose the bus has not arrived by the eighth minute, what is the conditional PMF of your waiting time X?

- You know that the bus did not arrive by the eight minute
- So, you know that the event {X > 8} occurred
- You want to find P[X=x | X > 8] or $P_{X|X>8}$
- P[X=x | X > 8] = P[X = x, X > 8]/P[X>8]
  - P[X = x, X > 8] for x = 9,10,…,20.
  - P[X = x, X > 8] for x = 0,1,…,8

Theorem 2.18

(a) For any $x \in B$, $P_{X|B}(x) \geq 0$.

(b) $\sum_{x \in B} P_{X|B}(x) = 1$.

(c) For any event $C \subset B$, $P[C|B]$, the conditional probability that X is in the set C, is

$$P\left[C|B\right] = \sum_{x \in C} P_{X|B}\left(x\right).$$

Conditional Expected Value

- Just replace the PMF $P_X(x)$ in the formula for expectation by the conditional PMF $P_{X|B}(x)$
- The same holds for the conditional expected value of a function of a RV too

Definition 2.20 ***Conditional Expected Value***

The conditional expected value *of random variable X given condition B is*

$$E[X|B] = \mu_{X|B} = \sum_{x \in B} x P_{X|B}(x).$$

Expectation of X in Terms of Conditional Expectations Over an Event Space

Theorem 2.19

For a random variable X resulting from an experiment with event space $B_1, \ldots, B_m$,

$$E[X] = \sum_{i=1}^{m} E[X|B_i]\, P[B_i].$$

$$\sum_{i=1}^{M} \underbrace{\left[\sum_{x \in B_i} x\, P[X=x|B_i]\right]}_{E[X|B_i]} P[B_i]$$

- We know from the definition of expectation that

$$E[X] = \sum_{x \in S_X} x P_X(x) = \sum_{x \in S_X} \sum_{i=1}^{M} x\, P[X=x, B_i]$$

- We want to express P[X=x] in terms of the conditional probabilities $P[X=x|B_i]$

$$E[X] = \sum_{x \in S_X} x \sum_{i=1}^{m} P_{X|B_i}(x) P[B_i]$$

Swap the summations and we get the result!

Conditional Expected Value of Y=g(X)

Theorem 2.20

The conditional expected value of $Y = g(X)$ given condition B is

$$E[Y|B] = E[g(X)|B] = \sum_{x \in B} g(x) P_{X|B}(x).$$

- Proof: Do as we did for E[Y] in Theorem 2.10.

Word Problem

Quiz 2.9

On the Internet, data is transmitted in packets. In a simple model for World Wide Web traffic, the number of packets N needed to transmit a Web page depends on whether the page has graphic images. If the page has images (event I), then N is uniformly distributed between 1 and 50 packets. If the page is just text (event T), then N is uniform between 1 and 5 packets. Assuming a page has images with probability $1/4$, find the

(1) conditional PMF $P_{N|I}(n)$

(2) conditional PMF $P_{N|T}(n)$

(3) PMF $P_N(n)$

(4) conditional PMF $P_{N|N\leq 10}(n)$

(5) conditional expected value $E[N|N \leq 10]$

(6) conditional variance $\mathrm{Var}[N|N \leq 10]$

$N|I \sim \mathrm{Unif}(1,50)$

$P[N=n|I] = \begin{cases} \frac{1}{50} & n \in \{1,2,\ldots,50\} \\ 0 & \text{otherwise} \end{cases}$

$\{I, T\}$

$N|T \sim \mathrm{Unif}(1,5)$

$P[N=n|T] = \begin{cases} \frac{1}{5} & n \in \{1,\ldots,5\} \\ 0 & \text{otherwise} \end{cases}$

$P[N=n|N \leq 10] = \dfrac{P[N=n, N \leq 10]}{P[N \leq 10]} = \begin{cases} \dfrac{P[N=n]}{P[N \leq 10]} & n \leq 10 \\ 0 & \text{otherwise} \end{cases}$

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

$E[N^2 | N \leq 10]$ $\quad (E[N | N \leq 10])^2$

Problem 2.7.7

A particular circuit works if all 10 of its component devices work. Each circuit is tested before leaving the factory. Each working circuit can be sold for k dollars, but each nonworking circuit is worthless and must be thrown away. Each circuit can be built with either ordinary devices or ultrareliable devices. An ordinary device has a failure probability of $q = 0.1$ while an ultrareliable device has a failure probability of $q/2$, independent of any other device. However, each ordinary device costs \$1 while an ultrareliable device costs \$3. Should you build your circuit with ordinary devices or ultrareliable devices in order to maximize your expected profit $E[R]$? Keep in mind that your answer will depend on k.

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2$$

$$E[N^2 | N \le 10] \qquad (E[N | N \le 10])^2$$

Problem 2.7.7

A particular circuit works if all 10 of its component devices work. Each circuit is tested before leaving the factory. Each working circuit can be sold for k dollars, but each nonworking circuit is worthless and must be thrown away. Each circuit can be built with either ordinary devices or ultrareliable devices. An ordinary device has a failure probability of $q = 0.1$ while an ultrareliable device has a failure probability of $q/2$, independent of any other device. However, each ordinary device costs \$1 while an ultrareliable device costs \$3. Should you build your circuit with ordinary devices or ultrareliable devices in order to maximize your expected profit $E[R]$? Keep in mind that your answer will depend on k.

Revenue obtained from a working ckt is k.

$$R_{ord} = \begin{cases} k - 10c & (0.9)^{10} \\ -10c & 1-(0.9)^{10} \end{cases}$$

$$E[R_{ord}] \qquad E[R_{rel}]$$

Problem 2.7.8

In the New Jersey state lottery, each \$1 ticket has six randomly marked numbers out of $1, \ldots, 46$. A ticket is a winner if the six marked numbers match six numbers drawn at random at the end of a week. For each ticket sold, 50 cents is added to the pot for the winners. If there are k winning tickets, the pot is divided equally among the k winners. Suppose you bought a winning ticket in a week in which $2n$ tickets are sold and the pot is n dollars.

(a) What is the probability q that a random ticket will be a winner?

(b) What is the PMF of K_n, the number of other (besides your own) winning tickets?

(c) What is the expected value of W_n, the prize you collect for your winning ticket?

$$W_n = \frac{n}{K_n+1} \qquad E[W_n] = E\left[\frac{n}{K_n+1}\right] = nE\left[\frac{1}{K_n+1}\right]$$

$$\neq \cancel{\frac{n}{E[K_n+1]}}$$

Problem 2.8.10

Let random variable X have PMF $P_X(x)$. We wish to guess the value of X before performing the actual experiment. If we call our guess $\hat{x}$, the expected square of the error in our guess is

$$e(\hat{x}) = E\left[(X - \hat{x})^2\right]$$

Show that $e(\hat{x})$ is minimized by $\hat{x} = E[X]$.

$$E\left[(X-\hat{x})^2\right] = \boxed{\sum_{x \in S_X} (x-\hat{x})^2 \underbrace{P[X=x]}_{P_X(x)}}$$

$$\sum_{x \in S_X} x P[X=x]$$

$$= \underbrace{\sum_{x \in S_X} \hat{x} P[X=x]}_{\hat{x}}$$

$$\sum_{x \in S_X} \cancel{2}(x-\hat{x}) P[X=x] = 0.$$

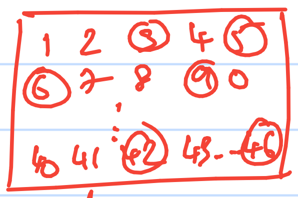
$$\hat{x} = ?$$

e) Problem 2.7.8.

e) Problem 2.7.8.

Price of lottery ticket = \$1

Ticket contains six randomly marked numbers out of 1, 2, ..., 46



We want to choose six numbers out of 46.

Total of ${}^{46}C_6$ ways of choosing a ticket.

Total tickets sold $= 2n$. $\Rightarrow$ Total revenue $= \$\,2n$

Total pot $= 2n(0.5) = n$.

(a) Probability a random ticket will be a winner $q = \dfrac{1}{{}^{46}C_6}$.

(b) Let k_n be the number of other (besides own) winning tickets.

We have a total of $2n$ tickets out of which one is own and is winning.

If $k_n = k$, we have $(k+1)$ winning tickets of which one is yours

$\underline{\ \ }_1\ \underline{\ \ }_2\ \underline{\ \ }_3\ \cdots\ \underline{\ \ }_{2n}$ (tickets) $\rightarrow$ own

$$P[k_n = k] = {}^{2n-1}C_k\, q^k (1-q)^{2n-1-k}$$

$$= {}^{2n-1}C_k\, q^k (1-q)^{2n-(k+1)}$$

(c) $W_n = \dfrac{n}{k_n+1}$

$$E[W_n] = E\left[\frac{n}{k_n+1}\right] = n\,E\left[\frac{1}{k_n+1}\right]$$

$$= n\sum_{k=0}^{2n-1}\left(\frac{1}{k+1}\right) P[k_n = k]$$

$$= n(1-q)^{2n}\sum_{k=0}^{2n-1} {}^{2n-1}C_k\, q^k (1-q)^{-(k+1)}\left(\frac{1}{k+1}\right)$$

$$= n(1-q)^{2n}\,\lfloor 2n-1 \sum_{k=0}^{2n-1}\frac{q^k(1-q)^{-(k+1)}}{\lfloor k\ \lfloor 2n-1-k\ (k+1)}$$

$$= \frac{n(1-q)^{2n}\,\lfloor 2n-1}{(1-q)}\sum_{k=0}^{2n-1}\frac{}{\lfloor k+1\ \lfloor 2n-(k+1)}\left(\frac{q}{1-q}\right)^k$$

$$= \frac{1}{2q}\sum_{k=0}^{2n-1} {}^{2n}C_{k+1}\, q^{k+1}(1-q)^{2n-(k+1)}$$

$$= \frac{1}{2q}\sum_{k+1=1}^{2n} {}^{2n}C_{k+1}\, q^{k+1}(1-q)^{2n-(k+1)}$$

$$= \frac{1}{2q}\left[1-(1-q)^{2n}\right] = \left(\frac{1-(1-q)^{2n}}{2q}\right)$$

[Good to simulate for diff n]

Probability and Random Processes

Continuous Random Variables

Sanjit Kaul

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Continuous Random Variables

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Intervals

- (x_1, x_2)
  - Includes neither x_1 nor x_2 but all real numbers in between
- $[x_1, x_2)$
- $(x_1, x_2]$
- $[x_1, x_2]$

A Very Narrow Carrom Board

- A very narrow board of length L can be approximated by a line
  - This is a terrible attempt on my part to emulate the example in RY
- You shoot the striker (red)
- Your strike is powerful
- This ensures that the striker stops at any point along the thin board with equal probability
- The striker stops at a point in [0,L]

A Very Narrow Carrom Board

- Let X be the RV that describes the final location of the pointer
- We have $S_X = [0,L]$
- The range is a continuum. Thus X is a CRV
- What is $P[X = x]$ where $x \in S_x$?

A Very Narrow Carrom Board

- Let's think in discrete terms

- We split the board into n equal intervals
- Define the RV Y with $S_Y = \{1,2,\ldots,n\}$
- Y is a discrete uniform RV
- Note that in the discrete problem all values of X in interval k correspond to Y=k.

$$P[Y = k] = P[k - 1 \leq X < k]$$

A Very Narrow Carrom Board

If $X = x$ then $Y = \lceil x/(L/n) \rceil = \lceil nx/L \rceil$

- Also

$$\{X = x\} \subset \{Y = \lceil nx/L \rceil\}$$

- Therefore

$$P[X = x] \leq P[Y = \lceil nx/L \rceil] = 1/n$$

A Very Narrow Carrom Board

- Getting as close as possible to the continuous case

$$P[X = x] \leq \lim_{n\to\infty} 1/n = 0$$

- Therefore

$$P[X = x] = 0$$

- Every outcome {X=x} has a probability 0!
  - Note that this is not true for an interval on the board
- The PMF is not a useful way of describing a continuous RV
- What about the CDF?

Def 3.1 Cumulative Distribution Function

- The CDF of a RV X is $F_X(x) = P[X \leq x]$
- Properties of a CDF are the same for all kinds of random variables
  - Discrete,
  - Continuous, and
  - Mixed.

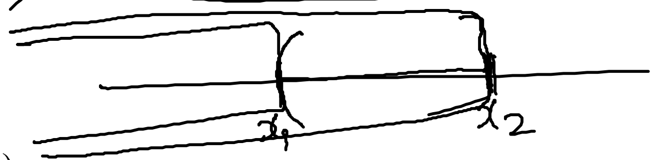
Theorem 3.1

For any random variable X,

$P[X \in (x_1, x_2]]$

(a) $F_X(-\infty) = 0$

(b) $F_X(\infty) = 1$

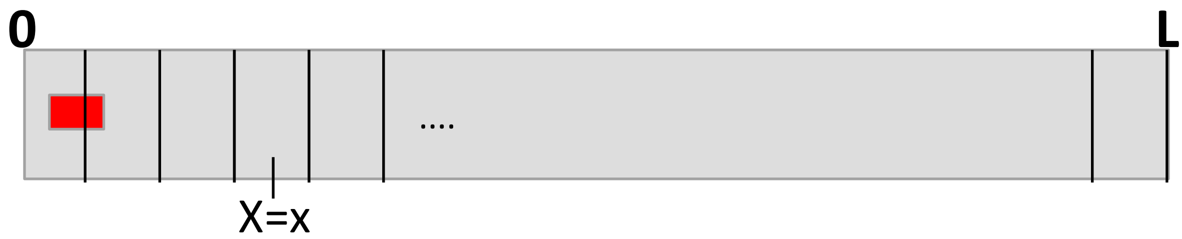


(c) $P[x_1 < X \leq x_2] = F_X(x_2) - F_X(x_1)$

Problem

- What is the CDF of X?
- P[X ≤ x] = ?, What is your guess?
- We can say that
  - P[X ≤ x] = 1, for x >= L
  - P[X ≤ x] = 0, for x <= 0
- The probability accumulates and goes from 0 to 1 as we go from x=0 to x=L.
- Given that all regions on the board are equally preferred, intuitively probability must accumulate from 0 to x/L as we go from 0 to x
  - P[X ≤ x] = x/L for 0 <= x <= L

Problem



- We have

$$\{Y \leq \lceil nx/L \rceil \;\; - 1\} \subset \{X \leq x\} \subset \{Y \leq \lceil nx/L \rceil\}$$

- Therefore

$$F_Y(\lceil nx/L \rceil \;\; - 1) \leq P[X \leq x] \leq F_Y(\lceil nx/L \rceil)$$

- Since Y is a discrete uniform distribution

$$\frac{\lceil nx/L \rceil - 1}{n} \leq P[X \leq x] \leq \frac{\lceil nx/L \rceil)}{n}$$

Calculating CDF for the Board Game

- As n goes to infinity, we get the CDF of the continuous uniform RV

$$P[X \leq x] = x/L \text{ for } 0 \leq x < L.$$

- We used the fact that

$$\lim_{n \to \infty} \frac{\lceil nx/L \rceil}{n} = \frac{x}{L}$$

CDF Discrete vs. Continuous

- The CDF of a discrete RV contained jumps
- The size of the jump (change) was the probability of the RV at the point of jump
- The larger the jump, the larger the probability
- The CDF of a continuous RV is continuous

Quiz 3.1

$f(x) = x^2$

The cumulative distribution function of the random variable Y is

$(-\infty, y]$

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ y/4 & 0 \le y \le 4 \\ 1 & y > 4 \end{cases} \tag{3.9}$$

Sketch the CDF of Y and calculate the following probabilities: $f(x=1)$

(1) $P[Y \le -1]$ $\quad F_Y(y=1) - F_Y(-\infty)$ $\quad f(1)$

$F_Y(y \le 1)$

(2) $P[Y \le 1]$

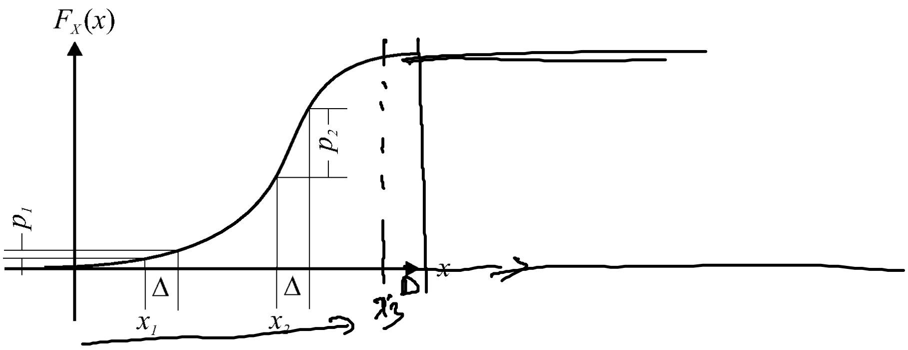
$P[Y \le 3] - P[Y \le 2]$

(3) $P[2 < Y \le 3]$ $\quad F_Y(3) - F_Y(2)$

(4) $P[Y > 1.5]$

Example CDF of a continuous RV

Figure 3.2



$F_X(x_3+\Delta) - F_X(x_3)$

$\leq$

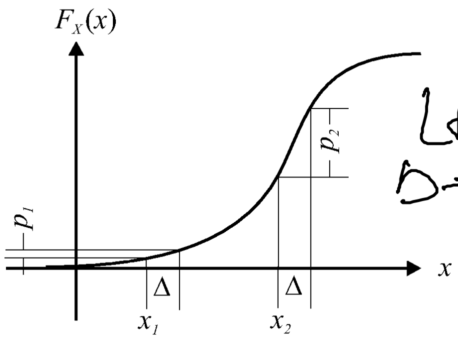
The graph of an arbitrary CDF $F_X(x)$.

- From the figure

$$F_X(x_2+\Delta) - F_X(x_2) > F_X(x_1+\Delta) - F_X(x_1)$$

Information in a CDF

Figure 3.2



$$\lim_{\Delta \to 0} \frac{F_X(x_2+\Delta) - F_X(x_2)}{(x_2+\Delta) - x_2}$$

$f_X(x_2)$ PDF

The graph of an arbitrary CDF $F_X(x)$.

- Rate of accumulation of probability in the region right of x_2 is greater than the in the region right of x_1
- As the interval Δ becomes smaller the rate of accumulation -> slope of the CDF

Probability Density Function

- Greater **the slope of the CDF** at any point x, the more the likelihood of an outcome in an interval around x

Probability Density Function

Definition 3.3 *(PDF)*

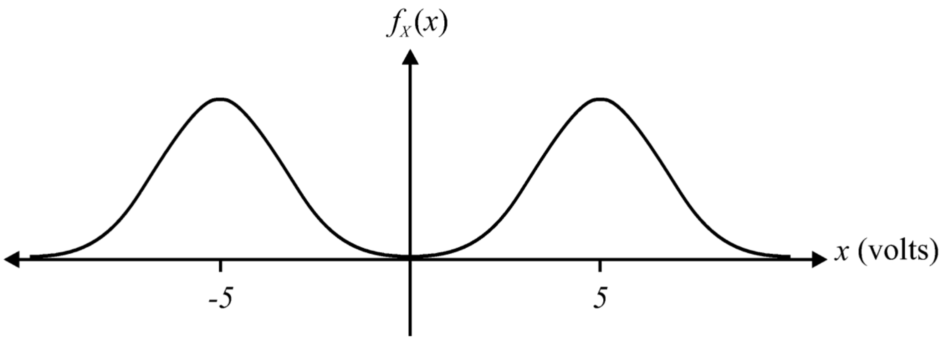
The probability density function *(PDF) of a continuous random variable* X *is*

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

Note the lowercase f

Example PDF

Figure 3.3



The PDF of the modem receiver voltage X.

- You are sending 0(s) and 1(s) mapped to -5 and +5
- Receiver sees the above PDF, thanks to Gaussian noise
- Which regions are most likely to be seen by the receiver and why?

Properties of a PDF

Theorem 3.2

For a continuous random variable X with PDF $f_X(x)$,

(a) $f_X(x) \geq 0$ for all x,

(b) $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$,

(c) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$.

- Why is (a) true?
  - CDF is a non-decreasing function and PDF is its slope
- Why (b)?
  - From definition of a PDF
- Why (c)?
  - The integral is $P[X < \infty] = F_X(\infty) = 1$.

Properties of a PDF

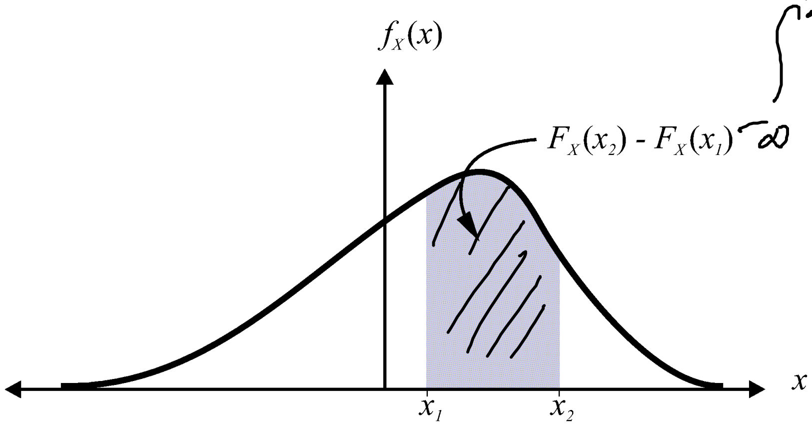
Theorem 3.3

$$P\left[x_1 < X \le x_2\right] = \int_{x_1}^{x_2} f_X(x)\, dx.$$

$P[X \le x_2] - P[X \le x_1]$

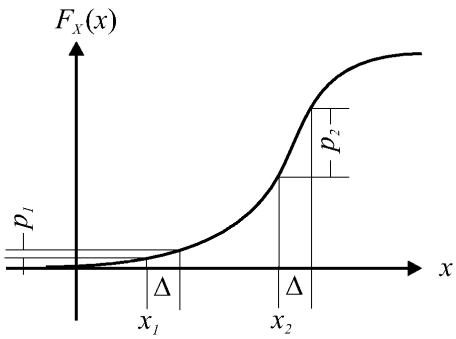
$\int_{-\infty}^{x_2} f_X(x)\, dx$

$- \int_{-\infty}^{x_1} f_X(x)\, dx$



The PDF and CDF of X.

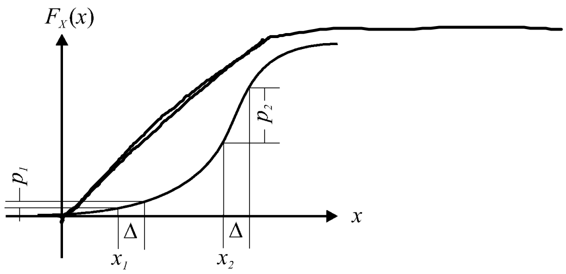
Figure 3.2



The graph of an arbitrary CDF $F_X(x)$.

$$P[x_2 < X \leq x_2 + \Delta] = \frac{F_X(x_2+\Delta) - F_X(x_2)}{\Delta}\Delta$$

Figure 3.2



The graph of an arbitrary CDF $F_X(x)$.

$$P[x_2 < X \leq x_2 + \Delta] = \frac{F_X(x_2+\Delta) - F_X(x_2)}{\Delta}\Delta$$

- As the interval Δ becomes smaller

$$\frac{F_X(x_2+\Delta) - F_X(x_2)}{\Delta} = f_X(x_2)$$

- In general we can write for any x and an infinitesimally small *dx*

$$P[x < X \leq x + dx] = f_X(x)dx$$

Intervals and CRV(s)

- Consider the four different events
- A = (0,1)
- B = (0,1]
- C = [0,1)
- D = [0,1]
- They belong to the range space of a **continuous RV X**
- What can we say about P[A], P[B], P[C], and P[D]?
- Since X is continuous, P[A] = P[B] = P[C] = P[D]
  - This is true because probability of a point (outcome) is 0 for a continuous RV

Problem on PDF

Quiz 3.2

Random variable X has probability density function

$$f_X(x) = \begin{cases} cxe^{-x/2} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.25}$$

Sketch the PDF and find the following:

(1) the constant c

(2) the CDF $F_X(x)$

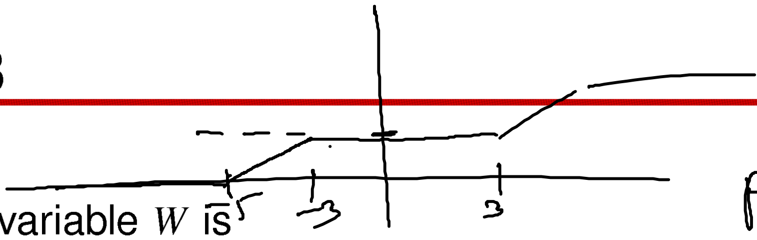
(3) $P[0 \leq X \leq 4]$

(4) $P[-2 \leq X \leq 2]$

(5) What is the mode of the random variable X?

Problem 3.1.3

The CDF of random variable W is



$P[W \le 3]$

$=$

$y = (x+5)/8$

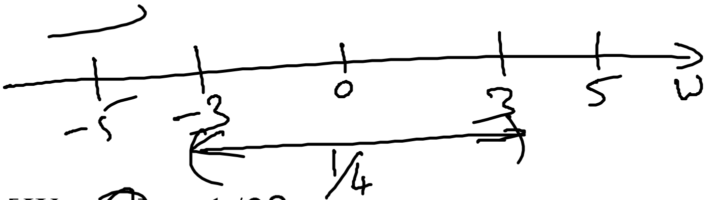
$$F_W(w) = P[W \le w] = \begin{cases} 0 & w < -5, \\ (w+5)/8 & -5 \le w < -3, \\ 1/4 & -3 \le w < 3, \\ 1/4 + 3(w-3)/8 & 3 \le w < 5, \\ 1 & w \ge 5. \end{cases}$$

(a) What is $P[W \le 4]$?

(b) What is $P[-2 < W \le 2]$?

(c) What is $P[W > 0]$?

(d) What is the value of a such that $P[W \le a] = 1/2$?



$\frac{1}{4} + 3(w-3)/8 = 1/2$

$w = 11/3$

$P[W \le 2.9] = 1/4$

Problem 3.2.4

For a constant parameter $a > 0$, a Rayleigh random variable X has PDF

$$f_X(x) = \begin{cases} a^2 x e^{-a^2x^2/2} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

What is the CDF of X?

Expected Value

- **Def 3.4** The expected value of a continuous random variable X is

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

- Summation for the discrete case is replaced by integration for the continuous case
- **Theorem 3.4** The expected value of a function g(X) of a continuous random variable X is

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

Example of g(X)

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

$P[X \leq x] = x/L$ for $0 \leq x < L$. (and 0 otherwise)

What is $f_X(x)$? ⇨ $f_X(x) = \begin{cases} \frac{1}{L} & 0 \leq x < L \\ 0 & \text{otherwise.} \end{cases}$

VVVS Problem

- What is E[X]?
  - L/2
- What is $E[X^2]$?
  - $L^2/3$
- What is Var[X]?
  - $L^2/3 - L^2/4 = L^2/12$

Theorem 3.5

For any random variable X,

(a) $E[X - \mu_X] =$

(b) $E[aX + b] =$

(c) $\mathrm{Var}[X] =$

(d) $\mathrm{Var}[aX + b] =$

Problem 3.4.1

Radars detect flying objects by measuring the power reflected from them. The reflected power of an aircraft can be modeled as a random variable Y with PDF

$Y > E[Y]$

$$f_Y(y) = \begin{cases} \frac{1}{P_0} e^{-y/P_0} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $P_0 > 0$ is some constant. The aircraft is correctly identified by the radar if the reflected power of the aircraft is larger than its average value. What is the probability $P[C]$ that an aircraft is correctly identified?

$\frac{1}{P_0}\int_0^\infty e^{-y/P_0}\,dy = 1$ $\quad \int_0^\infty y\, f_Y(y)\,dy$ $\quad P_0$ $\quad e^{-1}$ $\quad \int_{P_0}^\infty f_Y(y)\,dy$

Families of Continuous RV(s)

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

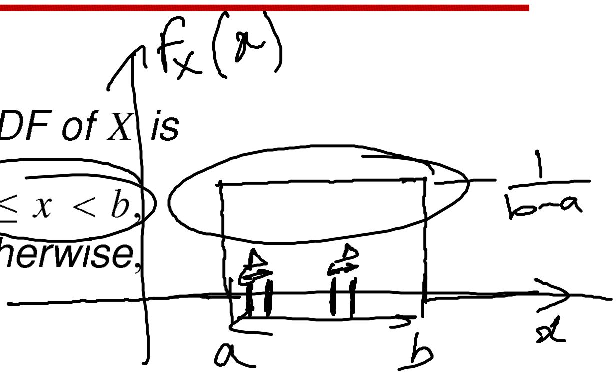
Uniform(a,b)

Definition 3.5 *Uniform Random Variable*

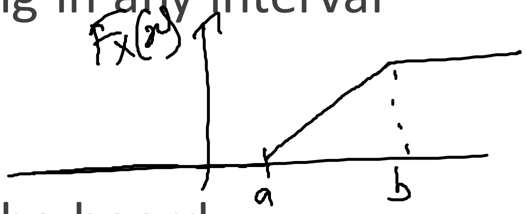
X is a uniform (a, b) random variable if the PDF of X is

$$f_X(x) = \begin{cases} 1/(b-a) & a \le x < b, \\ 0 & \text{otherwise,} \end{cases}$$

where the two parameters are $b > a$.



- The probability of an outcome being in any interval of a given size Δ is the same
- How can we conclude this?
- The final position of the striker in the board example was *uniformly distributed*

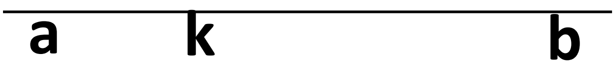


Theorem 3.7

Let X be a uniform (a, b) random variable, where a and b are both integers. Let $K = \lceil X \rceil$. Then K is a ______ random variable.

- You want to describe the RV K
- K is a discrete RV. It is described completely by its PMF P[K = k]

P[K=k]



Given the ceil function, all x that are > k-1 and <= k correspond to the outcome {K=k}

$$P[K = k] = P[k - 1 < x \leq k] = \frac{1}{b-a}$$

Discrete and Continuous Uniform RV(s)

$$P[K = k] = P[k - 1 < x \leq k] = \frac{1}{b-a}$$

- The above is true for k = a+1, a+2,…, b
- P[K=k] = 0 otherwise.
- Clearly K is a discrete uniform RV

Exponential(λ)

Definition 3.6 Exponential Random Variable

X is an exponential (λ) random variable *if the PDF of X is*

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

where the parameter $\lambda > 0$.

A

$\cancel{f_X(x)}$

$$f_Y(y) = \begin{cases} c\lambda e^{-\lambda y} & 0 \leq y < a \\ 0 & \text{otherwise} \end{cases}$$

- The inter-arrival time between two packet arrivals at a server may be modeled as an Exp RV
- What is the average inter-arrival time?

$c = \frac{1}{A}$

Theorem 3.9

If X is an exponential (λ) random variable, then $K = \lceil X \rceil$ is a geometric (p) random variable with $p = 1 - e^{-\lambda}$.

- Proof is similar to the case of Uniform distribution

Example 3.14 Problem

Phone company A charges \$$0.15$ per minute for telephone calls. For any fraction of a minute at the end of a call, they charge for a full minute. Phone Company B also charges \$$0.15$ per minute. However, Phone Company B calculates its charge based on the exact duration of a call. If T, the duration of a call in minutes, is an exponential $(\lambda = 1/3)$ random variable, what are the expected revenues per call $E[R_A]$ and $E[R_B]$ for companies A and B?

Erlang(n,λ)

Definition 3.7 Erlang Random Variable

X is an Erlang (n, λ) *random variable if the PDF of X is*

$$f_X(x) = \begin{cases} \dfrac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} & x \geq 0, \\ 0 & \textit{otherwise}. \end{cases}$$

where the parameter $\lambda > 0$, and the parameter $n \geq 1$ is an integer.

- The waiting time between n-occurrences of an event
- The waiting time between n customer arrivals
- n => Sum of n independent Exp(λ) Random Variables

Theorem 3.10

If X is an Erlang (n, λ) random variable, then

$X = X_1 + X_2 + \cdots + X_n$

$$E[X] = \frac{n}{\lambda}, \qquad \text{Var}[X] = \frac{n}{\lambda^2}.$$

$X_1, X_2, \ldots, X_n$

$X_i \sim f_{X_i}(x) \sim \text{Exp}(\lambda)$

$E[X_i] = \frac{1}{\lambda}$

$E[X]$

$= E[X_1] + \cdots + E[X_n] = \frac{n}{\lambda}$

$E\left[\left(X_1 + \cdots + X_n - E[X_1 + \cdots + X_n]\right)^2\right]$

$= E\left[(X_1 + X_2 + \cdots + X_n)^2\right] - \left(E\left[(X_1 + \cdots + X_n)\right]\right)^2$

Quiz 3.4

Continuous random variable X has $E[X] = 3$ and $\text{Var}[X] = 9$. Find the PDF, $f_X(x)$, if

(1) X has an exponential PDF,

(a, b)

(2) X has a uniform PDF.

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

$$\frac{(b-a)^2}{12} = 9$$

A, A^C $\quad A = \{M \le 20\}$ $\quad E[C_B] = E[C_B|A]P[A]$

Problem 3.4.9

$A^C = \{M > 20\}$ $\quad + \underbrace{E[C_B|A^C]}_{E[99+10(M-20)]} P[A^C]$

Long-distance calling plan A offers flat rate service at 10 cents per minute. Calling plan B charges 99 cents for every call under 20 minutes; for calls over 20 minutes, the charge is 99 cents for the first 20 minutes plus 10 cents for every additional minute. (Note that these plans measure your call duration exactly, without rounding to the next minute or even second.) If your long-distance calls have exponential distribution with expected value τ minutes, which plan offers a lower expected cost per call?

M

$C_A = 10M \qquad E[C_A] = 10E[M] = 10\tau.$

$P[M \le 20]$

$C_B = 99 + 10(M-20) \qquad M \ge 20$

$P[M > 20]$

$E[C_B] = \boxed{P[M \le 20]\,99 + P[M > 20]\,(E[99+10(M-20)])}$

99+

$M \sim Exp\left(\frac{1}{\tau}\right)$

$P[M > m] = e^{-m/\tau}$

$\underbrace{(M-20)}\ 10$

$\hookrightarrow$

$P[M > 20] = e^{-\left(\frac{1}{\tau}\right)20}$

$P[M - 20 > 10] = P[M > 30]$

$E[M - 20] = E[M'] = \int_0^\infty m'$

$f_{M|M>20}(m)$

Gaussian(μ,σ)

- The Bell-shaped distribution
- The Normal distribution
- Parameters are the mean μ and standard deviation is σ
  - Variance is σ^2
  - If X is Gaussian we often write X is N[μ, σ^2]

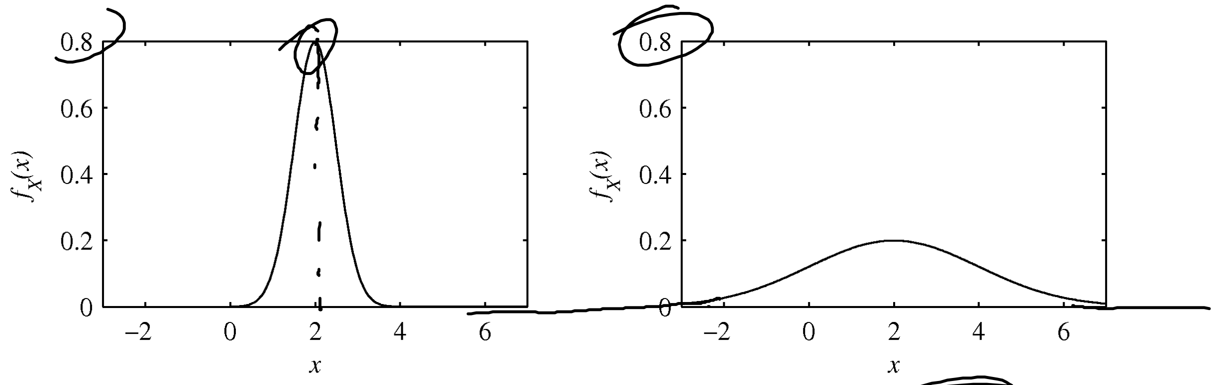
Definition 3.8 Gaussian Random Variable

X is a Gaussian (μ, σ) random variable if the PDF of X is

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

where the parameter μ can be any real number and the parameter $\sigma > 0$.

Figure 3.5



(a) $\mu = 2, \sigma = 1/2$ **(b)** $\mu = 2, \sigma = 2$

Two examples of a Gaussian random variable X with expected value μ and standard deviation σ.

Show that the Gaussian PDF is a valid PDF

- Let W be a Gaussian RV

$$\int_{-\infty}^{\infty} f_W(w)\, dw = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(w-\mu)^2/(2\sigma^2)}\, dw$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x)^2/(2\sigma^2)}\, dx$$

- The square of the integral is

$$I^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(y)^2/(2\sigma^2)}\, dy \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x)^2/(2\sigma^2)}\, dx$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} e^{(x^2+y^2)/(2\sigma^2)}\, dxdy$$

- Substitute x=r sin(ϕ) and y = rcos(ϕ) and show $I^2 = 1$

Derive E[W] and E[W^2]

- Use steps similar to showing area is unity
- You will also need to apply integration by parts

Linear Transformation of a Gaussian

Theorem 3.13

If X is Gaussian (μ, σ), $Y = aX + b$ is ________.

$a^2\sigma^2$

- Linear transformation of a Gaussian gives another Gaussian!
- How do show the above?
- If X is not Gaussian, what are E[Y] and E[Y^2]?

Y is Gaussian $(a\mu + b, a\sigma)$

Standard Normal Variable and CDF

- **Def 3.9** The standard random variable is Gaussian(0,1) – 0 mean and unit variance

$\uparrow \uparrow$

(μ, σ)

$$\int_{-\infty}^{x} f_X(x)\,dx$$

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left[(x-\mu)^2/(2\sigma^2)\right]}$$

***Definition 3.10** Standard Normal CDF*

The CDF of the standard normal random variable Z is

$$\boxed{\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\,du.}$$

$E(\sigma Z + \mu)$

$X \sim \text{Gaussian}(\mu, \sigma)$ $\quad Z \sim N(0,1)$ $\quad \frac{X-\mu}{\sigma} = Z$

Expressing a Gaussian CDF as a N[0,1] CDF

Theorem 3.14

$P[X \le 100] = P[X - \mu_X \le 100 - \mu_X]$

$= P\left[\frac{X-\mu_X}{\sigma_X} \le \frac{100-\mu_X}{\sigma_X}\right]$

If X is a Gaussian (μ, σ) random variable, the CDF of X is

$$\boxed{F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).}$$

$X \sim N(2, 20)$

The probability that X is in the interval $(a, b]$ is

$$P[a < X \le b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

$P[X \le 100] = \int_{-\infty}^{100} f_X(x)\,dx$

$\int_{-\infty}^{100} \frac{1}{\sqrt{2\pi(400)}} e^{-\left[(x-2)^2/2(400)\right]} dx$

- Show using definition of CDF and substitution of variables in the integral
- All I need are values of ϕ(.)
  - This is important as CDF integral calculations for finite limits can only be done numerically
  - Thankfully, all we now need is a tabulation of ϕ(.)

Problem 3.4.9.

We can use conditional PDF(s) or we can use what we know about calculating the $E[g(X)]$ where $g(X)$ is a function of the RV X.

We had defined C_B as the random variable of costs on using plan B.

We had said $C_B = \begin{cases} 99 & M \le 20 \\ 10(M-20) & M > 20 \end{cases}$

One way therefore write C_B as

$$C_B = 99 + 10(M-20)^+ \qquad \boxed{X^+ = X \text{ if } X > 0 \\ = 0 \text{ otherwise}}$$

We want $E[C_B] = E[99 + 10(M-20)^+]$

$= 99 + 10\,E[(M-20)^+]$

$E[(M-20)^+]$ is in the form $E[g(M)]$ where $g(M) = (M-20)^+$.

$$\therefore E[(M-20)^+] = \int_0^\infty (M-20)^+ f_M(m)\,dm \quad (1)$$

$$f_M(m) = \begin{cases} \frac{1}{\tau} e^{-M/\tau} & \tau \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

This is simply the PDF of an exponential RV with mean $\frac{1}{\tau}$. That M is $Exp(1/\tau)$ is given in the question.

Equation (1) is $E[(M-20)^+] =$

$$\int_0^\infty (m-20)^+ f_M(m)\,dm$$

$$= \int_{20}^\infty (M-20) \frac{1}{\tau} e^{-m/\tau}\,dm \quad (2)$$

Let $y = m - 20$. $\quad m = 20 \Rightarrow y = 0$, $\; m = \infty \Rightarrow y = \infty$

Substituting in (2) we get:

$$\int_0^\infty y \frac{1}{\tau} e^{-(y+20)/\tau}\,dy$$

$$= e^{-20/\tau} \underbrace{\int_0^\infty y \frac{1}{\tau} e^{-(y/\tau)}\,dy}_{\text{Expectation of } Exp(1/\tau)}$$

$= \tau e^{-20/\tau}$

$\therefore E[C_B] = 99 + 10\tau e^{-20/\tau}$.

(**) Bonus: $E[X^2]$ when $X \sim Exp(\mu)$.

$$\int_0^\infty x^2 \mu e^{-\mu x}\,dx$$

$$= -\int_0^\infty x^2\, de^{-\mu x}$$

$$= -\left[(x^2 e^{-\mu x})_0^\infty - \int_0^\infty 2x e^{-\mu x}\,dx\right]$$

Note that $\lim_{x \to \infty} x^2 e^{-\mu x} = 0$.

$$= \int_0^\infty 2x e^{-\mu x}\,dx$$

$$= 2\left[\int_0^\infty x e^{-\mu x}\,dx\right]$$

$$= \frac{-2}{\mu}\left[\int_0^\infty x\, de^{-\mu x}\right]$$

$$= \frac{-2}{\mu}\left[(x e^{-\mu x})_0^\infty - \int_0^\infty e^{-\mu x}\,dx\right]$$

$$= +\frac{2}{\mu}\int_0^\infty e^{-\mu x}\,dx$$

$$= \frac{2}{\mu^2}\underbrace{\int_0^\infty \mu e^{-\mu x}\,dx}_{\text{Integrating over PDF of an Exp RV.}} = \frac{2}{\mu^2}. \; //$$

Thm 3.13

Linear Transformation of a Gaussian gives another Gaussian.

$Y = aX + b$ where $X \sim \text{Gaussian}(\mu, \sigma)$

$$P(Y \le y) = P[aX + b \le y]$$

$$= P[X \le (y-b)/a]$$

$$= \int_{-\infty}^{(y-b)/a} f_X(x)\, dx$$

$$= \int_{-\infty}^{(y-b)/a} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} dx \quad \text{—} (1)$$

$$y = \frac{a(y-b)}{a} + b$$

let $ax + b = z$

$\therefore \quad x = \left(\frac{z-b}{a}\right) \Rightarrow z = ax + b.$

$dx = \frac{dz}{a}$

$\therefore$ (1) becomes

$$\int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{z-b}{a} - \mu\right)^2 / 2\sigma^2} dz$$

$$= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi(a\sigma)^2}} e^{-(z-(a\mu+b))/[2(a\sigma)^2]} dz.$$

$\Rightarrow \quad Y \sim \text{Gaussian}(a\mu + b, a\sigma)$

Hence proved.

$$Y = aX + b$$

$$f_Y(y) = f_Y(ax+b)$$

$$F_Y(y) = F_Y(ax+b)$$

$$\frac{F_Y(y)}{dy} = f_{ax+b}(y)dy = f_X\left(\frac{y-b}{a}\right)\frac{dy}{a}$$

$$= \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\left(\frac{\left(\frac{y-b}{a} - \mu_x\right)^2}{2\sigma_x^2}\right)} \frac{dy}{a}$$

$$= \frac{1}{\sqrt{2\pi\sigma_x^2 a^2}} e^{-\left(\frac{(y-b-a\mu_x)^2}{2\sigma_x^2 a^2}\right)} dy.$$

$X = 0 \Rightarrow Y = b.$ There is a one to one correspondence. We want rate of acc around $X=0$ to be same as around $Y=b$

Probability and Random Processes

Continuous Random Variables

Sanjit Kaul

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Continuous Random Variables

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Intervals

- (x_1, x_2)
  - Includes neither x_1 nor x_2 but all real numbers in between
- $[x_1, x_2)$
- $(x_1, x_2]$
- $[x_1, x_2]$

A Very Narrow Carrom Board

- A very narrow board of length L can be approximated by a line
  - This is a terrible attempt on my part to emulate the example in RY
- You shoot the striker (red)
- Your strike is powerful
- This ensures that the striker stops at any point along the thin board with equal probability
- The striker stops at a point in [0,L]

A Very Narrow Carrom Board

- Let X be the RV that describes the final location of the pointer
- We have $S_X = [0,L]$
- The range is a continuum. Thus X is a CRV
- What is $P[X = x]$ where $x \in S_x$?

A Very Narrow Carrom Board

- Let's think in discrete terms

- We split the board into n equal intervals
- Define the RV Y with $S_Y = \{1,2,\ldots,n\}$
- Y is a discrete uniform RV
- Note that in the discrete problem all values of X in interval k correspond to Y=k.

$$P[Y = k] = P[k - 1 \leq X < k]$$

A Very Narrow Carrom Board

If $X = x$ then $Y = \lceil x/(L/n) \rceil = \lceil nx/L \rceil$

- Also

$$\{X = x\} \subset \{Y = \lceil nx/L \rceil\}$$

- Therefore

$$P[X = x] \leq P[Y = \lceil nx/L \rceil] = 1/n$$

A Very Narrow Carrom Board

- Getting as close as possible to the continuous case

$$P[X = x] \leq \lim_{n\to\infty} 1/n = 0$$

- Therefore

$$P[X = x] = 0$$

- Every outcome {X=x} has a probability 0!
  - Note that this is not true for an interval on the board
- The PMF is not a useful way of describing a continuous RV
- What about the CDF?

Def 3.1 Cumulative Distribution Function

- The CDF of a RV X is $F_X(x) = P[X \leq x]$
- Properties of a CDF are the same for all kinds of random variables
  - Discrete,
  - Continuous, and
  - Mixed.

Theorem 3.1

For any random variable X,

$P[X \in (x_1, x_2]]$

(a) $F_X(-\infty) = 0$

(b) $F_X(\infty) = 1$

(c) $P[x_1 < X \leq x_2] = F_X(x_2) - F_X(x_1)$

Problem

- What is the CDF of X?
- $P[X \leq x]$ = ?, What is your guess?
- We can say that
  - $P[X \leq x] = 1$, for $x >= L$
  - $P[X \leq x] = 0$, for $x <= 0$
- The probability accumulates and goes from 0 to 1 as we go from x=0 to x=L.
- Given that all regions on the board are equally preferred, intuitively probability must accumulate from 0 to x/L as we go from 0 to x
  - $P[X \leq x] = x/L$ for $0 <= x <= L$

Problem

- We have

$$\{Y \leq \lceil nx/L \rceil \; - 1\} \subset \{X \leq x\} \subset \{Y \leq \lceil nx/L \rceil\}$$

- Therefore

$$F_Y(\lceil nx/L \rceil \; - 1) \leq P[X \leq x] \leq F_Y(\lceil nx/L \rceil)$$

- Since Y is a discrete uniform distribution

$$\frac{\lceil nx/L \rceil - 1}{n} \leq P[X \leq x] \leq \frac{\lceil nx/L \rceil)}{n}$$

Calculating CDF for the Board Game

- As n goes to infinity, we get the CDF of the continuous uniform RV

$$P[X \leq x] = x/L \text{ for } 0 \leq x < L.$$

- We used the fact that

$$\lim_{n\to\infty} \frac{\lceil nx/L \rceil}{n} = \frac{x}{L}$$

CDF Discrete vs. Continuous

- The CDF of a discrete RV contained jumps
- The size of the jump (change) was the probability of the RV at the point of jump
- The larger the jump, the larger the probability
- The CDF of a continuous RV is continuous

Quiz 3.1

$f(x) = x^2$

The cumulative distribution function of the random variable Y is

$(-\infty, 1]$

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ y/4 & 0 \le y \le 4 \\ 1 & y > 4 \end{cases} \tag{3.9}$$

Sketch the CDF of Y and calculate the following probabilities:

$f(x=1)$

$f(1)$

(1) $P[Y \le -1]$

$F_Y(y=1) - F_Y(-\infty)$

$F_Y(y \le 1)$

(2) $P[Y \le 1]$

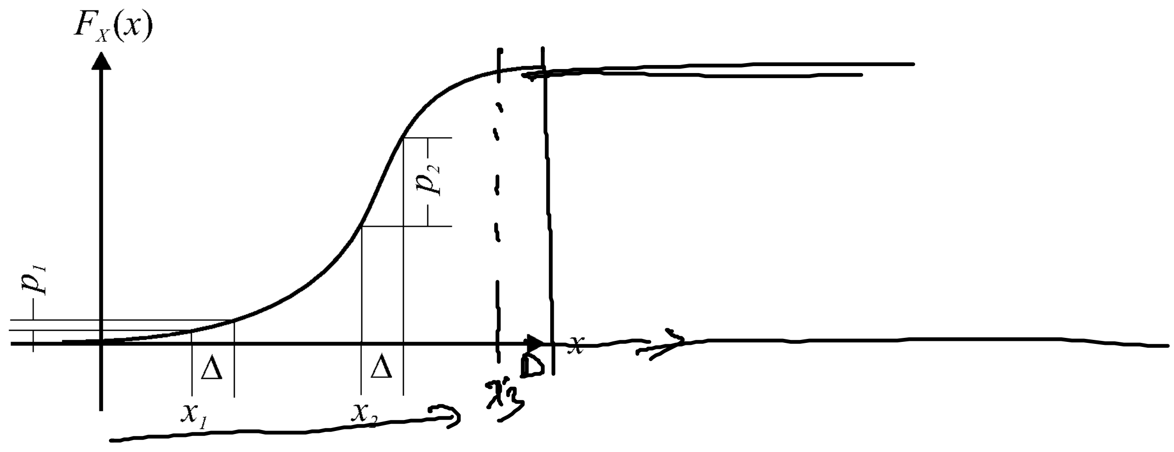
$P[Y \le 3) - P[Y \le 2]$

(3) $P[2 < Y \le 3]$ $\rightarrow$ $F_Y(3) - F_Y(2)$

(4) $P[Y > 1.5]$

Example CDF of a continuous RV

Figure 3.2



$F_X(x_3+\Delta) - F_X(x_3)$

$\leq$

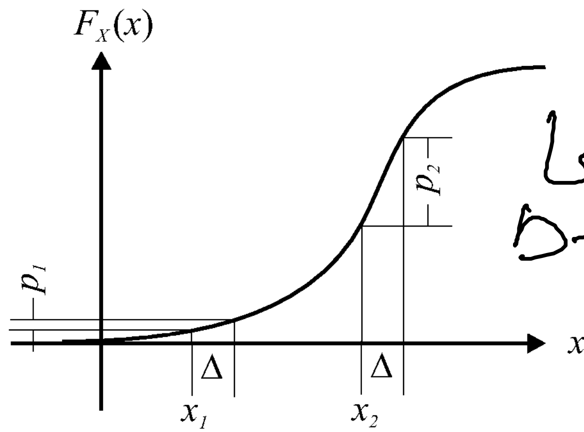
The graph of an arbitrary CDF $F_X(x)$.

- From the figure

$$F_X(x_2+\Delta) - F_X(x_2) > F_X(x_1+\Delta) - F_X(x_1)$$

Information in a CDF

Figure 3.2



$$\underset{\Delta \to 0}{Lt} \frac{F_X(x_2+\Delta) - F_X(x_2)}{(x_2+\Delta) - x_2}$$

$f_X(x_2)$ PDF

The graph of an arbitrary CDF $F_X(x)$.

- Rate of accumulation of probability in the region right of x_2 is greater than the in the region right of x_1
- As the interval Δ becomes smaller the rate of accumulation -> slope of the CDF

Probability Density Function

- Greater **the slope of the CDF** at any point x, the more the likelihood of an outcome in an interval around x

Probability Density Function

Definition 3.3 (PDF)

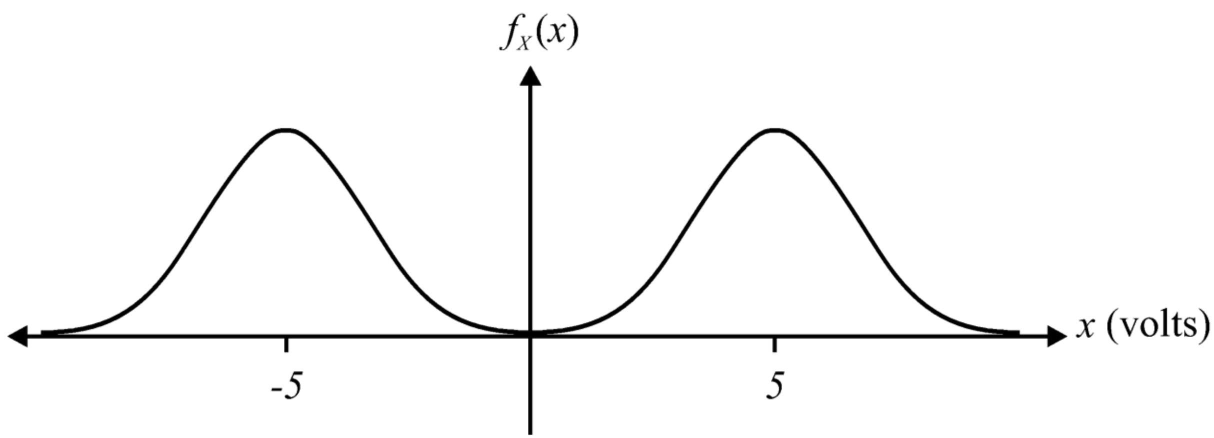
The probability density function *(PDF) of a continuous random variable* X *is*

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

Note the lowercase f

Example PDF

Figure 3.3



The PDF of the modem receiver voltage X.

- You are sending 0(s) and 1(s) mapped to -5 and +5
- Receiver sees the above PDF, thanks to Gaussian noise
- Which regions are most likely to be seen by the receiver and why?

Properties of a PDF

Theorem 3.2

For a continuous random variable X with PDF $f_X(x)$,

(a) $f_X(x) \geq 0$ for all x,

(b) $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$,

(c) $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$.

- Why is (a) true?
  - CDF is a non-decreasing function and PDF is its slope
- Why (b)?
  - From definition of a PDF
- Why (c)?
  - The integral is $P[X < \infty] = F_X(\infty) = 1$.

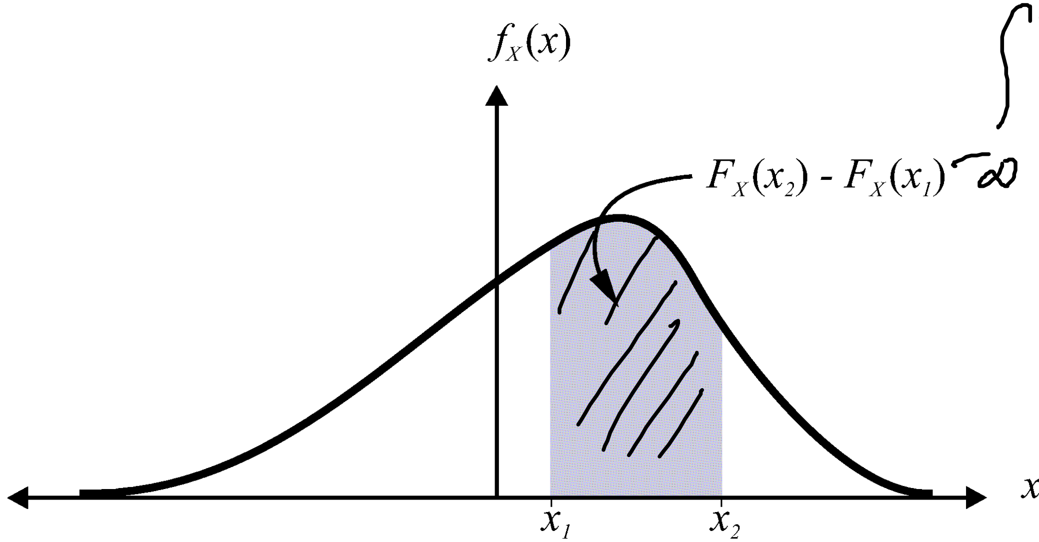
Properties of a PDF

Theorem 3.3

$$P[x_1 < X \le x_2] = \int_{x_1}^{x_2} f_X(x)\, dx.$$

$$P[X \le x_2] - P[X \le x_1]$$

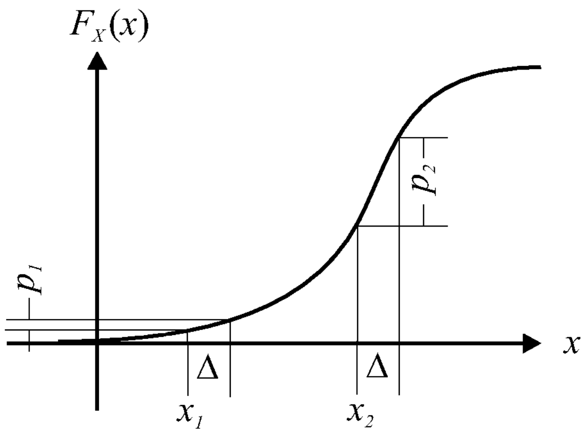
$$\int_{-\infty}^{x_2} f_X(x)\, dx - \int_{-\infty}^{x_1} f_X(x)\, dx$$



The PDF and CDF of X.

Properties of a PDF

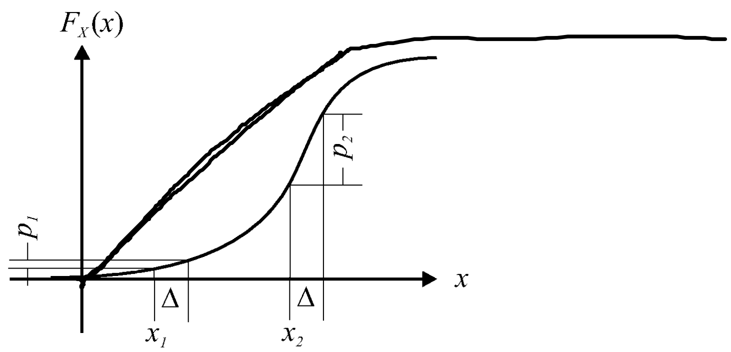
Figure 3.2



The graph of an arbitrary CDF $F_X(x)$.

$$P[x_2 < X \le x_2 + \Delta] = \frac{F_X(x_2+\Delta) - F_X(x_2)}{\Delta}\Delta$$

Figure 3.2



The graph of an arbitrary CDF $F_X(x)$.

$$P[x_2 < X \leq x_2 + \Delta] = \frac{F_X(x_2+\Delta) - F_X(x_2)}{\Delta}\Delta$$

- As the interval Δ becomes smaller

$$\frac{F_X(x_2+\Delta) - F_X(x_2)}{\Delta} = f_X(x_2)$$

- In general we can write for any x and an infinitesimally small *dx*

$$P[x < X \leq x + dx] = f_X(x)dx$$

Intervals and CRV(s)

- Consider the four different events
- A = (0,1)
- B = (0,1]
- C = [0,1)
- D = [0,1]
- They belong to the range space of a **continuous RV X**
- What can we say about P[A], P[B], P[C], and P[D]?
- Since X is continuous, P[A] = P[B] = P[C] = P[D]
  - This is true because probability of a point (outcome) is 0 for a continuous RV

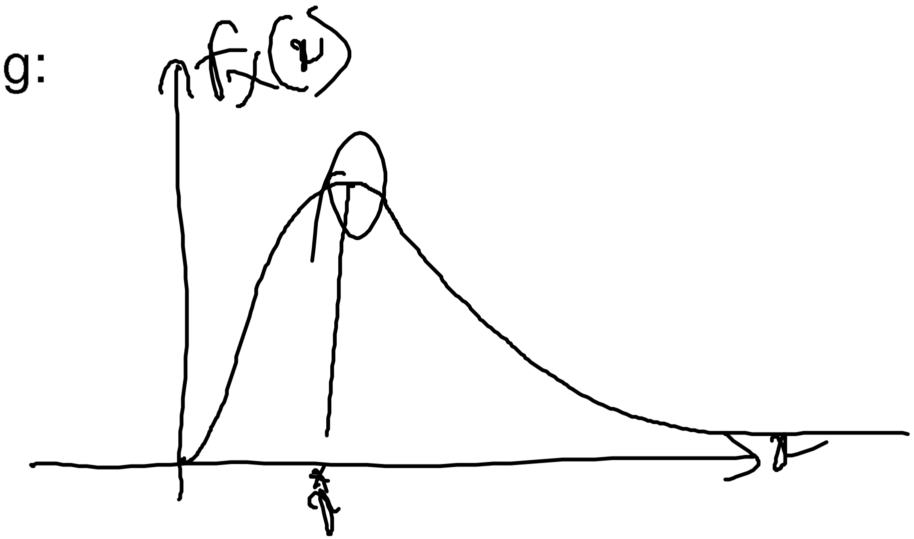
Problem on PDF

Quiz 3.2

Random variable X has probability density function

$$f_X(x) = \begin{cases} cxe^{-x/2} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.25}$$

Sketch the PDF and find the following:



(1) the constant c

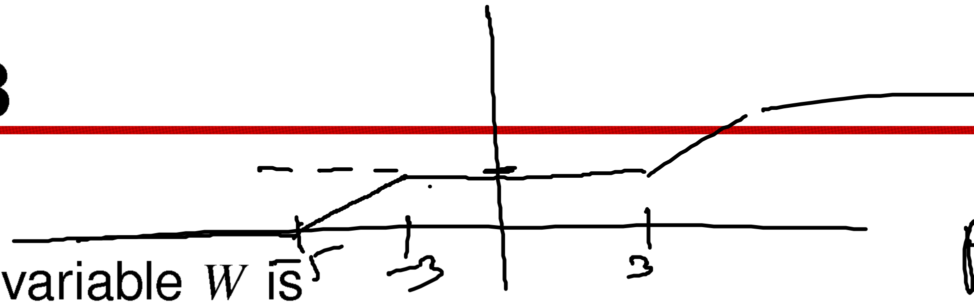
(2) the CDF $F_X(x)$

(3) $P[0 \leq X \leq 4]$

(4) $P[-2 \leq X \leq 2]$

(5) What is the mode of the random variable X?

Problem 3.1.3



$P[W \leq 3]$

$=$

The CDF of random variable W is

$y = (x+5)/8$

$$F_W(w) = \begin{cases} 0 & w < -5, \\ (w+5)/8 & -5 \leq w < -3, \\ 1/4 & -3 \leq w < 3, \\ 1/4 + 3(w-3)/8 & 3 \leq w < 5, \\ 1 & w \geq 5. \end{cases}$$

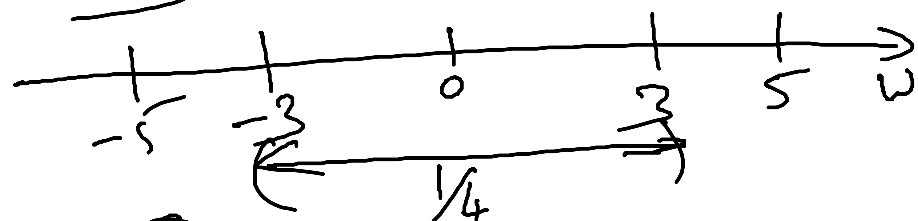
$P[W \leq w]$

(a) What is $P[W \leq 4]$?

(b) What is $P[-2 < W \leq 2]$?

(c) What is $P[W > 0]$?

(d) What is the value of a such that $P[W \leq a] = 1/2$?



$\frac{1}{4} + 3(w-3)/8 = 1/2$

$w = 11/3$

$P[W \leq 2.9] = 1/4$

Problem 3.2.4

For a constant parameter $a > 0$, a Rayleigh random variable X has PDF

$$f_X(x) = \begin{cases} a^2 x e^{-a^2 x^2/2} & x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

What is the CDF of X?

Expected Value

- **Def 3.4** The expected value of a continuous random variable X is

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

- Summation for the discrete case is replaced by integration for the continuous case
- **Theorem 3.4** The expected value of a function g(X) of a continuous random variable X is

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

Example of g(X)

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

$P[X \leq x] = x/L$ for $0 \leq x < L$. (and 0 otherwise)

What is $f_X(x)$? $\Rightarrow$ $f_X(x) = \begin{cases} \frac{1}{L} & 0 \leq x < L \\ 0 & \text{otherwise.} \end{cases}$

VVVS Problem

- What is $E[X]$?
  - $L/2$
- What is $E[X^2]$?
  - $L^2/3$
- What is $Var[X]$?
  - $L^2/3 - L^2/4 = L^2/12$

These Equalities are Valid for Both Continuous and Discrete RV(s)

Theorem 3.5

For any random variable X,

(a) $E[X - \mu_X] = \boxed{}$

(b) $E[aX + b] = \boxed{}$

(c) $\mathrm{Var}[X] = \boxed{}$

(d) $\mathrm{Var}[aX + b] = \boxed{\phantom{a^2 \mathrm{Var}[X]}}$

Problem 3.4.1

Radars detect flying objects by measuring the power reflected from them. The reflected power of an aircraft can be modeled as a random variable Y with PDF

$Y > E[Y]$

$$f_Y(y) = \begin{cases} \frac{1}{P_0} e^{-y/P_0} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $P_0 > 0$ is some constant. The aircraft is correctly identified by the radar if the reflected power of the aircraft is larger than its average value. What is the probability $P[C]$ that an aircraft is correctly identified?

$\frac{1}{P_0}\int_0^\infty e^{-y/P_0}\,dy = 1$ $\quad \int_0^\infty y\, f_Y(y)\,dy$ $\quad P_0$ $\quad e^{-1}$ $\quad \int_{P_0}^\infty f_Y(y)\,dy$

Families of Continuous RV(s)

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

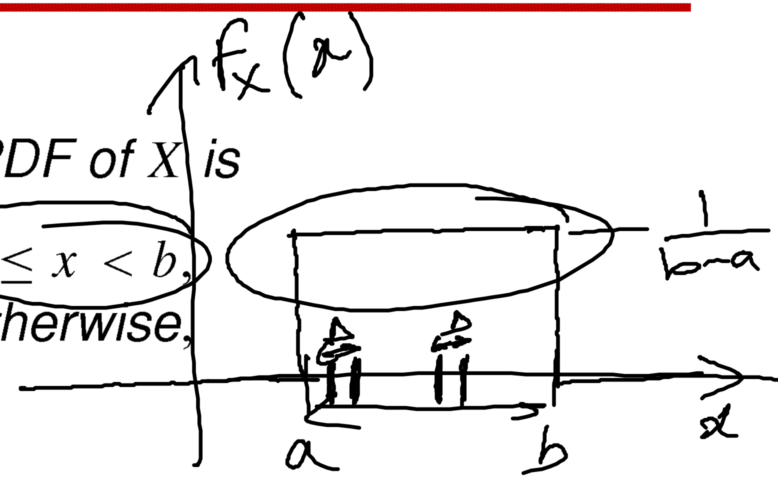
Uniform(a,b)

Definition 3.5 Uniform Random Variable

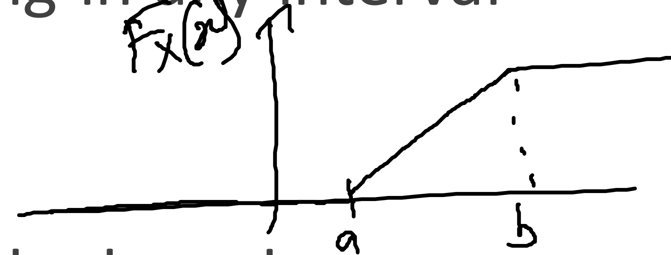
X is a uniform (a, b) random variable if the PDF of X is

$$f_X(x) = \begin{cases} 1/(b-a) & a \leq x < b, \\ 0 & \text{otherwise}, \end{cases}$$

where the two parameters are $b > a$.



- The probability of an outcome being in any interval of a given size Δ is the same
- How can we conclude this?
- The final position of the striker in the board example was *uniformly distributed*



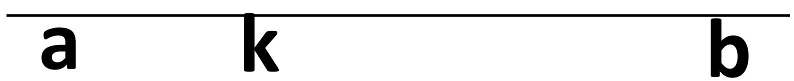
Discrete and Continuous Uniform RV(s)

Theorem 3.7

Let X be a uniform (a, b) random variable, where a and b are both integers. Let $K = \lceil X \rceil$. Then K is a ________ random variable.

- You want to describe the RV K
- K is a discrete RV. It is described completely by its PMF P[K = k]

$P[K=k]$



Given the ceil function, all x that are > k-1 and <= k correspond to the outcome {K=k}

$$P[K = k] = P[k - 1 < x \leq k] = \frac{1}{b-a}$$

Discrete and Continuous Uniform RV(s)

$$P[K = k] = P[k - 1 < x \leq k] = \frac{1}{b-a}$$

- The above is true for k = a+1, a+2,..., b
- P[K=k] = 0 otherwise.
- Clearly K is a discrete uniform RV

Exponential(λ)

Definition 3.6 *Exponential Random Variable*

X *is an* exponential (λ) random variable *if the PDF of* X *is*

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

where the parameter $\lambda > 0$.

A

$\cancel{f_X(x)}$

$$f_Y(y) = \begin{cases} c\lambda e^{-\lambda y} & 0 \leq y < a \\ 0 & \text{otherwise} \end{cases}$$

- The inter-arrival time between two packet arrivals at a server may be modeled as an Exp RV
- What is the average inter-arrival time?

$c = \frac{1}{A}$

Exponential and Geometric RV(s)

Theorem 3.9

If X is an exponential (λ) random variable, then $K = \lceil X \rceil$ is a geometric (p) random variable with $p = 1 - e^{-\lambda}$.

- Proof is similar to the case of Uniform distribution

Example 3.14 Problem

Phone company A charges \$$0.15$ per minute for telephone calls. For any fraction of a minute at the end of a call, they charge for a full minute. Phone Company B also charges \$$0.15$ per minute. However, Phone Company B calculates its charge based on the exact duration of a call. If T, the duration of a call in minutes, is an exponential $(\lambda = 1/3)$ random variable, what are the expected revenues per call $E[R_A]$ and $E[R_B]$ for companies A and B?

Erlang(n,λ)

Definition 3.7 Erlang Random Variable

X *is an* Erlang (n, λ) *random variable if the PDF of* X *is*

$$f_X(x) = \begin{cases} \dfrac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} & x \geq 0, \\ 0 & \textit{otherwise}. \end{cases}$$

where the parameter $\lambda > 0$, *and the parameter* $n \geq 1$ *is an integer.*

- The waiting time between n-occurrences of an event
- The waiting time between n customer arrivals
- n => Sum of n independent Exp(λ) Random Variables

Theorem 3.10

If X is an Erlang (n, λ) random variable, then

$X = X_1 + X_2 + \cdots + X_n$

$$E[X] = \frac{n}{\lambda}, \qquad \operatorname{Var}[X] = \frac{n}{\lambda^2}.$$

$X_1, X_2, \ldots, X_n$

$X_i \sim f_{X_i}(x) \sim \text{Exp}(\lambda)$

$E[X_i] = \frac{1}{\lambda}$

$E[X]$

$= E[X_1] + \cdots + E[X_n]$

$= \frac{n}{\lambda}$

$E[(X_1 + \cdots + X_n - E[X_1 + \cdots + X_n])^2]$

$= E[(X_1 + X_2 + \cdots + X_n)^2] - (E[(X_1 + \cdots + X_n)]^2$

Quiz 3.4

Continuous random variable X has $E[X] = 3$ and $\mathrm{Var}[X] = 9$. Find the PDF, $f_X(x)$, if

(1) X has an exponential PDF,

(a, b)

(2) X has a uniform PDF.

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

$$\frac{(b-a)^2}{12} = 9$$

$A, A^c \quad A = \{M \le 20\} \quad E[C_B] = E[C_B|A]P[A]$

Problem 3.4.9

$A^c = \{M > 20\}$

$+ \underbrace{E[C_B|A^c]}_{E[99+10(M-20)]} P[A^c]$

Long-distance calling plan A offers flat rate service at 10 cents per minute. Calling plan B charges 99 cents for every call under 20 minutes; for calls over 20 minutes, the charge is 99 cents for the first 20 minutes plus 10 cents for every additional minute. (Note that these plans measure your call duration exactly, without rounding to the next minute or even second.) If your long-distance calls have exponential distribution with expected value τ minutes, which plan offers a lower expected cost per call?

M

$E[C_A] = 10E[M] = 10\tau.$

$C_A = 10M$

$P[\underline{M \le 20}]$

$C_B = 99 + 10(M-20) \quad M \ge 20$

$P[M > 20]$

$E[C_B] = P[M \le 20]\,99 + P[M > 20]\,(\cancel{E}[99 + 10(M-20)])$

$99 +$

$P[M > 20] = e^{-\left(\frac{1}{\tau}\right)20}$

$(M - 20)\,10$

$M \sim Exp\left(\frac{1}{\tau}\right)$

$P[M - 20 > 10] = P[M > 30]$

$P[M > m] = e^{-M/\tau}$

$E[M - 20] = E[M'] = \int_0^\infty M'$

$f_{M \mid M > 20}(m)$

$C_B = 99 + (X - 20)^+ 10$

$E[C_B] = E[g(X)]$

$= E\left[99 + (X - 20)^+ 10\right]$

$\int_0^\infty \underbrace{g(x)}_{(x-20)^+} f_X(x)\,dx$

$= 99 + 10\,E\left[(X - 20)^+\right]$

$= \int_{20}^\infty (x - 20) f_X(x)\,dx$

Gaussian(μ,σ)

- The Bell-shaped distribution
- The Normal distribution
- Parameters are the mean μ and standard deviation is σ
  - Variance is σ^2
  - If X is Gaussian we often write X is N[μ, σ^2]

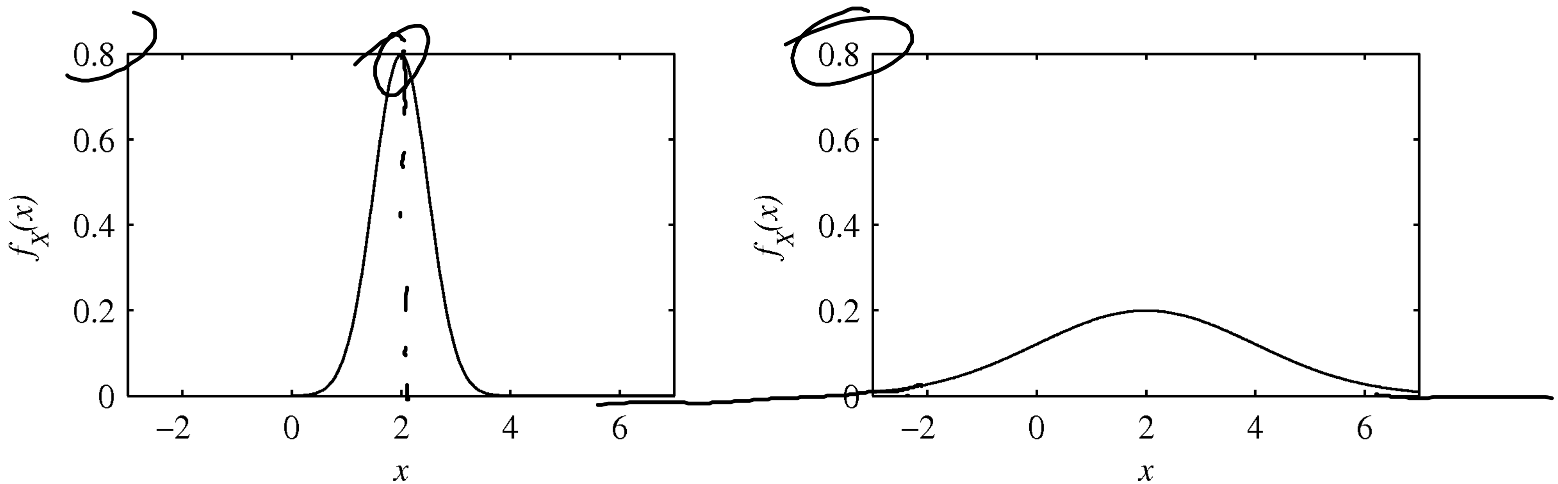
Definition 3.8 Gaussian Random Variable

X is a Gaussian (μ, σ) random variable if the PDF of X is

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

where the parameter μ can be any real number and the parameter $\sigma > 0$.

Figure 3.5



(a) $\mu = 2, \sigma = 1/2$ **(b)** $\mu = 2, \sigma = 2$

Two examples of a Gaussian random variable X with expected value μ and standard deviation σ.

Show that the Gaussian PDF is a valid PDF

- Let W be a Gaussian RV

$$\int_{-\infty}^{\infty} f_W(w)\,dw = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(w-\mu)^2/(2\sigma^2)}\,dw$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x)^2/(2\sigma^2)}\,dx$$

- The square of the integral is

$$I^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(y)^2/(2\sigma^2)}\,dy \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x)^2/(2\sigma^2)}\,dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} e^{(x^2+y^2)/(2\sigma^2)}\,dxdy$$

- Substitute $x=r\sin(\phi)$ and $y = r\cos(\phi)$ and show $I^2 = 1$

Derive $E[W]$ and $E[W^2]$

- Use steps similar to showing area is unity
- You will also need to apply integration by parts

Linear Transformation of a Gaussian

Theorem 3.13

If X is Gaussian (μ, σ), $Y = aX + b$ is .

$a^2\sigma^2$

- Linear transformation of a Gaussian gives another Gaussian!
- How do show the above?
- If X is not Gaussian, what are E[Y] and E[Y²]?

Y is Gaussian $(a\mu + b, a\sigma)$

Standard Normal Variable and CDF

- **Def 3.9** The standard random variable is Gaussian(0,1) – 0 mean and unit variance

$\int_{-\infty}^{x} f_X(x)\,dx$

(μ, σ)

$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)^2/(2\sigma^2)]}$

Definition 3.10 *Standard Normal CDF*

The CDF of the standard normal random variable Z *is*

$$\boxed{\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du.}$$

$E(\sigma Z + \mu)$

$X \sim \text{Gaussian}(\mu, \sigma)$ $\quad Z \sim N(0,1)$ $\quad \frac{X-\mu}{\sigma} = Z$

Expressing a Gaussian CDF as a N[0,1] CDF

Theorem 3.14

$P[X \le 100] = P[X - \mu_X \le 100 - \mu_X]$

$= P\left[\frac{X-\mu_X}{\sigma_X} \le \frac{100-\mu_X}{\sigma_X}\right]$

If X is a Gaussian (μ, σ) random variable, the CDF of X is

$$\boxed{F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).}$$

$X \sim N(2, 20)$

The probability that X is in the interval $(a, b]$ is

$$P[a < X \le b] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

$P[X \le 100] = \int_{-\infty}^{100} f_X(x)\, dx$

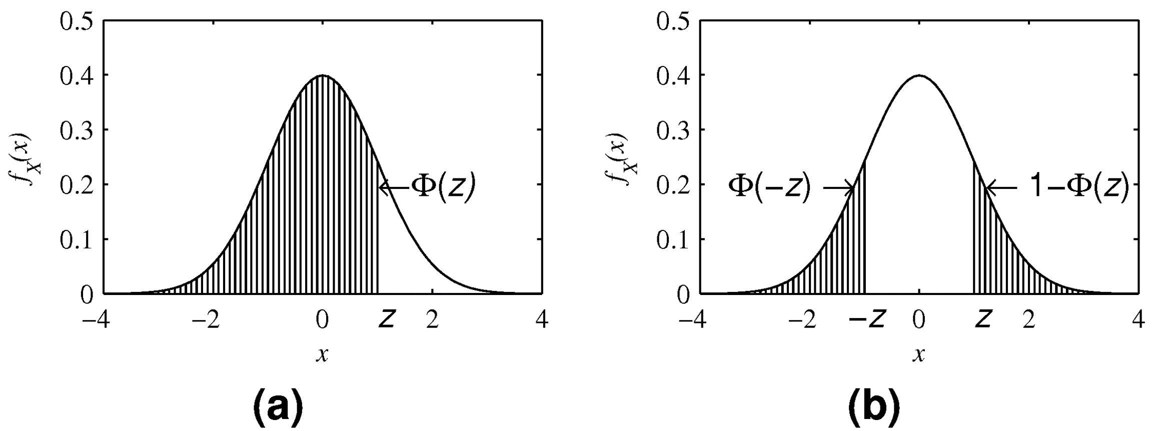
$\int_{-\infty}^{100} \frac{1}{\sqrt{2\pi(400)}} e^{-\left((x-2)^2/2(400)\right)} dx$

- Show using definition of CDF and substitution of variables in the integral
- All I need are values of ϕ(.)
  - This is important as CDF integral calculations for finite limits can only be done numerically
  - Thankfully, all we now need is a tabulation of ϕ(.)

Standard Normal CDF

- In fact $\phi(-z) = 1 - \phi(z)$
  - Why?
  - Is this true for any Gaussian Distribution?
  - When is it true?
- All I need is $\phi(z)$ for $z >= 0$

Figure 3.6



$$\Phi(-z) = \int_{-\infty}^{-z} f_Z(z)\,dz$$

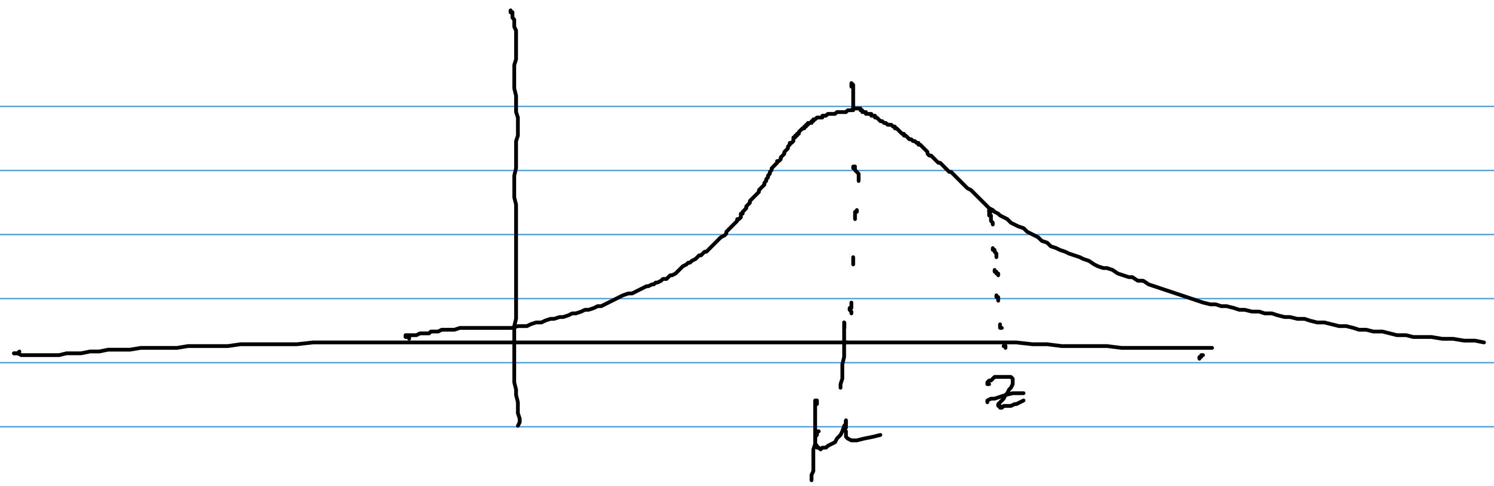
$$= \int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi(1)}} e^{-((z-0)^2/(2(1)))}\,dz$$

$$\int_{+\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}(-dy)$$

$$= \int_{-\infty}^{-z} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$$

$$1-\Phi(z) = \int_{-\infty}^{\infty} f_Z(z) - \int_{-\infty}^{z} f_Z(z)\,dz$$

$$\int_{z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$$



μ
z

Example 3.16 Problem

If X is the Gaussian $(61, 10)$ random variable, what is $P[X \leq 46]$?

$$F_X(46) = \Phi\left(\frac{46 - 61}{10}\right)$$

$$= \Phi(-1.5) = 1 - \Phi(1.5)$$

$$P[X \le 46] = P[X - \mu_x \le 46 - \mu_x]$$

$$= P\left[\frac{X - \mu_x}{\sigma_x} \le \frac{46 - \mu_x}{\sigma_x}\right]$$

$$= \Phi\left(\frac{46 - \mu_x}{\sigma_x}\right)$$

$$Z \sim N(0, 1)$$

$$\mu_Z \uparrow \quad \uparrow \sigma_Z$$

$$X \sim N(\mu_X, \sigma_X)$$

$$\frac{X - \mu_X}{\sigma_X} \to Z$$

Example 3.17 Problem

If X is a Gaussian random variable with $\mu = 61$ and $\sigma = 10$, what is $P[51 < X \leq 71]$?

$$P[51 < X \leq 71] = P[X \leq 71] - P[X \leq 51]$$

$$\frac{X - \mu}{\sigma_X} \qquad \frac{X - \mu}{\sigma_X^2}$$

Standard Normal CCDF

Definition 3.11

CCDF

The standard normal complementary CDF *is*

$$Q(z) = P[Z > z] = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-u^2/2}\, du = 1 - \Phi(z).$$

$$\Phi(3) = 0.9987, \;\; \Phi(4) = 0.9999768$$

$$Q(3) = 1.35 \times 10^{-3}, Q(4) = 3.17 \times 10^{-5}$$

Problem 3.5.6

A professor pays 25 cents for each blackboard error made in lecture to the student who points out the error. In a career of n years filled with blackboard errors, the total amount in dollars paid can be approximated by a Gaussian random variable Y_n with expected value $40n$ and variance $100n$. What is the probability that Y_{20} exceeds 1000? How many years n must the professor teach in order that $P[Y_n > 1000] > 0.99$?

$$P[Y_{20} > 1000] = P\left[\frac{Y_{20} - 40(20)}{10\sqrt{20}} > \frac{1000 - 40(20)}{10\sqrt{20}}\right]$$

$$P\left[\frac{Y_n - 40n}{10\sqrt{n}} > \frac{1000 - 40n}{10\sqrt{n}}\right] > 0.99$$

Problem 3.5.7

Suppose that out of 100 million men in the United States, 23,000 are at least 7 feet tall. Suppose that the heights of U.S. men are independent Gaussian random variables with a expected value of $5'10''$. Let N equal the number of men who are at least $7'6''$ tall.

(a) Calculate σ_X, the standard deviation of the height of men in the United States.

(b) In terms of the $\Phi(\cdot)$ function, what is the probability that a randomly chosen man is at least 8 feet tall?

(c) What is the probability that there is no man alive in the U.S. today that is at least $7'6''$ tall?

(d) What is $E[N]$?

$$P[X > 7'] = \frac{23000}{10^8}$$

$$P\left[\frac{X - 5'10''}{\sigma_X} > \frac{7' - 5'10''}{\sigma_X}\right] = 23 \times 10^{-5}$$

- Six Sigma Event?
  - Use MATLAB's qfunc

$(\mu - 6\sigma, \mu + 6\sigma)$

$(\mu - \sigma, \mu + \sigma)$

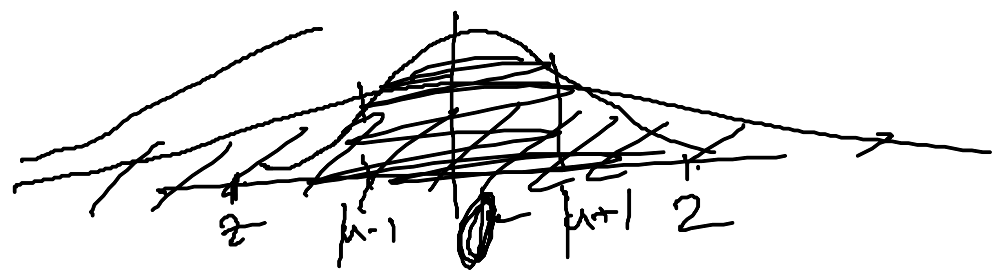
$(\mu - 2\sigma, \mu + 2\sigma)$

Quiz 3.5

X is the Gaussian $(0, 1)$ random variable and Y is the Gaussian $(0, 2)$ random variable.

$(\mu - \sigma, \mu + \sigma)$

(1) Sketch the PDFs $f_X(x)$ and $f_Y(y)$ on the same axes.



(2) What is $P[-1 < X \leq 1]$?

(3) What is $P[-1 < Y \leq 1]$?

(4) What is $P[X > 3.5]$?

(5) What is $P[Y > 3.5]$?

Conditioning on a Continuous RV

Definition 3.15 *Conditional PDF given an Event*

For a random variable X *with PDF* $f_X(x)$ *and an event* $B \subset S_X$ *with* $P[B] > 0$, *the* conditional PDF of X given B *is*

$$f_{X|B}(x) = \begin{cases} \dfrac{f_X(x)}{P[B]} & x \in B, \\ 0 & \textit{otherwise.} \end{cases}$$

$f_{X|B}(x)\,\Delta x$
$= P[X \le x + \Delta x \mid B]$
$- P[X \le x \mid B]$

- Note that the conditional pdf can be written as

$$f_{X|B}(x)\,dx = P[x \le X < x + dx | B]$$
$$= \frac{P[x \le X < x + dx, B]}{P[B]}$$

$X \sim f_X(x), \quad x \in R \qquad X \in (-\infty, x]$

B

$$f_{X|B}(x) = \frac{d}{dx} P[\{X \le x\} | B]$$

$$f_{X|B}(x) \simeq \frac{P[X \le x + \Delta x | B] - P[X \le x | B]}{\Delta x}$$

$$f_X(x)\,\Delta x = \frac{P[x \le X < x + \Delta x | B]}{\Delta x}$$

$(x, x+\Delta x)$

$$\frac{P[x \le X \le x + \Delta x, B]}{P[B]\,\Delta x}$$

$x \in B$

Conditioning on a Continuous RV

- What if the event {x <= X < x + dx} is in B?

$$f_{X|B}(x)\,dx = \frac{P[x \leq X < x + dx]}{P[B]}$$
$$= \frac{f_X(x)\,dx}{P[B]}$$

- = 0, when event {x <= X < x + dx} is **NOT** in B
- The definition comprises of the above two cases

Theorem 3.23

$P[X \in (x, x+\Delta x)] = \sum_i P[X \in (x, x+\Delta x), B_i]$

Given an event space $\{B_i\}$ and the conditional PDFs $f_{X|B_i}(x)$,

$$f_X(x) = \sum_i f_{X|B_i}(x) P[B_i].$$

Δt

t

0

$V=0$

a

$a\Delta t$

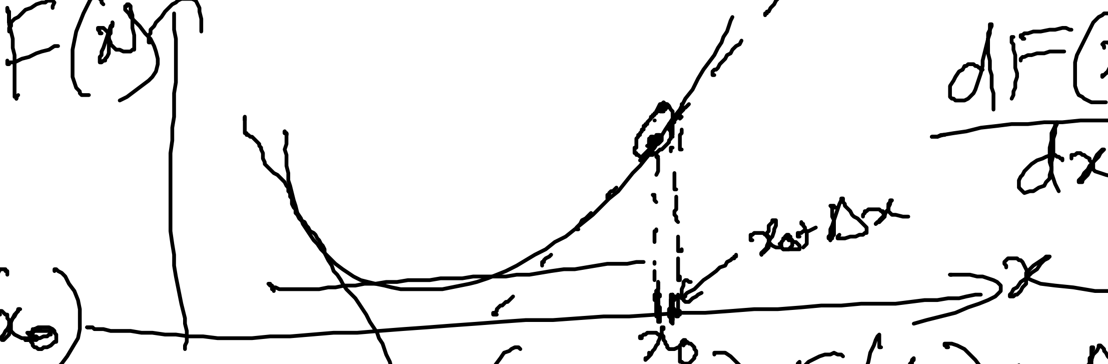
$$\frac{P[x \le X < x + \Delta x]}{\Delta x}$$

$\Delta x \, P[x \le X < x+\Delta x \mid B_i]$

$P[B_i]$

$X \in (x, x+\Delta x)$

$F(x_0+\Delta x) - F(x_0) \simeq \Delta x \, F'(x_0)$



$\left.\frac{dF(x)}{dx}\right|_{x=x_0}$

$F(x_0+\Delta x) = F(x_0) + \Delta x \, F'(x_0)$

Example 3.32 Problem

Continuing Example 3.3, when symbol “0” is transmitted (event B_0), X is the Gaussian $(-5, 2)$ random variable. When symbol “1” is transmitted (event B_1), X is the Gaussian $(5, 2)$ random variable. Given that symbols “0” and “1” are equally likely to be sent, what is the PDF of X?

$$f_{X|B_0}(x) = \frac{1}{\sqrt{2\pi(2)^2}} e^{-(x-(-5))^2/2(2)^2}$$

$f_{X|B_1}$

$$f_X(x) = f_{X|B_0}(x)P[B_0] + f_{X|B_1}(x)P[B_1] \qquad x \in R$$

Conditional Expected Value

Definition 3.16 Given an Event

If $\{x \in B\}$, *the conditional expected value of* X *is*

$$E[X|B] = \int_{-\infty}^{\infty} x f_{X|B}(x)\, dx.$$

$Y = g(X)$

$E[X]$

$E[Y|B]$

$= \int_{-\infty} g(x) f_X(x)\, dx$

Y=g(X), where X is a Continuous Random Variable

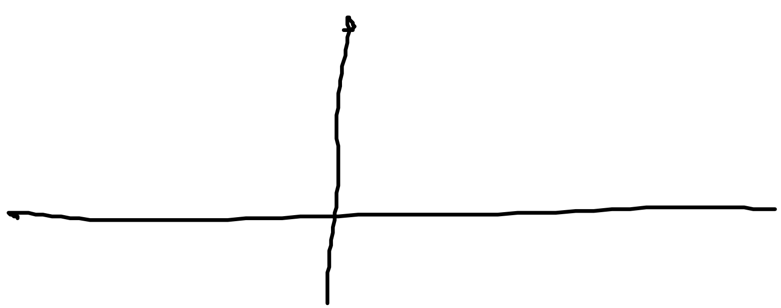
- We want to find $f_Y(y)$
- You want a complete probability description of Y
  - Just the expectation or some moment of Y is not enough
- Two step procedure
  - Find the CDF $F_Y(y)$
  - Find $f_Y(y)$. How??

$$Y = aX + b + 2b^2$$

$$Y = aX$$

$$P[Y \le y] = P[aX \le y] = P[X \le y/a]$$

$$= F_X(y/a)$$



Scaling a Random Variable

Theorem 3.19

If $Y = aX$, where $a > 0$, then Y has CDF and PDF

$$F_Y(y) = F_X(y/a), \qquad f_Y(y) = \frac{1}{a} f_X(y/a).$$

- We want P[Y<=y]
- P[Y <= y] = P[aX <= y] = P[X <= y/a] = F_X(y/a)

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(y/a)$$

$$= \frac{d}{dx}F_X(y/a)\frac{d}{dy}(y/a) = \frac{1}{a}f_X(y/a)$$

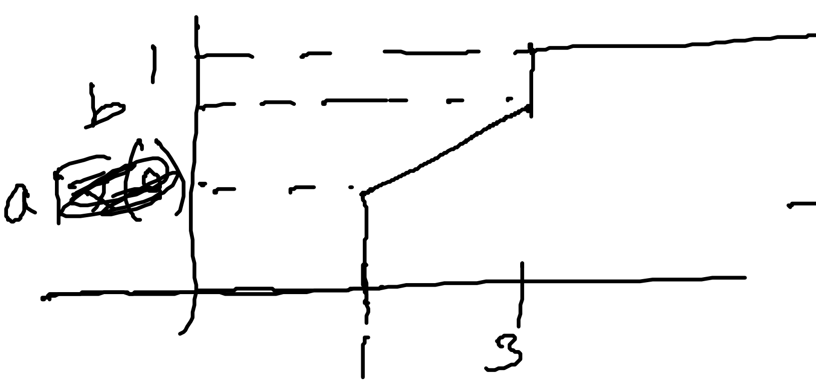
Adding a Constant

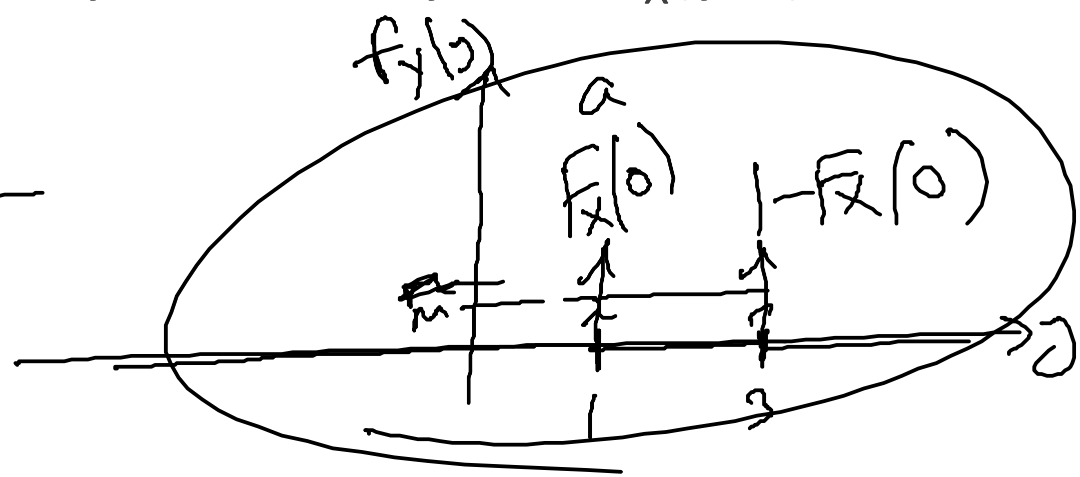
Theorem 3.21

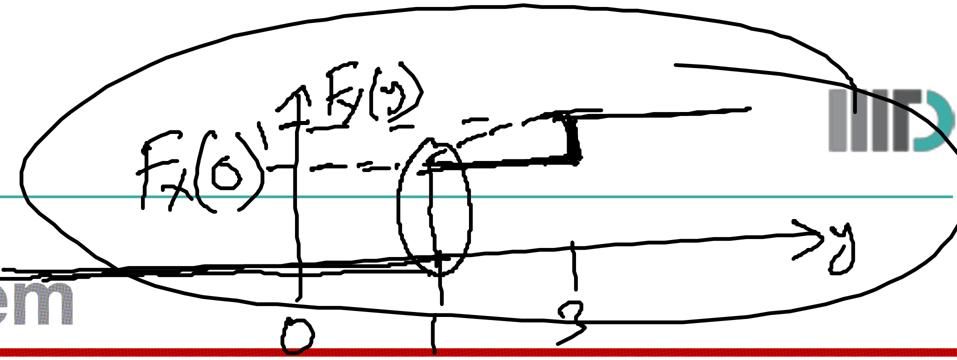
If $Y = X + b$,

$$F_Y(y) = F_X(y - b), \qquad f_Y(y) = f_X(y - b).$$

- We want P[Y<=y]
- P[Y <= y] = P[X + b <= y] = P[X <= y-b] = F_X(y-b)
- f_Y(y) = f_X(y-b)





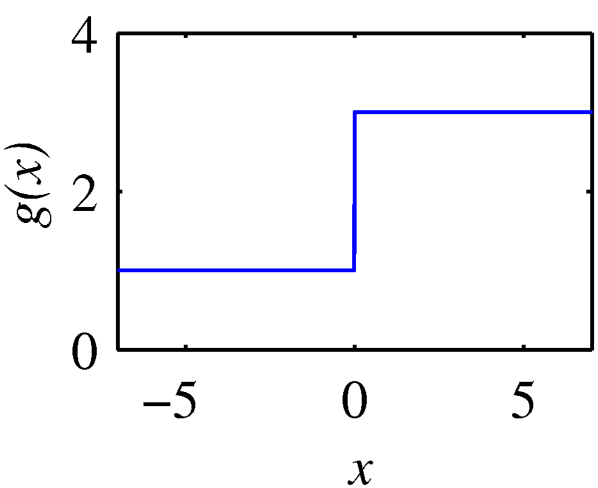


Example 3.24 Problem

Let X be a random variable with CDF $F_X(x)$. Let Y be the output of a clipping circuit with the characteristic $Y = g(X)$ where

$P[Y \le 3] =$

$P[Y \le y]$



$P[Y \le 1]$

$$g(x) = \begin{cases} 1 & x \le 0, \\ 3 & x > 0. \end{cases}$$

$= P[X \le 0] = F_X(0)$

Express $F_Y(y)$ and $f_Y(y)$ in terms of $F_X(x)$ and $f_X(x)$.

$P[Y \le 2] = F_X(0)$ $\quad P[Y < 1] = 0$

Detour...

Definition 3.12 Unit Impulse (Delta) Function

Let

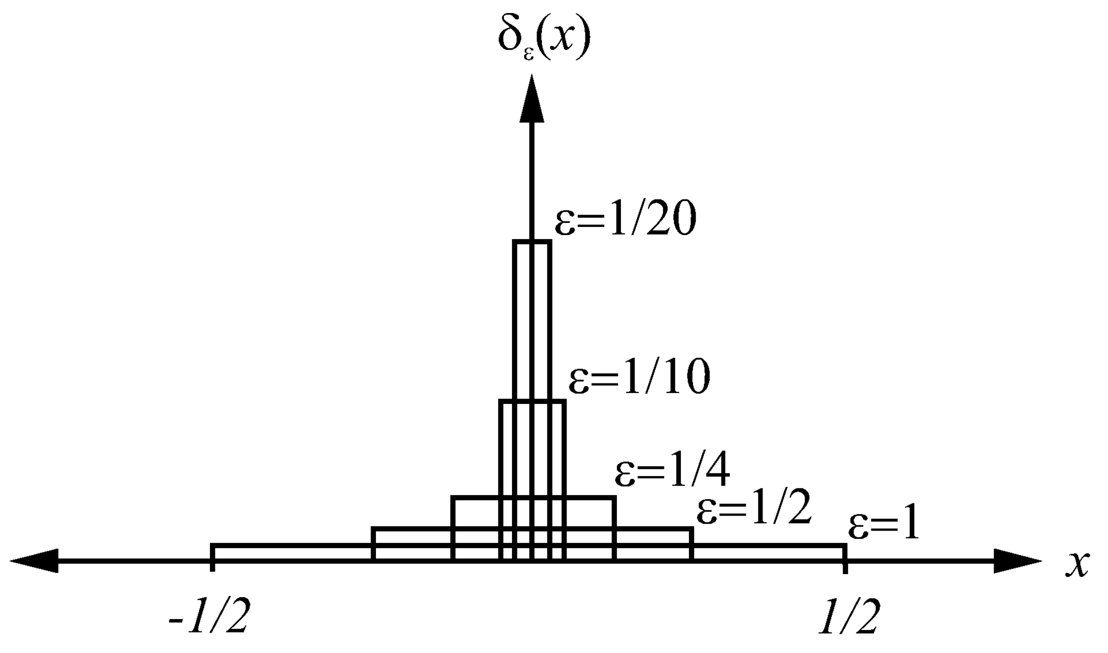
$$d_\epsilon(x) = \begin{cases} 1/\epsilon & -\epsilon/2 \leq x \leq \epsilon/2, \\ 0 & \textit{otherwise.} \end{cases}$$

The unit impulse function *is*

$$\delta(x) = \lim_{\epsilon \to 0} d_\epsilon(x).$$

Delta Function

Figure 3.7



As $\epsilon \to 0$, $d_\epsilon(x)$ approaches the delta function $\delta(x)$. For each ϵ, the area under the curve of $d_\epsilon(x)$ equals 1.

Properties of the Delta Function

- The delta function δ(x) is infinite at x=0 and zero everywhere else.
- The area under the delta function is unity

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1$$

- It has the sifting property. For a continuous function f(x)

$$\int_{-\infty}^{\infty} f(x)\delta(x - x_0)\, dx = f(x_0)$$

- This is because $\delta(x-x_0)$ is $\delta(x)$ shifted by x_0. Specifically, we can write

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} f(x) d_\epsilon(x - x_0)\, dx = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_{x_0-\epsilon/2}^{x_0+\epsilon/2} f(x)\, dx$$

Properties of the Delta Function

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} f(x) d_\epsilon(x - x_0)\, dx = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \int_{x_0 - \epsilon/2}^{x_0 + \epsilon/2} f(x)\, dx$$

- The integral on the right is the average of the function f(x) over the interval of length Є
- The average -> the constant f(x_0) as Є -> 0
- Thus

$$\int_{-\infty}^{\infty} f(x) \delta(x - x_0)\, dx = f(x_0)$$

The Unit Step Function

Definition 3.13 *Unit Step Function*

The unit step function *is*

$$u(x) = \begin{cases} 0 & x < 0, \\ 1 & x \geq 0. \end{cases}$$

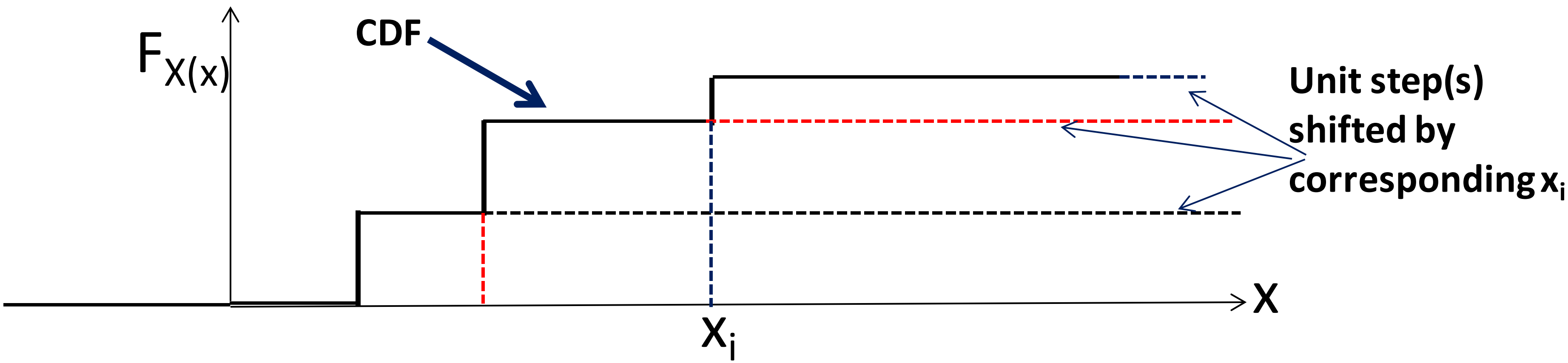
Note that u(0) = 1. This is by definition and leads to a jump (discontinuity) at x=0

Theorem 3.17

$$\int_{-\infty}^{x} \delta(v)\, dv = u(x).$$

In other words the delta function is the derivative of the unit step function

CDF of a Discrete RV

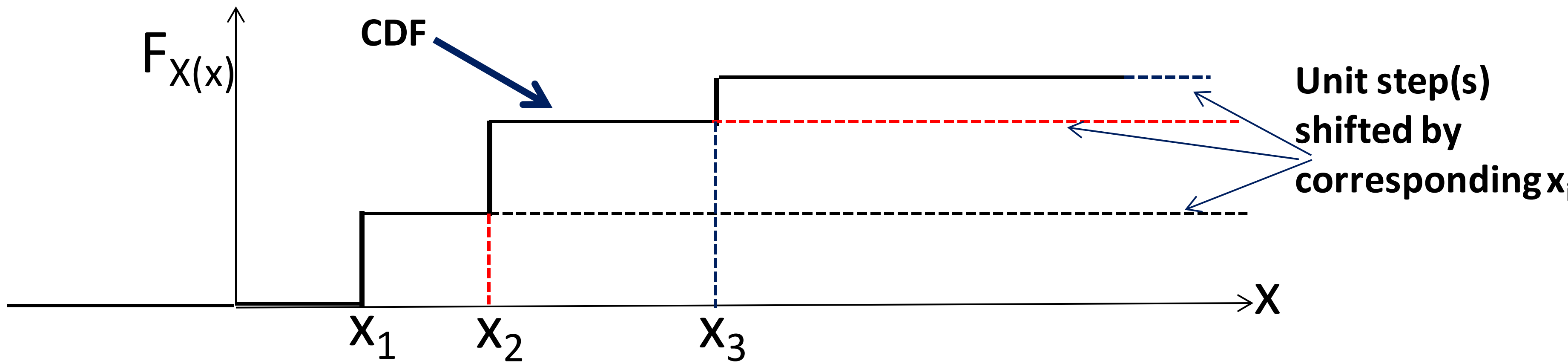


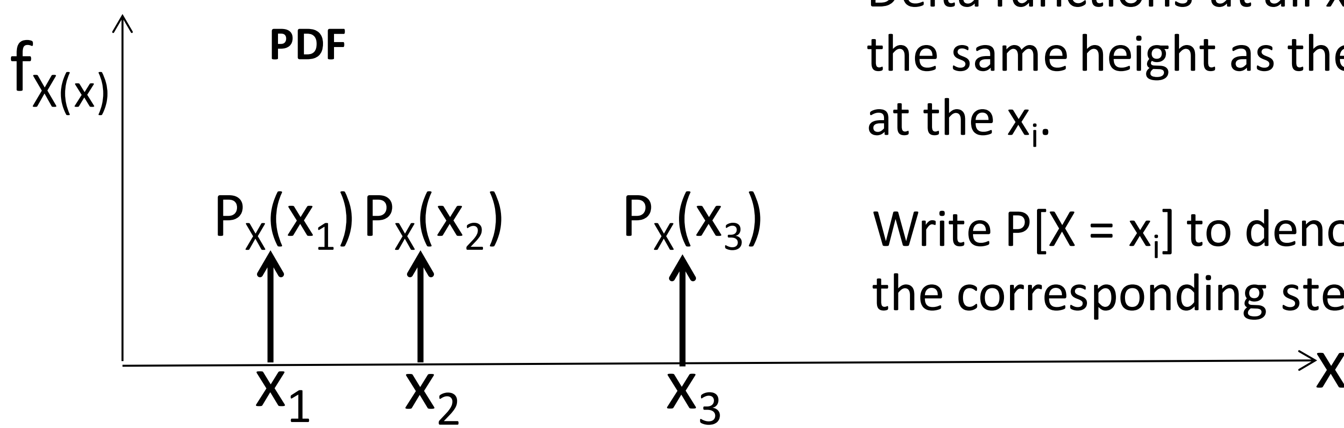
$$F_X(x) = \sum_{x_i \in S_X} P_X(x_i)\,u(x - x_i)$$

For any given x, $u(x - x_i) = 1$ for $x >= x_i$

PDF of a Discrete RV

- Having defined the unit step, we can now define a pdf for a discrete random variable!





Delta functions at all x_i are shown of the same height as they all are infinite at the x_i.

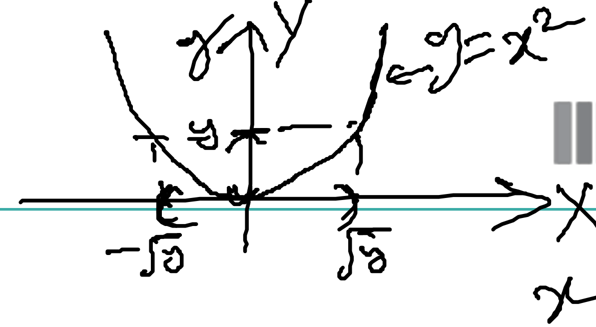
Write $P[X = x_i]$ to denote the size of the corresponding step in the CDF.

- Jumps in CDF correspond to delta functions in the pdf!

$Y = X^2$

- X has a discrete uniform distribution and the range $S_X = \{1,2,3,4,5,6\}$
- What is $F_Y(y)$ and $f_Y(y)$?
  - $S_Y = \{1,4,9,16,25,36\}$
  - $P_Y(y) = 1/6$, the pdf will contain impulses with weight 1/6 at each point in S_Y
- X has a discrete uniform distribution and the range $S_X = \{-2, -1, 0, 1,2,3\}$
- What is $F_Y(y)$ and $f_Y(y)$?
  - $S_Y = \{0, 1, 4, 9\}$
  - $P_Y(y) = 1/6, 1/3, 1/3, 1/6$ for $Y = 0,1,4$, and 9 respectively

$Y = X^2$



- For a continuous RV X
- $P[Y <= y] = P[X^2 <= y] = 0$, for $y < 0$.
- For $y >= 0$ we have

$$P[Y \leq y] = P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}]$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

- Therefore

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}, y > 0$$

- $f_Y(y) = 0$ for $y <= 0$.

$y \in (0,1) \rightarrow P[Y \leq y]$

$\downarrow$

$P[Y \in (0,y)]$

$y \in (1,9)$

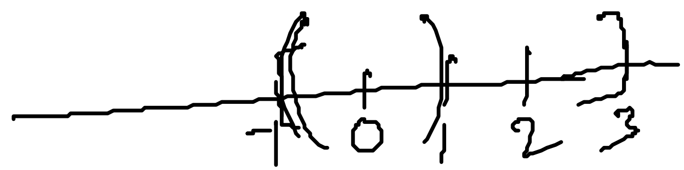
$Y = X^2$

$P[Y \le y]$ a $P[Y \le 0]$

$P[Y \le 9]$ $X \in (-\sqrt{y}, \sqrt{y})$

Example 3.26 Problem

Suppose X is uniformly distributed over $[-1, 3]$ and $Y = X^2$. Find the CDF $F_Y(y)$ and the PDF $f_Y(y)$.



- We know

$$P[Y \le y] = P[X^2 \le y] = P[-\sqrt{y} \le X \le \sqrt{y}]$$
$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

- Therefore

$$P[Y \le y] = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)\, dx$$

$Y = X^2$

- However X is Uniform over [-1,3]
  - While [-1,1] is a subset of S_X, the elements in the set (1,3] have no corresponding negative values

- Therefore, for y > 0

$$P[Y \leq y] = \int_{\max(-1,-\sqrt{y})}^{\sqrt{y}} f_X(x)\, dx$$

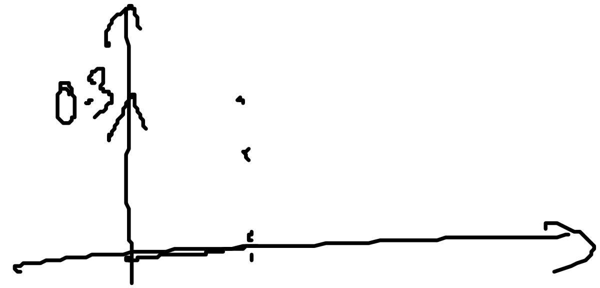
- For y < 0, we have P[Y <= y] = 0.

$$P[Y \leq y] = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)\, dx = \frac{\sqrt{y}}{2}, 0 \leq y \leq 1$$

$Y=X^2$

- Also

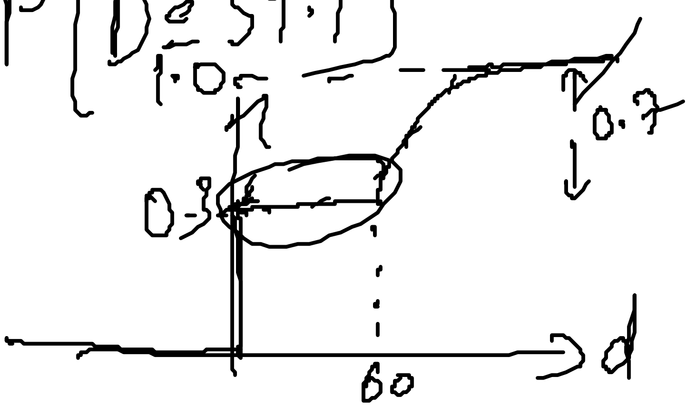
$$P[Y \leq y] = \int_{-1}^{\sqrt{y}} f_X(x)\,dx = \frac{\sqrt{y}+1}{4}, 1 < y \leq 9$$



Problem 3.6.8

With probability 0.7, the toss of an Olympic shot-putter travels $D = 60 + X$ feet, where X is an exponential random variable with expected value $\mu = 10$. Otherwise, with probability 0.3, a foul is committed by stepping outside of the shot-put circle and we say $D = 0$. What are the CDF and PDF of random variable D?

$P[D \le 59.9]$

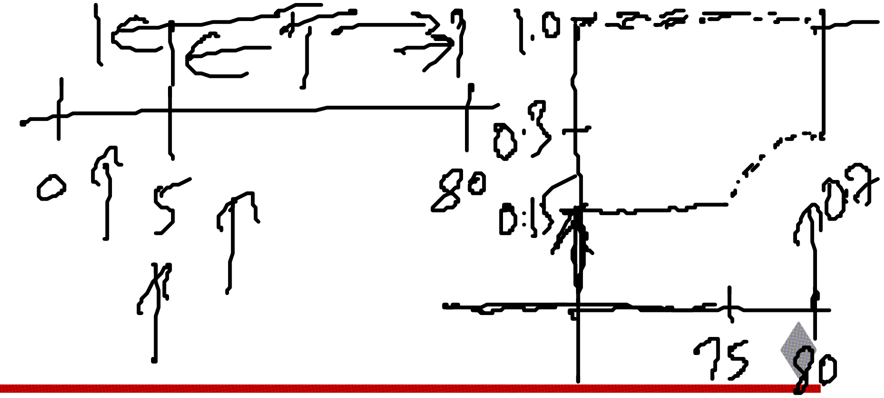


$P[D \le d] = 0 \qquad d < 0$

$P[D \le d] = P[D < 0] + P[D = 0] \qquad d \le 0$

$= 0.3$

$T=0 \Leftrightarrow \{L \geq 5\}$

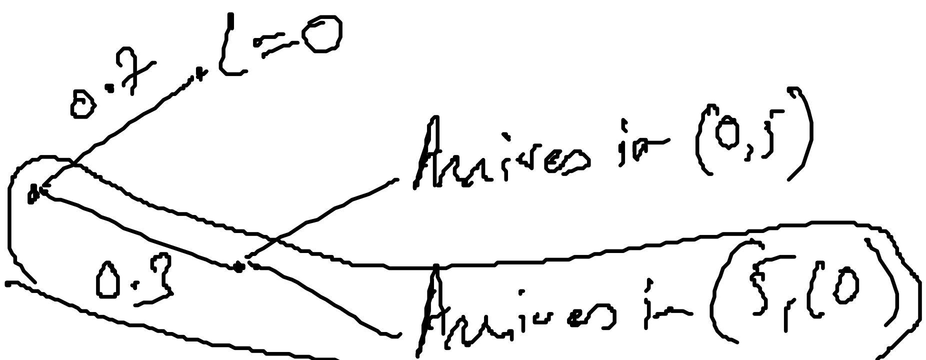


Problem 3.6.9

For 70% of lectures, Professor Y arrives on time. When Professor Y is late, the arrival time delay is a continuous random variable uniformly distributed from 0 to 10 minutes. Yet, as soon as Professor Y is 5 minutes late, all the students get up and leave. (It is unknown if Professor Y still conducts the lecture.) If a lecture starts when Professor Y arrives and always ends 80 minutes after the scheduled starting time, what is the PDF of T, the length of time that the students observe a lecture.

$P[T<0]=0$

$P[T\leq 0]=0.15$



$(0.3)(0.5)$

Problem 3.2.5

For constants a and b, random variable X has PDF

$$f_X(x) = \begin{cases} ax^2 + bx & 0 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

What conditions on a and b are necessary and sufficient to guarantee that $f_X(x)$ is a valid PDF?

Problem 3.4.14

This problem outlines the steps needed to show that a nonnegative continuous random variable X has expected value

$$E[X] = \int_0^\infty x f_X(x)\, dx = \int_0^\infty [1 - F_X(x)]\, dx.$$

Problem 3.5.10

In mobile radio communications, the radio channel can vary randomly. In particular, in communicating with a fixed transmitter power over a "Rayleigh fading" channel, the receiver signal-to-noise ratio Y is an exponential random variable with expected value γ. Moreover, when $Y = y$, the probability of an error in decoding a transmitted bit is $P_e(y) = Q(\sqrt{2y})$ where $Q(\cdot)$ is the standard normal complementary CDF. The average probability of bit error, also known as the bit error rate or BER, is

$$\overline{P}_e = E\left[P_e(Y)\right] = \int_{-\infty}^{\infty} Q(\sqrt{2y}) f_Y(y)\, dy.$$

Find a simple formula for the BER $\overline{P}_e$ as a function of the average SNR γ.

Birthday Problem (Example 3.5 RN)

- Assume that birthdays occur uniformly over all 365 days of the year. In a group of k people what is the probability that no two people have the same birthday?
- HW: Plot this probability as a function of k
- Probability is the ratio of the number of favorable outcomes to total number of possibilities, which is

$$\frac{(n)_k}{n^k}$$

$$\frac{365 \times 364 \times \cdots (365-k+1)}{(365)^k}$$

Coincidences

- Generalization of the birthday problem
- You have n possible events
  - n=365 in birthday problem
- In a set of k events, the probability that no two events are the same is

$$\frac{(n)_k}{n^k} = \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right)$$

- Fix k and let n -> ∞
  - The probability goes to 1!
- Note as n becomes large for a fixed k, it does not matter whether you permute with or without replacement
  - Even when you permute with replacement the probability of picking the same item from a large number n of items is small for small k

Monty Hall Problem

http://mathworld.wolfram.com/MontyHallProblem.html

$\to P[\text{MH picks } 3 \mid \text{Car behind } 2, \text{You pick } 1]\, P[\text{Car is behind } 2 \mid \text{You pick } 1]$

- Assume that a room is equipped with three doors. Behind two are goats, and behind the third is a shiny new car. You are asked to pick a door, and will win whatever is behind it. Let's say you pick door 1.
- Before the door is opened, however, someone who knows what's behind the doors (Monty Hall) opens *one of the other* two doors, revealing a goat, and asks you if you wish to change your selection to the third door (i.e., the door which neither you picked nor he opened). The Monty Hall problem is deciding whether you do.

$P[1] = P[2] = P[3] = 1/3$

1 2 3

You pick

Picks

$$P[\text{Car is behind } 2 \mid \text{You pick } 1, \text{MH picks } 3] = \frac{P[\text{Car is behind } 2, \text{MH picks } 3 \mid \text{You pick } 1]}{P[\text{MH picks } 3 \mid \text{You pick } 1]}$$

Summary For a RV X

- CDF is given by $F_X(x) = P[X <= x]$
  - Nondecreasing
  - Starts at 0 and ends at 1
  - Is continuous for a continuous RV
  - Contains steps for a discrete RV. Step size at x is the PMF at x, $P_X(x) = P[X=x]$
- PDF $f_X(x)$ is the derivative of $F_X(x)$
  - Always >= 0
  - It is the slope of the CDF at x
  - Area under PDF is 1

$$\begin{aligned} P[x_1 \leq X \leq x_2] &= P[X \leq x_2] - P[X < x_1] \\ &= F_X(x_2) - F_X(x_1) \\ &= \int_{x_1}^{x_2} f_X(x)\, dx \end{aligned}$$

Summary For a RV X

- Moments of RV X

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

- When conditioning on an event E

$$P_X(x) = P_{X|B}(x)P[B] + P_{X|B^c}(x)P[B^c]$$

$$F_X(x) = F_{X|B}(x)P[B] + F_{X|B^c}(x)P[B^c]$$

$$f_X(x) = f_{X|B}(x)P[B] + f_{X|B^c}(x)P[B^c]$$

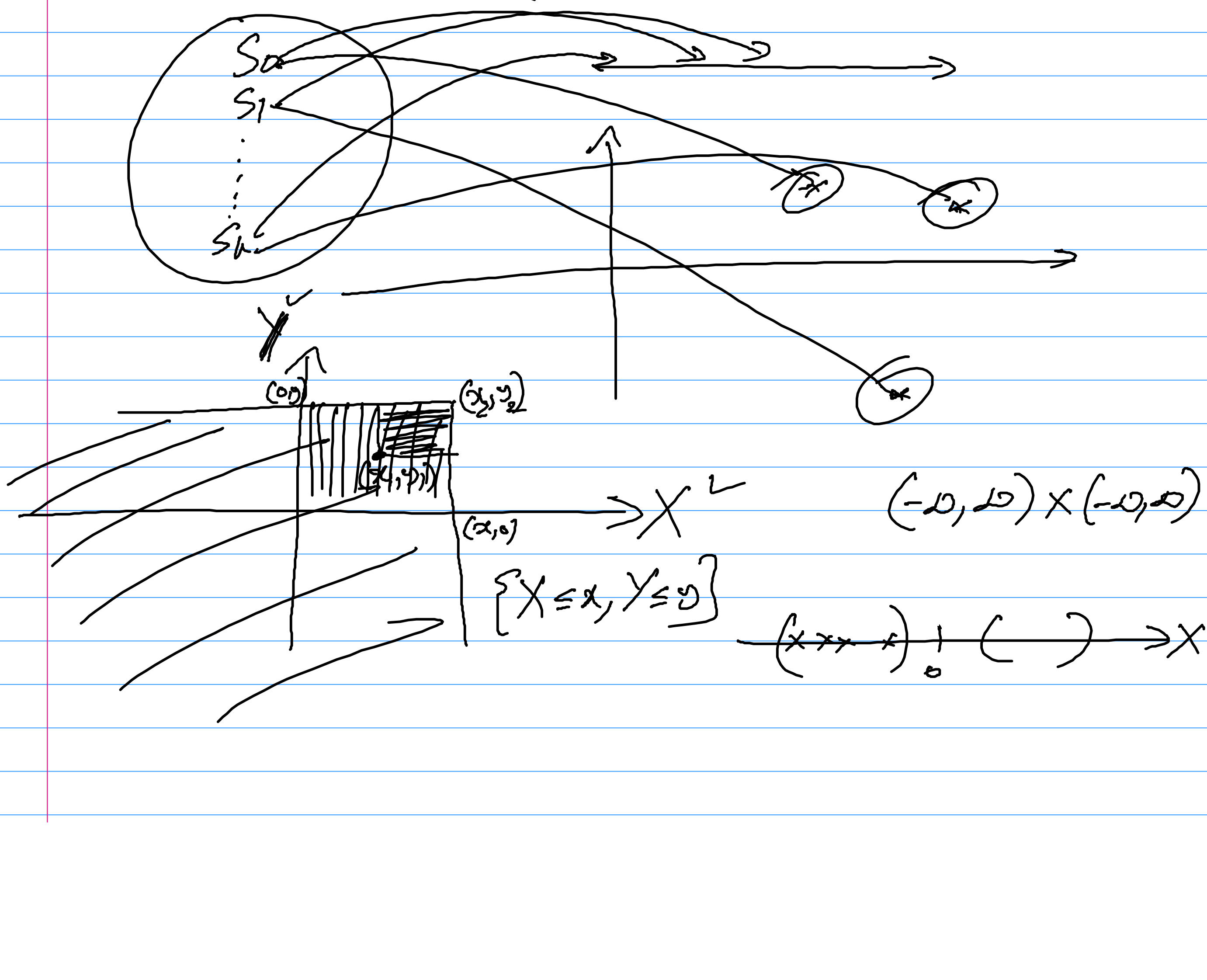
$$E[g(x)] = E[g(x)|B]P[B] + E[g(x)|B^c]P[B^c]$$

Pairs of Random Variables

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

$s \in S$

$X(s)$



S_0
S_1
S_n
Y
X
$(0,y)$
(x_2,y_2)
(x_1,y_1)
$(x,0)$
$\{X \le x, Y \le y\}$
$(-\infty,\infty)\times(-\infty,\infty)$
0
X

We started with…

- An Experiment that contains a
  - Procedure, Observations, and Model
- Later we mapped outcomes to numbers
  - One or more outcomes to a point on the line
- We called the numbers random variables
- We will now map an outcome to a pair of numbers
  - One or more outcomes to a point on a plane
- The numbers in the pair correspond to RVs X and Y
- For example, a transmitted sinusoid that is received with a random amplitude and random phase
  - Outcomes can be described by X = amplitude and Y = phase

Pair of RVs

- We defined a CDF for a single RV X
- For the pair of RVs we define a ***joint*** CDF

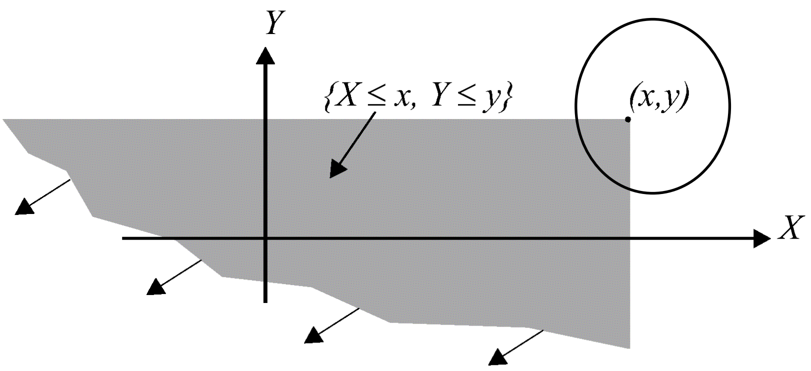
Definition 4.1 ***Joint Cumulative Distribution Function (CDF)***

The joint cumulative distribution function *of random variables* X *and* Y *is*

$$F_{X,Y}(x, y) = P[X \le x, Y \le y].$$

Joint CDF

Figure 4.1



The area of the (X, Y) plane corresponding to the joint cumulative distribution function $F_{X,Y}(x, y)$.

- We are interested in the probability of the intersection of the events {X<=x} and {Y<=y}

Joint CDF

S_Y

Theorem 4.1

$\{X \le x\}$ $\{Y \le y,\ Y > y\}$

For any pair of random variables, X, Y,

(a) $\boxed{0} \le F_{X,Y}(x, y) \le \boxed{1}$ $\{X \le x \cap Y \le y\} \cup$

(b) $F_X(x) = \boxed{F_{X,Y}(x,\infty)}$, $\{X \le x, Y \le \infty\} = \{X \le x\}$ $\{X \le x, Y > y\}$ $P[X \le x, Y < \infty]$ $P[Y \le \infty]$ $= P[X \le x]$

(c) $F_Y(y) = \boxed{}$,

$\{X \le -\infty, Y \le y\}$

(d) $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = \boxed{}$ $A \subseteq B$

$P[X \le x, Y \le y]$ $P[X \le x_1, Y \le y_1]$

(e) If $x \le x_1$ and $y \le y_1$, then $F_{X,Y}(x, y) \boxed{\le} F_{X,Y}(x_1, y_1)$, $P[A] \le P[B]$

(f) $F_{X,Y}(\infty, \infty) = \boxed{1}$.

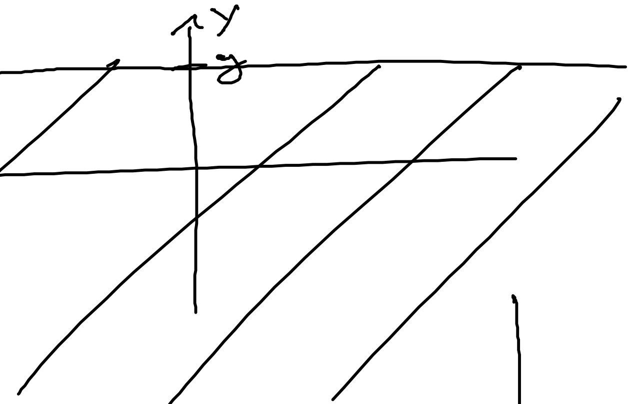
(g) $F_{X,Y}(\infty, -\infty) = ?$

Joint CDF

Theorem 4.1

For any pair of random variables, X, Y,

(a) $0 \leq F_{X,Y}(x, y) \leq 1$,



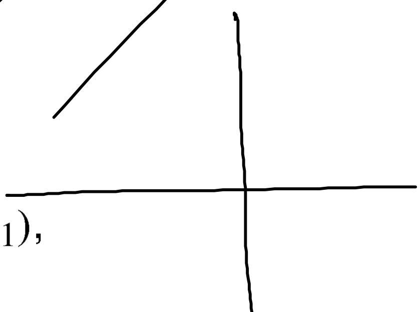
(b) $F_X(x) = F_{X,Y}(x, \infty)$,

(c) $F_Y(y) = F_{X,Y}(\infty, y)$,

$P[X \leq -\infty, Y \leq y]$

(d) $\underline{F_{X,Y}(-\infty, y)} = \underline{F_{X,Y}(x, -\infty)} = 0$,

(e) If $x \leq x_1$ and $y \leq y_1$, then $F_{X,Y}(x, y) \leq F_{X,Y}(x_1, y_1)$,



(f) $F_{X,Y}(\infty, \infty) = 1$.

(g) $F_{X,Y}(\infty, -\infty) = 0$.

Joint PMF $\frac{a}{b}$

$(0,1)$

Joint Probability Mass Function

Definition 4.2 (PMF) $S_X \times S_Y$

The joint probability mass function *of discrete random variables* X *and* Y *is*

$$P_{X,Y}(x, y) = P[X = x, Y = y].$$

- This is the probability of the intersection of the events {X=x} and {Y=y}
- We are saying that an observation was made that maps to X=x and Y=y

Example of a Joint PMF

Example 4.1 Problem

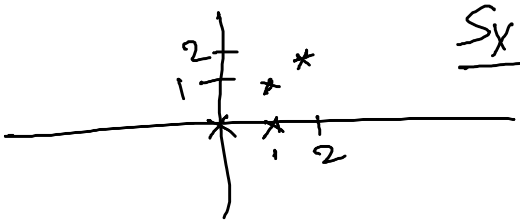
$\{a, r\}$

Test two integrated circuits one after the other. On each test, the possible outcomes are a (accept) and r (reject). Assume that all circuits are acceptable with probability 0.9 and that the outcomes of successive tests are independent. Count the number of acceptable circuits X and count the number of successful tests Y before you observe the first reject. (If both tests are successful, let $Y = 2$.) Draw a tree diagram for the experiment and find the joint PMF of X and Y.

$S_X = \{0, 1, 2\}$ $S_Y = \{0, 1, 2\}$

- What are the possible observations (sequence of observations)?
  - {aa, ar, ra, rr}

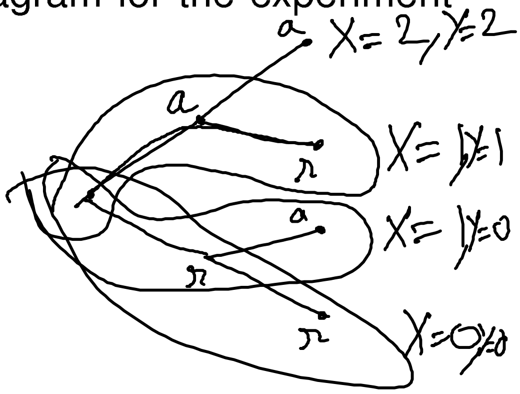
$\underline{S_X \times S_Y}$



Example of a Joint PMF

Example 4.1 Problem

Test two integrated circuits one after the other. On each test, the possible outcomes are a (accept) and r (reject). Assume that all circuits are acceptable with probability 0.9 and that the outcomes of successive tests are independent. Count the number of acceptable circuits X and count the number of successful tests Y before you observe the first reject. (If both tests are successful, let $Y = 2$.) Draw a tree diagram for the experiment and find the joint PMF of X and Y.

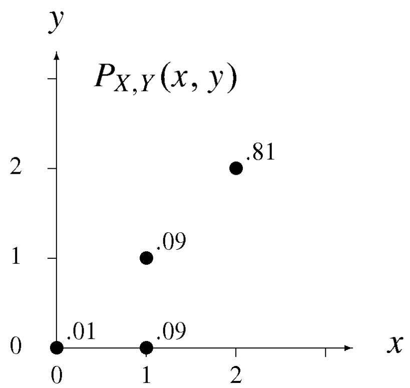


- aa => {X = 2, Y = 2} ; P[X = 2, Y = 2] = 0.81
- ar => {X = 1, Y = 1} ; P[X = 1, Y = 1] = 0.09
- ra => {X = 1, Y = 0} ; P[X = 1, Y = 0] = 0.09
- rr => {X = 0, Y = 0} ; P[X = 0, Y = 0] = 0.01

Example of a Joint PMF

Three possible representations of a Joint PMF

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.01	0	0
$x=1$	0.09	0.09	0
$x=2$	0	0	0.81



$$P_{X,Y}(x,y) = \begin{cases} 0.81 & x=2, y=2, \\ 0.09 & x=1, y=1, \\ 0.09 & x=1, y=0, \\ 0.01 & x=0, y=0. \\ 0 & \text{otherwise} \end{cases}$$

Joint PMF

$$\sum_{x \in S_X} \sum_{y \in S_Y} P_{X,Y}(x,y) = ?$$

- 1. Because you are summing over the probabilities of all mutually exclusive events {x,y} in S
- In other words, you are saying that at least one pair of {x,y} will be observed where x is in range of X and y in range of Y
- Note that the event {X=x1, Y=y1} is mutually exclusive of the event {X=x2, Y=y2} if x1 ≠ x2 or y1 ≠ y2

Probability of an Event B that is in the set S_X *x* S_Y

Theorem 4.2

For discrete random variables X and Y and any set B in the X, Y plane, the probability of the event $\{(X, Y) \in B\}$ is

$$P[B] = \sum_{(x,y)\in B} P_{X,Y}(x, y).$$

Quiz 4.2

The joint PMF $P_{Q,G}(q,g)$ for random variables Q and G is given in the following table:

$\{Q=0\} = (\{Q=0\} \cap \{G=0\}) \cup$

$P_{Q,G}(q,g)$	$g=0$	$g=1$	$g=2$	$g=3$
$q=0$	0.06	0.18	0.24	0.12
$q=1$	0.04	0.12	0.16	0.08

(4.12)

Calculate the following probabilities:

(1) $P[Q=0] \longrightarrow$

(2) $P[Q=G]$

(3) $P[G>1]$

(4) $P[G>Q]$

The event B in Theorem 4.2 can be defined to be any of the events in (1) to (4).

Marginal PMF

- For discrete RVs X and Y with joint PMF $P_{X,Y}(x,y)$, $P_X(x)$ and $P_Y(y)$ are defined as the marginal PMFs of X and Y respectively

$P[X=x, Y=y]$

$P[X=x] = \sum_{y \in S_Y} P[X=x | Y=y] P[Y=y]$ $\quad P[X=x] = \sum_{y \in S_Y} P[X=x, Y=y]$

Quiz 4.3

The probability mass function $P_{H,B}(h,b)$ for the two random variables H and B is given in the following table. Find the marginal PMFs $P_H(h)$ and $P_B(b)$.

$P_{H,B}(h,b)$	$b=0$	$b=2$	$b=4$
$h=-1$ ✓	0	0.4	0.2
$h=0$ ✓	0.1	0	0.1
$h=1$ ✓	0.1	0.1	0

(4.20)

$P[H=-1]$ ——

$P[H=0]$ ——

$P[H=1]$ ——

$P[\text{Any other } H] = 0$

Theorem 4.3

For discrete random variables X and Y with joint PMF $P_{X,Y}(x, y)$,

$$P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x, y), \qquad P_Y(y) = \sum_{x \in S_X} P_{X,Y}(x, y).$$

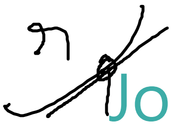
- We looked at the joint CDF and the joint PMF. What do you expect next?

$$\lim_{\Delta x, \Delta y \to 0} \frac{P[x \le X < x+\Delta x,\ y \le Y < y+\Delta y]}{\Delta x \Delta y} = \lim_{\Delta x, \Delta y \to 0} \frac{P[X \le x+\Delta x, Y \le y+\Delta y] - P[X \le x, Y \le y+\Delta y] - P[X \le x+\Delta x, Y \le y] + P[X \le x, Y \le y]}{\Delta x \Delta y}$$

$F_{X,Y}(x,y)$

$F_{X,Y}(x+\Delta x, y+\Delta y)$

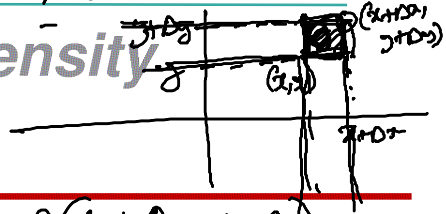
Joint PDF



$$\frac{\partial}{\partial y}\left(\frac{\partial}{\partial x} F_{X,Y}(x,y)\right) = \lim_{\Delta x \to 0} \frac{F_{X,Y}(x+\Delta x, y) - F_{X,Y}(x,y)}{\Delta x}$$

$F_{X,Y}(x,y)$

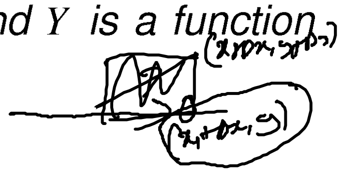
$f(x) \quad f(x_0) \quad x_0 + \Delta x$



Definition 4.3 Joint Probability Density Function (PDF)

$f(x_0 + \Delta x) \approx f(x_0) + \Delta x f'(x_0)$ $\quad f(x_0 + \Delta x, y_0 + \Delta y)$

The joint PDF of the continuous random variables X and Y is a function $f_{X,Y}(x, y)$ with the property



$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, dv\, du.$$

Theorem 4.4

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

Properties of a Joint PDF

Theorem 4.6

A joint PDF $f_{X,Y}(x, y)$ has the following properties corresponding to first and second axioms of probability (see Section 1.3):

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

Properties of a Joint PDF

Theorem 4.6

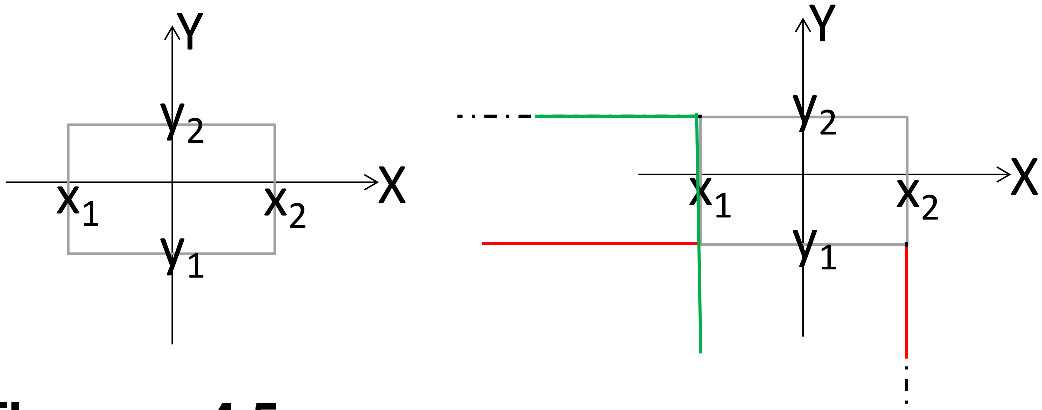
A joint PDF $f_{X,Y}(x, y)$ has the following properties corresponding to first and second axioms of probability (see Section 1.3):

(a) $f_{X,Y}(x, y) \geq 0$ for all (x, y),

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

- Nonnegative
- Area under it is unity

Theorem 4.5



Theorem 4.5

$$P\left[x_1 < X \leq x_2, y_1 < Y \leq y_2\right] = F_{X,Y}\left(x_2, y_2\right) - F_{X,Y}\left(x_2, y_1\right) - F_{X,Y}\left(x_1, y_2\right) + F_{X,Y}\left(x_1, y_1\right)$$

- Note that the theorem is valid since we assumed that X takes values along the x-axis and Y along the y-axis.
  - Outcome {X=x,Y=y} is a point in the *xy-plane*

Properties of a Joint PDF

- For a point (x,y) in the xy plane, we know that

$$\frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{F_{X,Y}(x+\Delta x, y+\Delta y) - F_{X,Y}(x, y+\Delta y) - F_{X,Y}(x+\Delta x, y) + F_{X,Y}(x,y)}{\Delta x \Delta y}$$

- Therefore we can write

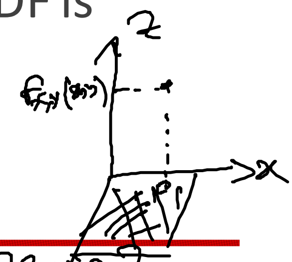
$$f_{X,Y}(x,y)\Delta x \Delta y \approx F_{X,Y}(x+\Delta x, y+\Delta y) - F_{X,Y}(x, y+\Delta y) - F_{X,Y}(x+\Delta x, y) + F_{X,Y}(x,y)$$

- Hence

$$f_{X,Y}(x,y)\Delta x \Delta y \approx P[x < X \leq x+\Delta x, y < Y \leq y+\Delta y]$$

Using a Joint PDF

- Even when the event corresponds to a rectangular area calculating probability of the event using the CDF is nontrivial!
- Using a PDF is usually more convenient

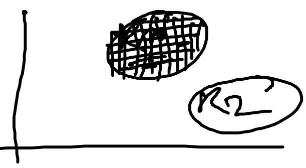


Theorem 4.7 $z = x^2 + y^2$

$A = \{R_1 \cup R_2\}$

The probability that the continuous random variables (X, Y) are in A is

$$P[A] = \iint_A f_{X,Y}(x, y)\, dx\, dy.$$



- A could be the rectangular region we looked at earlier
- A can be an area of any shape or size

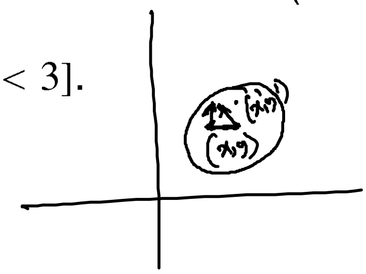
VVVS Problem

Example 4.4 Problem

Random variables X and Y have joint PDF

$$f_{X,Y}(x, y) = \begin{cases} c & 0 \leq x \leq 5, 0 \leq y \leq 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.22}$$

Find the constant c and $P[A] = P[2 \leq X < 3, 1 \leq Y < 3]$.



- c=1/15
- P[A] = 2/15
- Note that A is a rectangular area and so the problem is very straightforward

Calculating a CDF from a PDF!

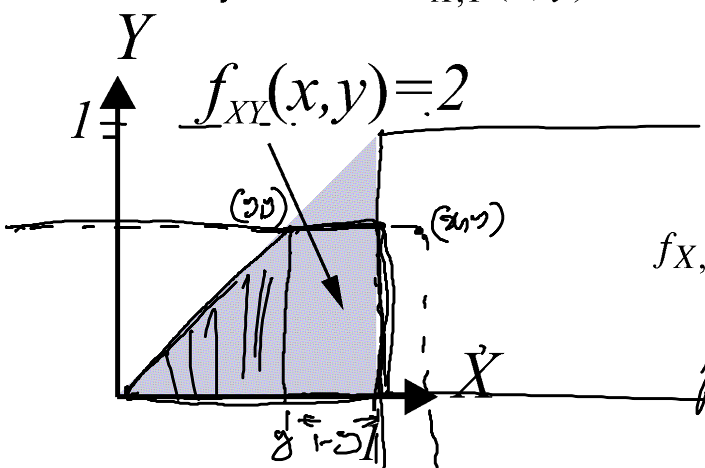
Example 4.5 Problem

$P[X \le x, Y \le y] = 1$ $\quad x \ge 1, y \ge 1$

$x \ge 1, 0 \le y < 1$

Find the joint CDF $F_{X,Y}(x, y)$ when X and Y have joint PDF

$P[X \le x, Y \le y]$ $\quad \int_0^1 \int_0^y dy\, dx$



$$f_{X,Y}(x, y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$\iint_A f_{X,Y}(x,y)\, dx\, dy$

$2 \iint_A dx\, dy$

$2(y)(1-y)$

$+ 2\left(\frac{1}{2}y^2\right)$

$= 2y - 2y^2 + y^2$

- All valid outcomes lie in the triangle

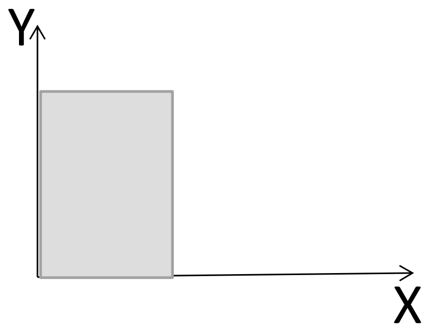
Quiz 4.4

The joint probability density function of random variables X and Y is

$$f_{X,Y}(x, y) = \begin{cases} cxy & 0 \le x \le 1, 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.33}$$

Find the constant c. What is the probability of the event $A = X^2 + Y^2 \le 1$?

- We want to find the probability that a pair (x,y) lies inside the unit circle
- Use polar coordinates
  - dx dy = r dr dϕ



Marginal PDF

Theorem 4.8

If X and Y are random variables with joint PDF $f_{X,Y}(x, y)$,

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

- Note that P[X <= x] = P[X<=x, Y<∞]
- We have

$$P[X \leq x] = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\, dv\, du$$

- Taking the derivative with respect to x (and using the fundamental theorem of calculus) we get our desired result

Quiz 4.5

The joint probability density function of random variables X and Y is

$$f_{X,Y}(x, y) = \begin{cases} 6(x + y^2)/5 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.40}$$

Find $f_X(x)$ and $f_Y(y)$, the marginal PDFs of X and Y.

- You must get

$$f_X(x) = \begin{cases} \frac{6x+2}{5} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

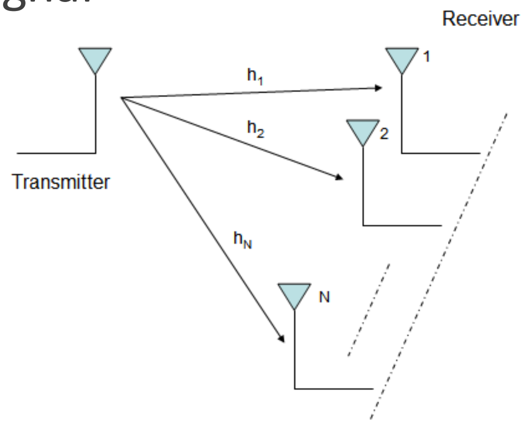
$$f_Y(y) = \begin{cases} \frac{3+6y^2}{5} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Functions of Two Random Variables

- Assume life was easier and you had just a midterm (30% weight) and a final exam.
- Let X be the RV that denotes your midterm marks
  - $S_X = [0,30]$
- Let Y be the RV that denotes your final exam marks
  - $S_Y = [0,70]$
- Let G be the RV that denotes your grade
- G = g(X,Y)
  - Where g is determined by yours truly
  - Example $G = 0.3\,X + 0.7\,Y$

A More Useful Example

- More than one antenna at your receiver receives a transmitted signal



Pic courtesy http://www.dsplog.com/2008/09/06/receiver-diversity-selection-diversity/

- For N=2, we have received signals (RVs) R_1 and R_2
- The resultant received signal $R = w_1R_1 + w_2R_2$
- We could keep it simpler and only select the stronger signal
  - $R = \max(R_1,R_2)$

Functions of Two Random Variables

- We want to find the probability model that corresponds to the resulting RV
- **We know the joint probability model**
- W = X + Y
  - PMF, CDF, PDF of W?
- As always if X and Y are discrete, life is easy

Theorem 4.9

For discrete random variables X and Y, the derived random variable $W = g(X, Y)$ has PMF

$$P_W(w) = \sum_{(x,y):g(x,y)=w} P_{X,Y}(x, y).$$

Continuous Function of Continuous RVs

- W = g(X,Y)
  - g is continuous and so are X and Y.
- W is thus a continuous RV

Theorem 4.10

For continuous random variables X and Y, the CDF of $W = g(X, Y)$ is

$$F_W(w) = P[W \le w] = \iint_{g(x,y)\le w} f_{X,Y}(x, y)\, dx\, dy.$$

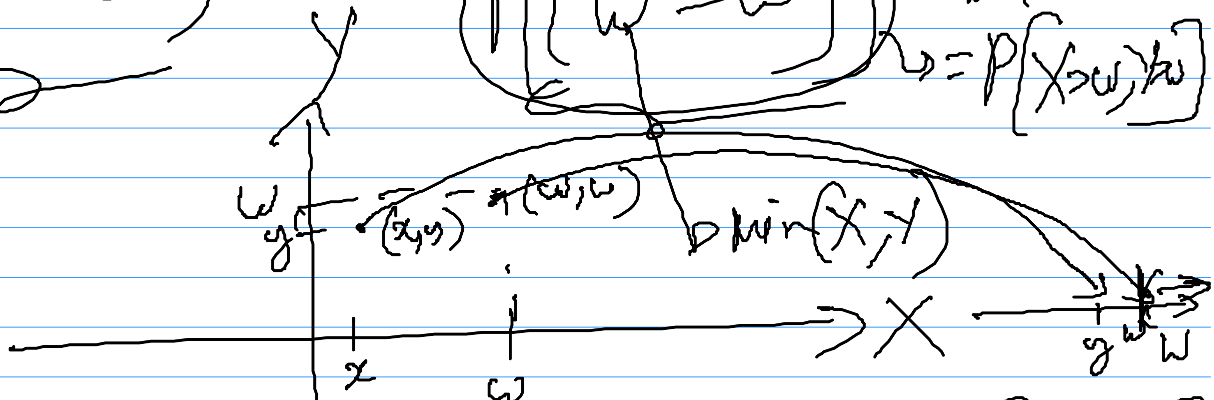
X, Y

$\{X \le w, Y \le w\}$ $\quad W = \max(X, Y)$

$\{W \le w\}$ $\quad P[\{W > w\}]$ $\quad \bar{F}_W(w)$

$= P[X > w, Y > w]$

$\to \min(X, Y)$



$F_W(w) = F_{X,Y}(w, w)$

$P[W \le w] = 1 - P[X > w, Y > w]$

Theorem 4.11

For continuous random variables X and Y, the CDF of $W = \max(X, Y)$ is

$$F_W(w) = F_{X,Y}(w, w) = \int_{-\infty}^{w} \int_{-\infty}^{w} f_{X,Y}(x, y)\, dx\, dy.$$

What if W = min(X,Y)?

$$P[W > w] = P[X > w, Y > w]$$

Expected Values!

$E[W] = \sum_{w \in S_W} w \, P[W=w]$

Theorem 4.12

For random variables X and Y, the expected value of $W = g(X, Y)$ is

Discrete: $$E[W] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x, y) P_{X,Y}(x, y)$$

$\sum_{w \in S_W} w \sum \sum_{g(x,y)=w} P[X=x, Y=y]$

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

Theorem 4.13

$$E[g_1(X, Y) + \cdots + g_n(X, Y)] = E[g_1(X, Y)] + \cdots + E[g_n(X, Y)].$$

- A very useful theorem!

Suppose $g(X,Y) = X = W$

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

$$= \int_{-\infty}^{\infty} x \underbrace{\left[\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\right]}_{f_X(x)} dx$$

$$\boxed{E[W] = \sum_{w \in S_W} w\, P[W=w]}$$

$$= \sum_{w \in S_W} w \boxed{\sum_{\{g(x,y)=w\}} P[X=x, Y=y]}$$

$$E[W] = \int_{-\infty}^{\infty} w \underbrace{f_W(w)}\, dw \qquad \{(x,y) : g(x,y) = w\}$$

$$\iint_{g(x,y)=w} f_{X,Y}(x,y)\, dx\, dy$$

Expected Values

- E[X + Y] = ?
- E[(X + Y)²] = ?

Theorem 4.15

The variance of the sum of two random variables is

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] + 2E[(X-\mu_X)(Y-\mu_Y)].$$

Definition 4.4 *Covariance*

The covariance *of two random variables X and Y is*

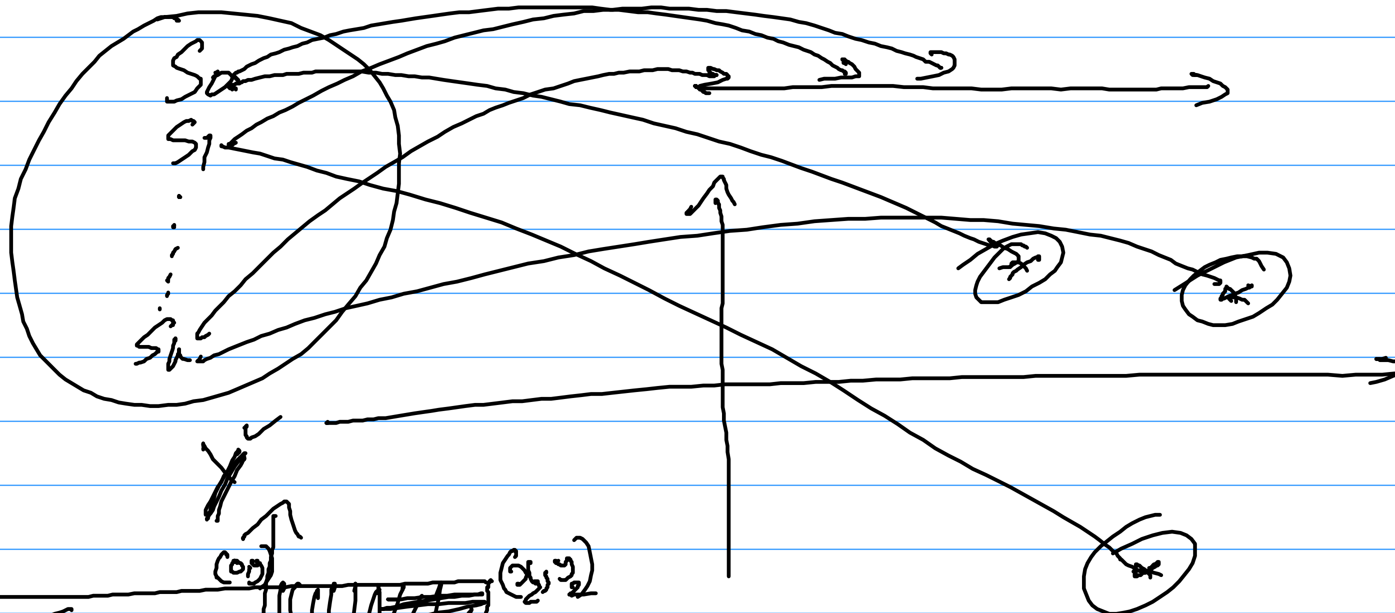
$$\text{Cov}[X,Y] = E[(X-\mu_X)(Y-\mu_Y)].$$

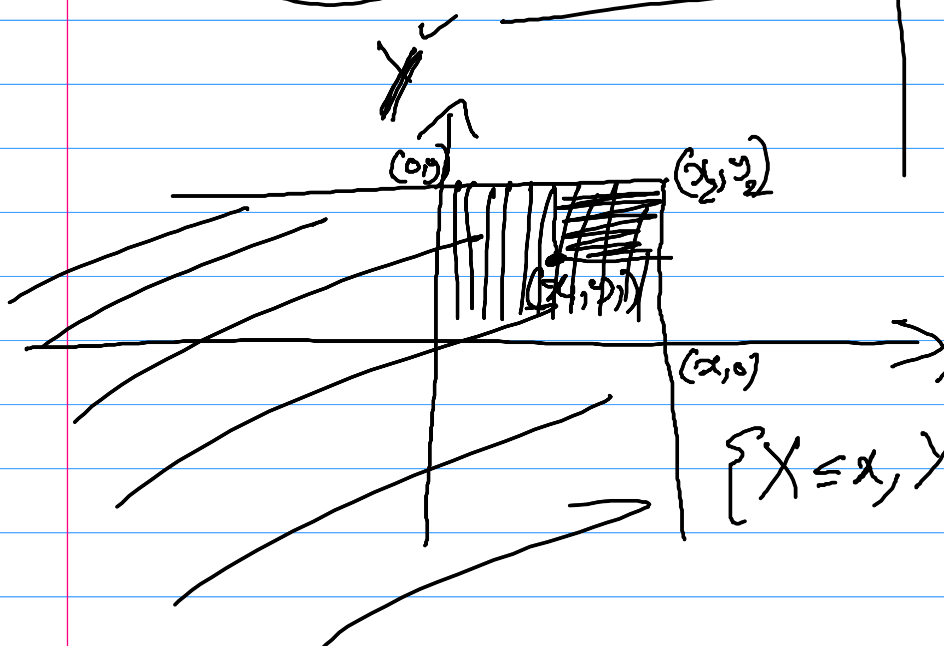
Pairs of Random Variables

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

$s \in S$

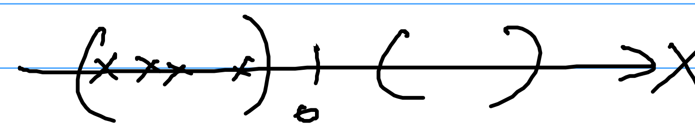
$X(s)$





$\{X \leq x, Y \leq y\}$

$(-\infty, \infty) \times (-\infty, \infty)$



We started with…

- An Experiment that contains a
  - Procedure, Observations, and Model
- Later we mapped outcomes to numbers
  - One or more outcomes to a point on the line
- We called the numbers random variables
- We will now map an outcome to a pair of numbers
  - One or more outcomes to a point on a plane
- The numbers in the pair correspond to RVs X and Y
- For example, a transmitted sinusoid that is received with a random amplitude and random phase
  - Outcomes can be described by X = amplitude and Y = phase

Pair of RVs

- We defined a CDF for a single RV X
- For the pair of RVs we define a ***joint*** CDF

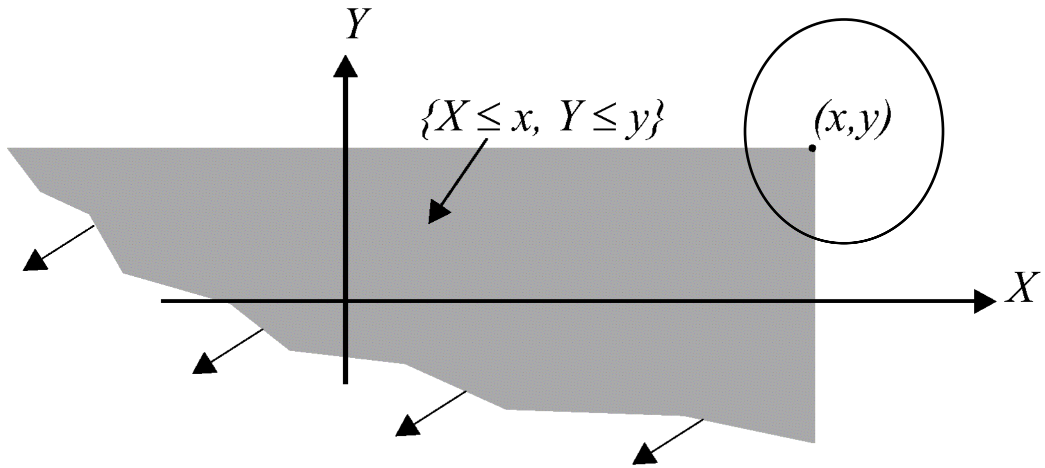
Definition 4.1 Joint Cumulative Distribution Function (CDF)

The joint cumulative distribution function *of random variables X and Y is*

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

Joint CDF

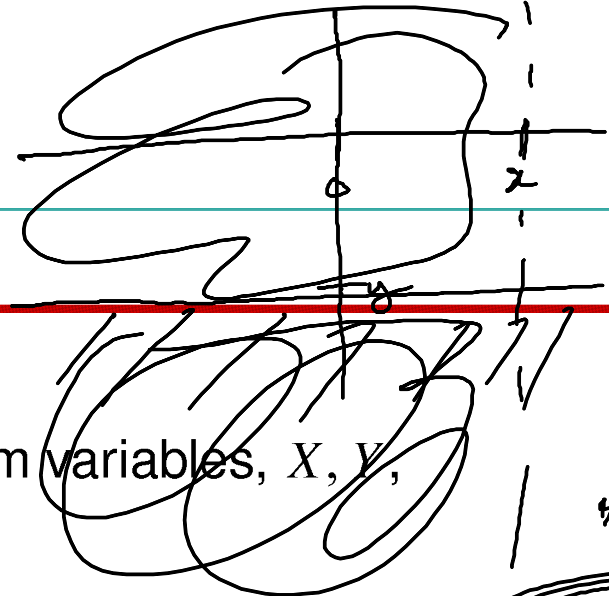
Figure 4.1



The area of the (X, Y) plane corresponding to the joint cumulative distribution function $F_{X,Y}(x, y)$.

- We are interested in the probability of the intersection of the events {X<=x} and {Y<=y}

Joint CDF



S_Y

Theorem 4.1

For any pair of random variables, X, Y,

$\{X \le x\}$

$\{Y \le y,\ Y > y\}$

$\{X \le x \cap Y \le y\} \cup \{X \le x,\ Y > y\}$

(a) $\boxed{0} \le F_{X,Y}(x, y) \le \boxed{1}$

(b) $F_X(x) = \boxed{F_{X,Y}(x,\infty)}$,

$\{X \le x,\ Y \le \infty\} = \{X \le x\}$

$P[X \le x,\ Y < \infty]\ P[Y < \infty] = P[X \le x]$

(c) $F_Y(y) = \boxed{}$,

$\{X \le -\infty,\ Y \le y\}$

(d) $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = \boxed{}$

$P[X \le x, Y \le y]$ $P[X \le x_1, Y \le y_1]$

$A \subseteq B$

$P[A] \le P[B]$

(e) If $x \le x_1$ and $y \le y_1$, then $F_{X,Y}(x, y) \boxed{\le} F_{X,Y}(x_1, y_1)$,

(f) $F_{X,Y}(\infty, \infty) = \boxed{1}$.

(g) $F_{X,Y}(\infty, -\infty) = ?$

Joint CDF

Theorem 4.1

For any pair of random variables, X, Y,

(a) $0 \le F_{X,Y}(x, y) \le 1$,

(b) $F_X(x) = F_{X,Y}(x, \infty)$,

(c) $F_Y(y) = F_{X,Y}(\infty, y)$,

$P[X \le -\infty, Y \le y]$

(d) $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$,

(e) If $x \le x_1$ and $y \le y_1$, then $F_{X,Y}(x, y) \le F_{X,Y}(x_1, y_1)$,

(f) $F_{X,Y}(\infty, \infty) = 1$.

(g) $F_{X,Y}(\infty, -\infty) = 0$.

Joint PMF

$\frac{a}{b}$

$(0,1)$

Joint Probability Mass Function

$S_X \times S_Y$

Definition 4.2 **(PMF)**

The joint probability mass function *of discrete random variables* X *and* Y *is*

$$P_{X,Y}(x, y) = P[X = x, Y = y].$$

- This is the probability of the intersection of the events {X=x} and {Y=y}
- We are saying that an observation was made that maps to X=x and Y=y

Example of a Joint PMF

Example 4.1 Problem

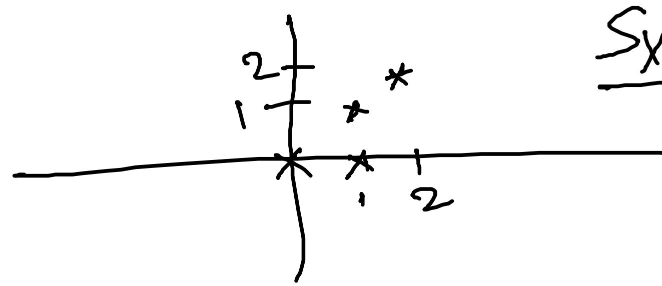
$\{a, r\}$

Test two integrated circuits one after the other. On each test, the possible outcomes are a (accept) and r (reject). Assume that all circuits are acceptable with probability 0.9 and that the outcomes of successive tests are independent. Count the number of acceptable circuits X and count the number of successful tests Y before you observe the first reject. (If both tests are successful, let $Y = 2$.) Draw a tree diagram for the experiment and find the joint PMF of X and Y.

$S_X = \{0, 1, 2\}$ $S_Y = \{0, 1, 2\}$

- What are the possible observations (sequence of observations)?
  - {aa, ar, ra, rr}

$\underline{S_X \times S_Y}$

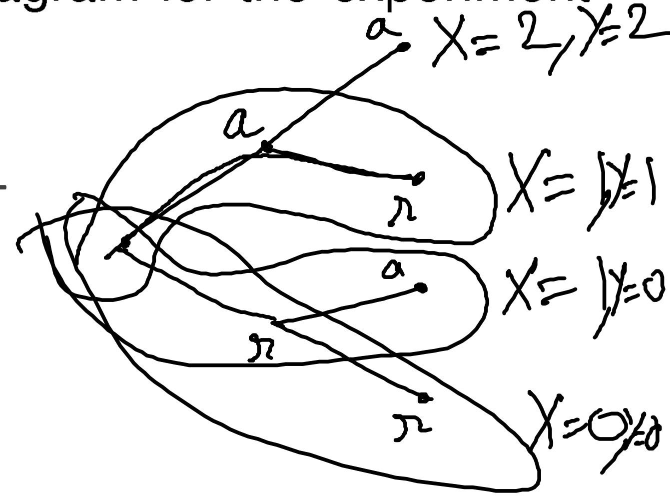


Example of a Joint PMF

Example 4.1 Problem

Test two integrated circuits one after the other. On each test, the possible outcomes are a (accept) and r (reject). Assume that all circuits are acceptable with probability 0.9 and that the outcomes of successive tests are independent. Count the number of acceptable circuits X and count the number of successful tests Y before you observe the first reject. (If both tests are successful, let $Y = 2$.) Draw a tree diagram for the experiment and find the joint PMF of X and Y.

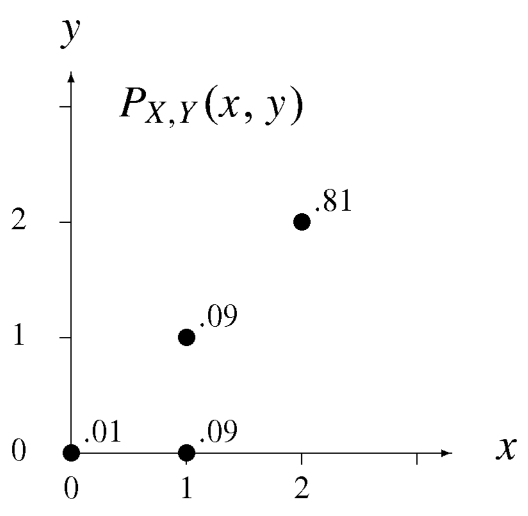
- aa => {X = 2, Y = 2} ; P[X = 2, Y = 2] = 0.81
- ar => {X = 1, Y = 1} ; P[X = 1, Y = 1] = 0.09
- ra => {X = 1, Y = 0} ; P[X = 1, Y = 0] = 0.09
- rr => {X = 0, Y = 0} ; P[X = 0, Y = 0] = 0.01



Example of a Joint PMF

Three possible representations of a Joint PMF

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.01	0	0
$x=1$	0.09	0.09	0
$x=2$	0	0	0.81



$$P_{X,Y}(x,y)=\begin{cases} 0.81 & x=2, y=2, \\ 0.09 & x=1, y=1, \\ 0.09 & x=1, y=0, \\ 0.01 & x=0, y=0. \\ 0 & \text{otherwise} \end{cases}$$

Joint PMF

$$\sum_{x \in S_X} \sum_{y \in S_Y} P_{X,Y}(x, y) = ?$$

- 1. Because you are summing over the probabilities of all mutually exclusive events {x,y} in S
- In other words, you are saying that at least one pair of {x,y} will be observed where x is in range of X and y in range of Y
- Note that the event {X=x1, Y=y1} is mutually exclusive of the event {X=x2, Y=y2} if x1 ≠ x2 or y1 ≠ y2

Probability of an Event B that is in the set $S_X \times S_Y$

Theorem 4.2

For discrete random variables X and Y and any set B in the X, Y plane, the probability of the event $\{(X, Y) \in B\}$ is

$$P[B] = \sum_{(x,y)\in B} P_{X,Y}(x, y).$$

VVVS Problem

Quiz 4.2

The joint PMF $P_{Q,G}(q,g)$ for random variables Q and G is given in the following table:

$\{Q=0\} = (\{Q=0\} \cap \{G=0\}) \cup$

$P_{Q,G}(q,g)$	$g=0$	$g=1$	$g=2$	$g=3$
$q=0$	0.06	0.18	0.24	0.12
$q=1$	0.04	0.12	0.16	0.08

(4.12)

Calculate the following probabilities:

(1) $P[Q=0]$ $\longrightarrow$
(2) $P[Q=G]$
(3) $P[G>1]$
(4) $P[G>Q]$

The event B in Theorem 4.2 can be defined to be any of the events in (1) to (4).

Marginal PMF

- For discrete RVs X and Y with joint PMF $P_{X,Y}(x,y)$, $P_X(x)$ and $P_Y(y)$ are defined as the marginal PMFs of X and Y respectively

$P[X=x, Y=y]$

$P[X=x] = \sum_{y \in S_Y} P[X=x \mid Y=y] \, P[Y=y]$

$P[X=x] = \sum_{y \in S_Y} P[X=x, Y=y]$

Quiz 4.3

The probability mass function $P_{H,B}(h,b)$ for the two random variables H and B is given in the following table. Find the marginal PMFs $P_H(h)$ and $P_B(b)$.

$P_{H,B}(h,b)$	$b=0$	$b=2$	$b=4$
$h=-1$	0	0.4	0.2
$h=0$	0.1	0	0.1
$h=1$	0.1	0.1	0

(4.20)

$P[H=-1]$ ___

$P[H=0]$ ___

$P[H=1]$ ___

$P[\text{Any other } H] = 0$

Marginal PMF

Theorem 4.3

For discrete random variables X and Y with joint PMF $P_{X,Y}(x, y)$,

$$P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x, y), \qquad P_Y(y) = \sum_{x \in S_X} P_{X,Y}(x, y).$$

- We looked at the joint CDF and the joint PMF. What do you expect next?

$$\lim_{\Delta x, \Delta y \to 0} \frac{P[x \le X < x+\Delta x,\ y \le Y < y+\Delta y]}{\Delta x \,\Delta y} = \lim_{\Delta x, \Delta y \to 0} \frac{P[X \le x+\Delta x,\ Y \le y+\Delta y] - P[X \le x,\ Y \le y+\Delta y] - P[X \le x+\Delta x,\ Y \le y] + P[X \le x,\ Y \le y]}{\Delta x \,\Delta y}$$

$F_{X,Y}(x,y)$

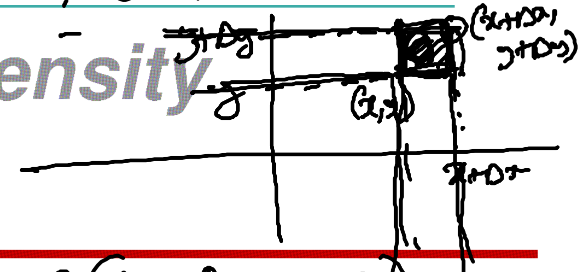
$F_{X,Y}(x+\Delta x, y+\Delta y)$

Joint PDF

$$\frac{\partial}{\partial y}\left(\frac{\partial}{\partial x} F_{X,Y}(x,y)\right) = \lim_{\Delta x \to 0} \frac{F_{X,Y}(x+\Delta x, y) - F_{X,Y}(x,y)}{\Delta x}$$

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

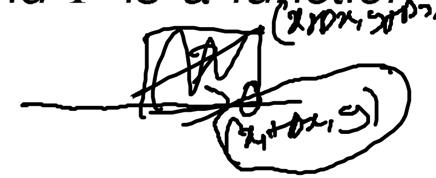
$f(x)$ $\quad$ $f(x_0)$ $\quad$ $x_0 + \Delta x$



Definition 4.3 Joint Probability Density Function (PDF)

$$f(x_0 + \Delta x) \cong f(x_0) + \Delta x f'(x_0) \qquad f(x_0 + \Delta x, y_0 + \Delta y)$$

The joint PDF of the continuous random variables X and Y is a function $f_{X,Y}(x,y)$ with the property



$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)\, dv\, du.$$

Theorem 4.4

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

Properties of a Joint PDF

Theorem 4.6

A joint PDF $f_{X,Y}(x, y)$ has the following properties corresponding to first and second axioms of probability (see Section 1.3):

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

Properties of a Joint PDF

Theorem 4.6

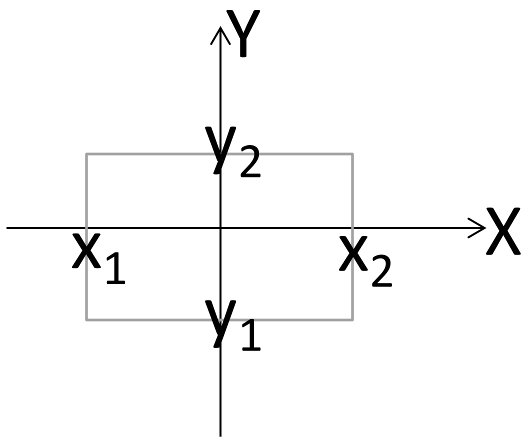
A joint PDF $f_{X,Y}(x, y)$ has the following properties corresponding to first and second axioms of probability (see Section 1.3):

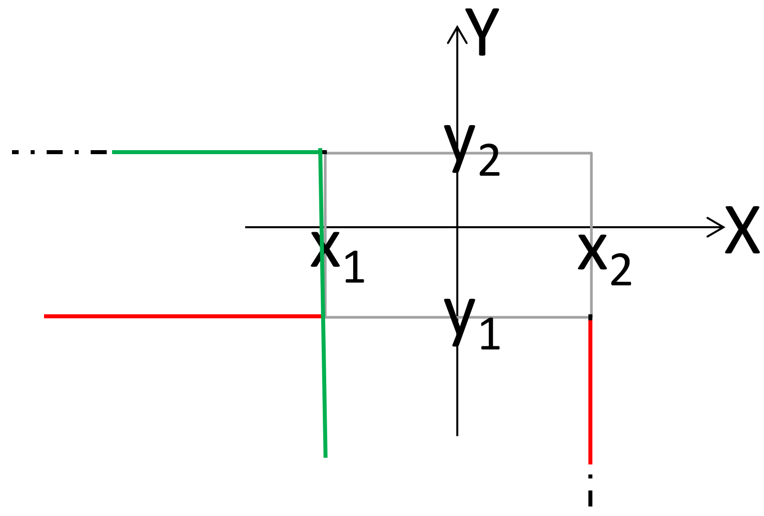
(a) $f_{X,Y}(x, y) \geq 0$ for all (x, y),

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

- Nonnegative
- Area under it is unity

Theorem 4.5





Theorem 4.5

$$P[x_1 < X \le x_2, y_1 < Y \le y_2] = F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1)$$

- Note that the theorem is valid since we assumed that X takes values along the x-axis and Y along the y-axis.
  - Outcome {X=x,Y=y} is a point in the *xy-plane*

Properties of a Joint PDF

- For a point (x,y) in the xy plane, we know that

$$\frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = \lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \frac{F_{X,Y}(x+\Delta x, y+\Delta y) - F_{X,Y}(x, y+\Delta y) - F_{X,Y}(x+\Delta x, y) + F_{X,Y}(x,y)}{\Delta x \Delta y}$$

- Therefore we can write

$$f_{X,Y}(x,y)\Delta x \Delta y \approx F_{X,Y}(x+\Delta x, y+\Delta y) - F_{X,Y}(x, y+\Delta y) - F_{X,Y}(x+\Delta x, y) + F_{X,Y}(x,y)$$

- Hence

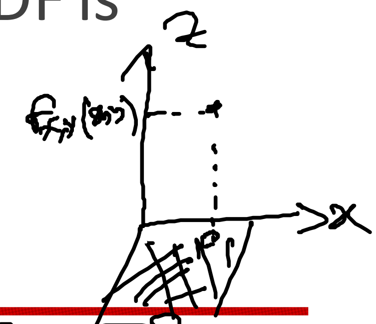
$$f_{X,Y}(x,y)\Delta x \Delta y \approx P[x < X \leq x + \Delta x, y < Y \leq y + \Delta y]$$

Using a Joint PDF

- Even when the event corresponds to a rectangular area calculating probability of the event using the CDF is nontrivial!
- Using a PDF is usually more convenient

Theorem 4.7

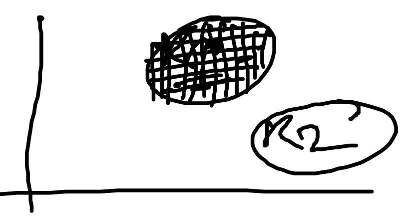
$z = x^2 + y^2$



$A = \{R_1 \cup R_2\}$
$\emptyset$

The probability that the continuous random variables (X, Y) are in A is

$$P[A] = \iint_A f_{X,Y}(x, y)\, dx\, dy.$$



- *A* could be the rectangular region we looked at earlier
- *A* can be an area of any shape or size

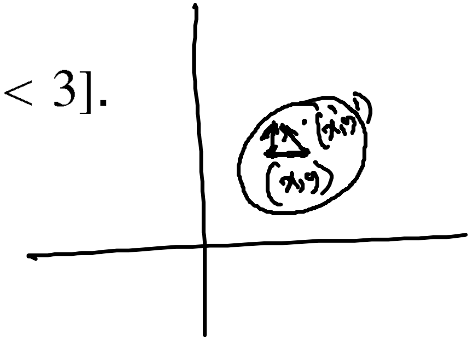
VVVS Problem

Example 4.4 Problem

Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} c & 0 \le x \le 5, 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.22}$$

Find the constant c and $P[A] = P[2 \le X < 3, 1 \le Y < 3]$.



- c=1/15
- P[A] = 2/15
- Note that A is a rectangular area and so the problem is very straightforward

Calculating a CDF from a PDF!

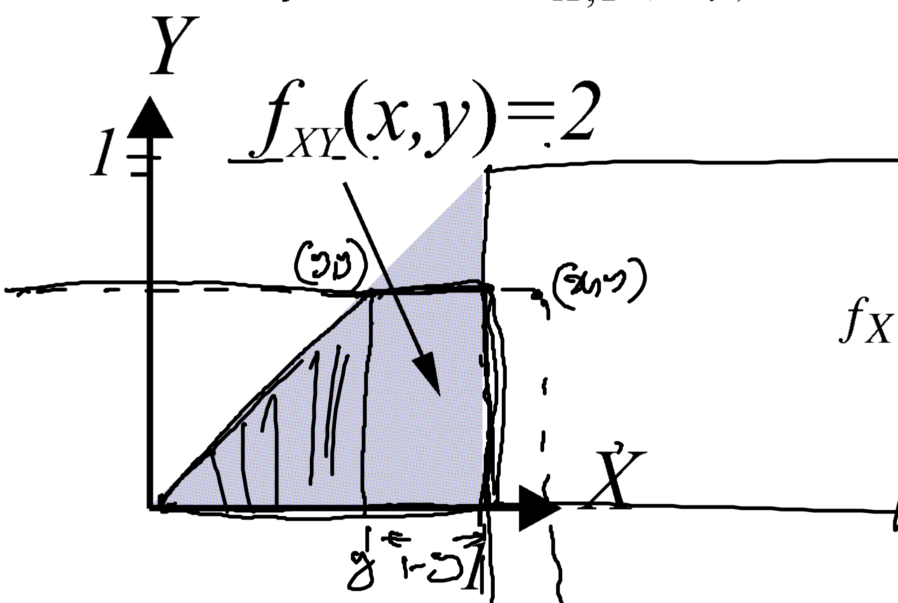
Example 4.5 Problem

$P[X \le x, Y \le y] = 1 \quad x > 1, y > 1$

$x > 1, 0 \le y < 1$

Find the joint CDF $F_{X,Y}(x, y)$ when X and Y have joint PDF

$P[X \le x, Y \le y] \quad \int_0^1 \int_0^y dy\,dx$



$$f_{X,Y}(x, y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

$\iint_A f_{X,Y}(x,y)\,dx\,dy$

$2 \iint_A dx\,dy$

$2(y)(1-y) + 2\left(\frac{1}{2}y^2\right) = 2y - 2y^2 + y^2$

- All valid outcomes lie in the triangle

VS Problem

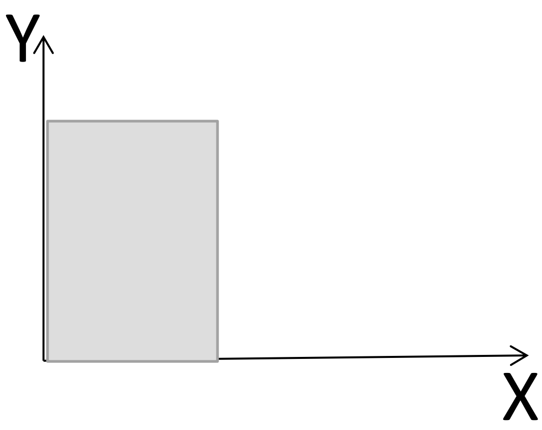
Quiz 4.4

The joint probability density function of random variables X and Y is

$$f_{X,Y}(x, y) = \begin{cases} cxy & 0 \le x \le 1, 0 \le y \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.33}$$

Find the constant c. What is the probability of the event $A = X^2 + Y^2 \le 1$?

- We want to find the probability that a pair (x,y) lies inside the unit circle
- Use polar coordinates
  - dx dy = r dr dϕ



Marginal PDF

Theorem 4.8

If X and Y are random variables with joint PDF $f_{X,Y}(x, y)$,

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

- Note that P[X <= x] = P[X<=x, Y<∞]
- We have

$$P[X \leq x] = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\, dv\, du$$

- Taking the derivative with respect to x (and using the fundamental theorem of calculus) we get our desired result

VVVS Problem

Quiz 4.5

The joint probability density function of random variables X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 6(x+y^2)/5 & 0 \le x \le 1, 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.40}$$

Find $f_X(x)$ and $f_Y(y)$, the marginal PDFs of X and Y.

- You must get

$$f_X(x) = \begin{cases} \frac{6x+2}{5} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

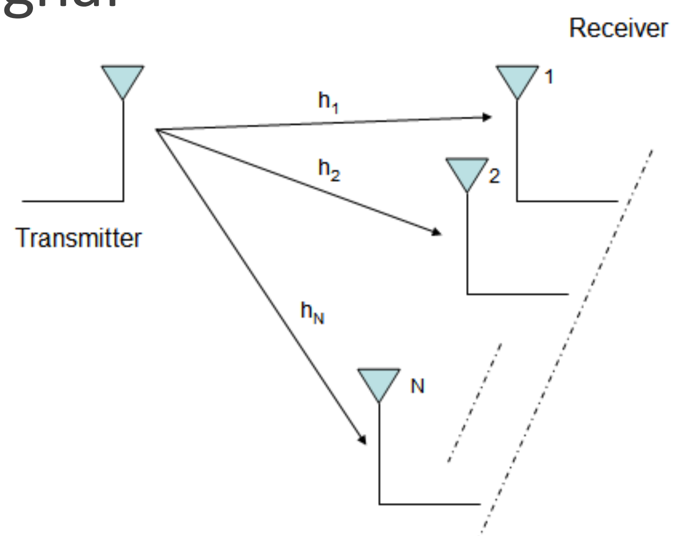
$$f_Y(y) = \begin{cases} \frac{3+6y^2}{5} & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Functions of Two Random Variables

- Assume life was easier and you had just a midterm (30% weight) and a final exam.
- Let X be the RV that denotes your midterm marks
  - $S_X = [0,30]$
- Let Y be the RV that denotes your final exam marks
  - $S_Y = [0,70]$
- Let G be the RV that denotes your grade
- $G = g(X,Y)$
  - Where g is determined by yours truly
  - Example $G = 0.3\ X + 0.7\ Y$

A More Useful Example

- More than one antenna at your receiver receives a transmitted signal



Pic courtesy
http://www.dsplog.com/2008/09/06/receiver-diversity-selection-diversity/

- For N=2, we have received signals (RVs) R_1 and R_2
- The resultant received signal $R = w_1R_1 + w_2R_2$
- We could keep it simpler and only select the stronger signal
  - $R = \max(R_1, R_2)$

Functions of Two Random Variables

- We want to find the probability model that corresponds to the resulting RV
- **We know the joint probability model**
- W = X + Y
  - PMF, CDF, PDF of W?
- As always if X and Y are discrete, life is easy

Theorem 4.9

For discrete random variables X and Y, the derived random variable $W = g(X, Y)$ has PMF

$$P_W(w) = \sum_{(x,y):g(x,y)=w} P_{X,Y}(x, y).$$

Continuous Function of Continuous RVs

- W = g(X,Y)
  - g is continuous and so are X and Y.
- W is thus a continuous RV

Theorem 4.10

For continuous random variables X and Y, the CDF of $W = g(X, Y)$ is

$$F_W(w) = P[W \leq w] = \iint_{g(x,y) \leq w} f_{X,Y}(x, y)\, dx\, dy.$$

X, Y

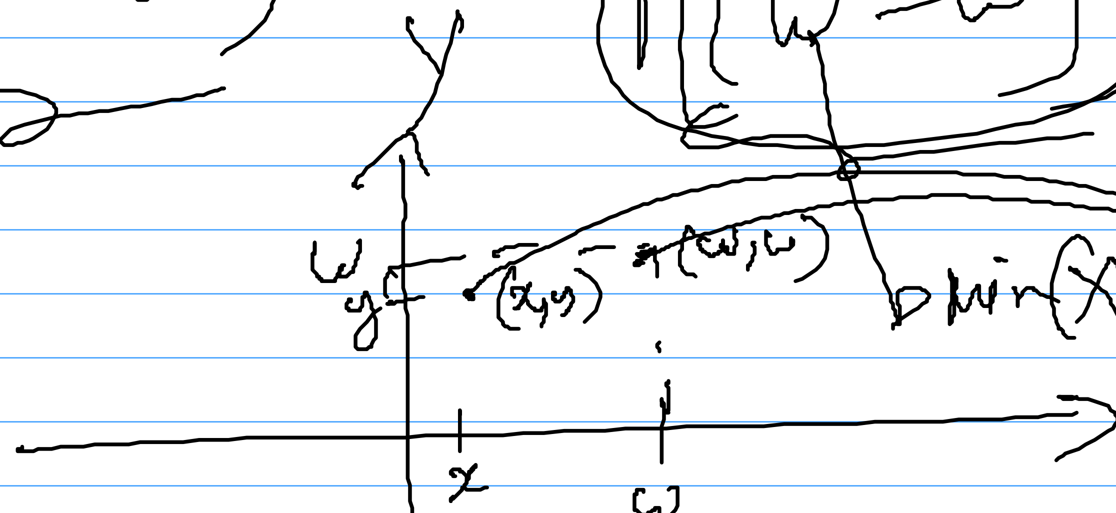
$$W = \max(X, Y)$$

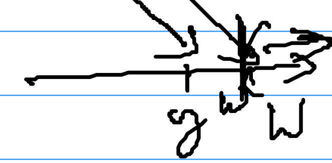
$$\{X \le w, Y \le w\}$$

$$\{W \le w\}$$

$$P[W > w] = \bar{F}_W(w) = P[X > w, Y > w]$$

$\min(X, Y)$





$$F_W(w) = F_{X,Y}(w, w)$$

$$P[W \le w] = 1 - P[X > w, Y > w]$$

Maximum of Two RVs

Theorem 4.11

For continuous random variables X and Y, the CDF of $W = \max(X, Y)$ is

$$F_W(w) = F_{X,Y}(w, w) = \int_{-\infty}^{w} \int_{-\infty}^{w} f_{X,Y}(x, y)\, dx\, dy.$$

What if W = min(X,Y)?

$$P[W > w] = P(X > w, Y > w)$$

Expected Values!

$$E[W] = \sum_{w \in S_W} w \, P[W=w]$$

Theorem 4.12

For random variables X and Y, the expected value of $W = g(X, Y)$ is

Discrete: $E[W] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x, y) P_{X,Y}(x, y)$

$$\sum_{w \in S_W} w \sum_{g(x,y)=w} \sum P[X=x, Y=y]$$

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

Theorem 4.13

$$E[g_1(X, Y) + \cdots + g_n(X, Y)] = E[g_1(X, Y)] + \cdots + E[g_n(X, Y)].$$

- A very useful theorem!

Suppose $g(X,Y) = X = W$

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

$$= \int_{-\infty}^{\infty} x \underbrace{\left[\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy\right]}_{f_X(x)} dx$$

$$\boxed{E[W] = \sum_{w \in S_W} w \, P[W = w]}$$

$$= \sum_{w \in S_W} w \boxed{\sum_{\{g(x,y)=w\}} P[X = x, Y = y]}$$

$$E[W] = \int_{-\infty}^{\infty} w \, f_W(w) \, dw \qquad \{(x,y) : g(x,y) = w\}$$

$$\iint_{g(x,y)=w} f_{X,Y}(x,y) \, dx \, dy$$

Expected Values

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

Theorem 4.15

The variance of the sum of two random variables is

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2E[(X-\mu_X)(Y-\mu_Y)].$$

Definition 4.4 Covariance

The covariance *of two random variables X and Y is*

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

$$\mathrm{Cov}[X,Y] = E[(X-\mu_X)(Y-\mu_Y)].$$

Definition 4.5 Correlation

The correlation *of* X *and* Y *is* $r_{X,Y} = E[XY]$

Theorem 4.16

(a) $\text{Cov}[X, Y] = r_{X,Y} - \mu_X \mu_Y$.

(b) $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\,\text{Cov}[X, Y]$.

(c) If $X = Y$, $\text{Cov}[X, Y] = \text{Var}[X] = \text{Var}[Y]$ and $r_{X,Y} = E[X^2] = E[Y^2]$.

Example 4.12 Problem

For the integrated circuits tests in Example 4.1, we found in Example 4.3 that the probability model for X and Y is given by the following matrix.

$P_{X,Y}(x, y)$	$y = 0$	$y = 1$	$y = 2$	$P_X(x)$
$x = 0$	0.01	0	0	
$x = 1$	0.09	0.09	0	
$x = 2$	0	0	0.81	
$P_Y(y)$				

(4.73)

Find $r_{X,Y}$ and $\mathrm{Cov}[X, Y]$.

VS Problem

Example 4.12 Problem

For the integrated circuits tests in Example 4.1, we found in Example 4.3 that the probability model for X and Y is given by the following matrix.

$P_{X,Y}(x, y)$	$y = 0$	$y = 1$	$y = 2$	$P_X(x)$
$x = 0$	0.01	0	0	0.01
$x = 1$	0.09	0.09	0	0.18
$x = 2$	0	0	0.81	0.81
$P_Y(y)$	0.10	0.09	0.81	

(4.73)

Find $r_{X,Y}$ and $\mathrm{Cov}[X, Y]$.

- 3.33 and 0.252

More Definitions!

Definition 4.6 Orthogonal Random Variables

Random variables X *and* Y *are* orthogonal *if* $r_{X,Y} = 0$.

$(x_1, y_1) \quad (x_2, y_2) \quad (x_3, y_3) \quad \ldots \quad (x_n, y_n)$

Definition 4.7 Uncorrelated Random Variables

$x^T y = (x_1 \; x_2 \ldots x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ $\quad \frac{x_1 y_1 + x_2 y_2 + x_3 y_3 + \cdots + x_n y_n}{n}$

Random variables X *and* Y *are* uncorrelated *if* $\mathrm{Cov}[X, Y] = 0$. $E(XY) = E[X]E[Y]$

Definition 4.8 Correlation Coefficient

The correlation coefficient *of two random variables* X *and* Y *is*

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{\mathrm{Cov}[X,Y]}{\sigma_X \sigma_Y}.$$

Correlation Coefficient

Theorem 4.17

$$-1 \leq \rho_{X,Y} \leq 1.$$

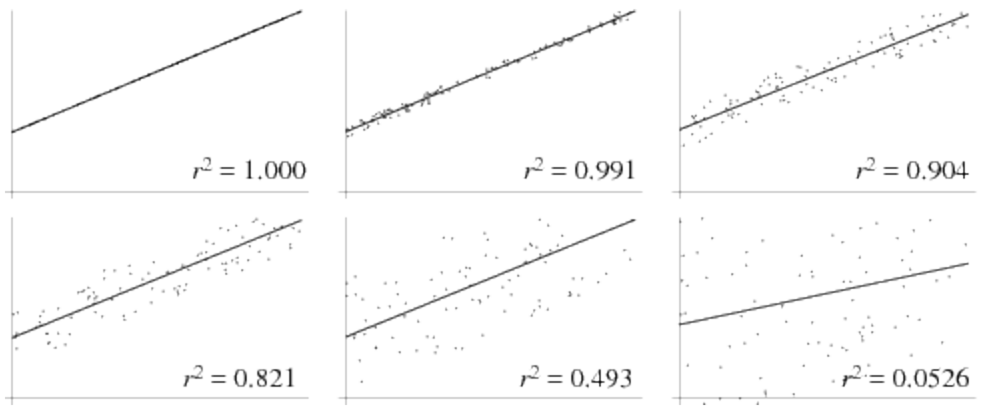
- Let W = X – aY, for any constant a
- We have Var[W] = Var[X] – 2a Cov[X,Y] + a^2 Var[Y]
- Use the fact that Var[W] >= 0
- Let a = σ_x/σ_y to get the upper bound
- Let a = - σ_x/σ_y to get the lower bound

Correlation Coefficient

- A positive correlation coefficient implies that when X is high wrt E[X], even Y tends to be high wrt E[Y]
  - When we observe high values of X it is likely that Y too is high
- A negative coefficient implies that when X is low, Y is likely to be high and vice-versa
- If X and Y are uncorrelated then no such trend is observed

ρ_{XY}^2

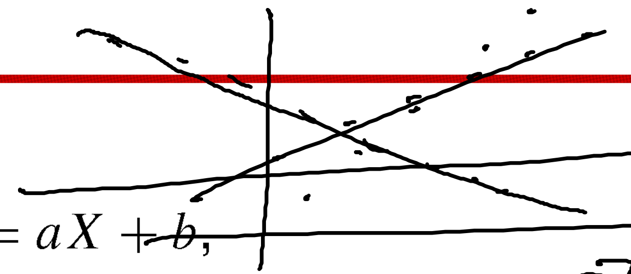
Examples (Scatter Plots)



Weisstein, Eric W. "Correlation Coefficient." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/CorrelationCoefficient.html

Y is an Affine Function of X

Theorem 4.18



If X and Y are random variables such that $Y = aX + b$,

$$\rho_{X,Y} = \begin{cases} -1 & a < 0,\\ 0 & a = 0,\\ 1 & a > 0.\end{cases}$$

$\mu_Y = E[Y]$

$= aE[X] + b$

$E[Y - \mu_Y]$

$= E[Y]$

$E[X(aX+b)]$

$= E[aX^2 + bX]$

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

$$\frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

$$\frac{E[XY] - \mu_X \mu_Y}{\sigma_X \sigma_Y}$$

$$\frac{aE[X^2] + bE[X] - E[X](aE[X]+b)}{}$$

Conditioning on an Event

Definition 4.9 Conditional Joint PMF

For discrete random variables X and Y and an event, B with $P[B] > 0$, the conditional joint PMF *of X and Y given B is*

$$P_{X,Y|B}(x, y) = P[X = x, Y = y|B].$$

$\{X=x, Y=y\}$

$$P[X=x, Y=y|B] = \frac{P[X=x, Y=y, B]}{P[B]}$$

Theorem 4.19

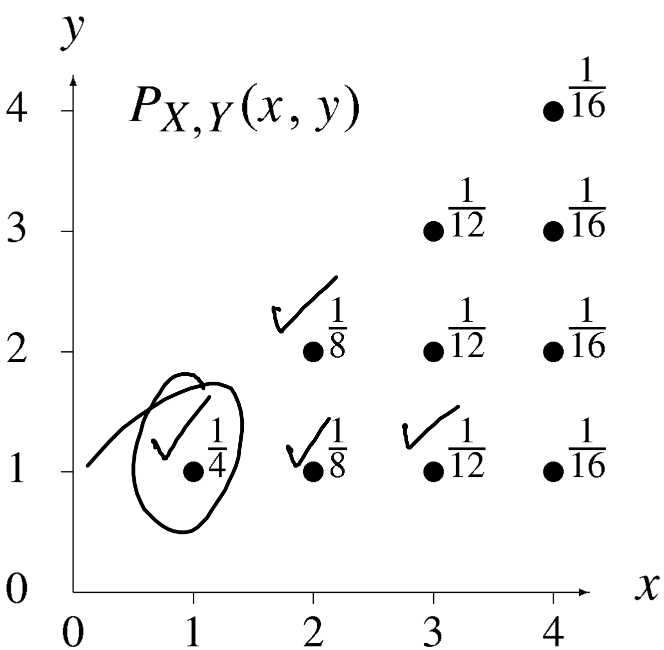
For any event B, a region of the X, Y plane with $P[B] > 0$,

$$P_{X,Y|B}(x, y) = \begin{cases} \frac{P_{X,Y}(x, y)}{P[B]} & (x, y) \in B, \\ 0 & \text{otherwise.} \end{cases}$$

VS Problem

Example 4.13 Problem

$$P[X=1, Y=1 \mid B] = \frac{1/4}{\left(\frac{1}{4}+\frac{1}{8}+\frac{1}{8}+\frac{1}{12}\right)}$$



Random variables X and Y have the joint PMF $P_{X,Y}(x, y)$ as shown. Let B denote the event $X + Y \leq 4$. Find the conditional PMF of X and Y given B.

$P[B] =$

- P[B] = 7/12

Conditioning on an Event

Definition 4.10 Conditional Joint PDF

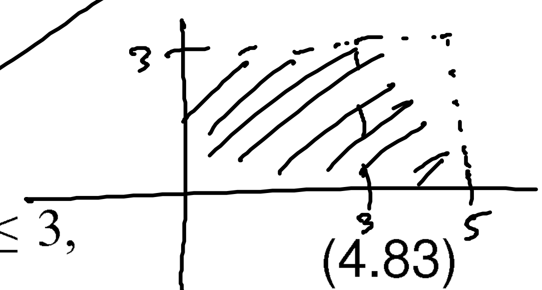
Given an event B with $P[B] > 0$, the conditional joint probability density function of X and Y is

$$f_{X,Y|B}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{P[B]} & (x,y) \in B, \\ 0 & \text{otherwise.} \end{cases}$$

$$\frac{P[x \le X < x+\Delta x,\ y \le Y < y+\Delta x]}{\Delta x \Delta y}$$

$$\approx \frac{P[x \le X < x+\Delta x,\ y \le Y < y+\Delta y \mid B]}{\Delta x \Delta y}$$

Example 4.14 Problem



X and Y are random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 1/15 & 0 \le x \le 5, 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases} \quad (4.83)$$

$y \ge 4 - x$

Find the conditional PDF of X and Y given the event $B = \{X + Y \ge 4\}$.

Conditioning on an Event

Theorem 4.20 Conditional Expected Value

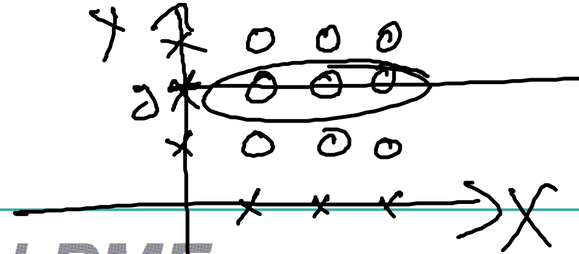
For random variables X and Y and an event B of nonzero probability, the conditional expected value of $W = g(X, Y)$ given B is

Discrete: $$E[W|B] = \sum_{x \in S_X} \sum_{y \in S_Y} g(x, y) P_{X,Y|B}(x, y)$$

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

- Nothing special about this
- You use the conditional PMF and PDF respectively

Conditioning on a RV



Definition 4.12 *Conditional PMF*

$P[Y=y]$

For any event $Y = y$ *such that* $P_Y(y) > 0$, *the* conditional PMF *of* X *given* $Y = y$ *is*

$$P_{X|Y}(x|y) = P[X = x|Y = y].$$

Theorem 4.22

For random variables X and Y with joint PMF $P_{X,Y}(x, y)$, and x and y such that $P_X(x) > 0$ and $P_Y(y) > 0$,

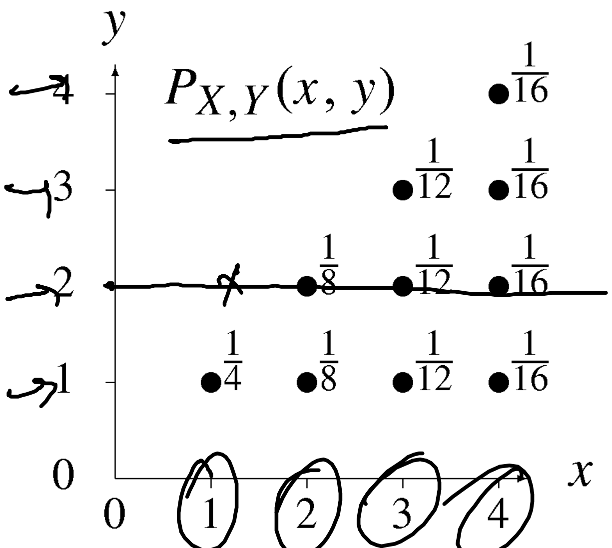
$$\underline{P_{X,Y}(x, y)} = \underbrace{P_{X|Y}(x|y)\,P_Y(y)} = \underbrace{P_{Y|X}(y|x)\,P_X(x)}.$$

$P[X=x|Y=y]$

- Proof anyone??

VS Problem

Example 4.17 Problem



Random variables X and Y have the joint PMF $P_{X,Y}(x, y)$, as given in Example 4.13 and repeated in the accompanying graph. Find the conditional PMF of Y given $X = x$ for each $x \in S_X$.

$$P[X=1 \mid Y=2] = \frac{P[X=1, Y=2]}{P[Y=2]}$$

- You have the joint PMF
- From it you can calculate the PMF of X and hence the conditional PMF of Y
- How many conditional PMF(s) of Y do we have?

VVVVS Problem

$S_Y = \{1,2,3,4\}$

- For a given y

$$P_{Y|X}(y|1) + P_{Y|X}(y|2) + P_{Y|X}(y|3) + P_{Y|X}(y|4) = ?$$

$$P[X=1 | S_X] + P[Y=2 | S_X] + \cdots + P[X=4 | S_X]$$

- For a given x

$$P_{Y|X}(1|x) + P_{Y|X}(2|x) + P_{Y|X}(3|x) + P_{Y|X}(4|x) = ?$$

$$P[Y=1 | X=x] + P[Y=2 | X=x] + \cdots + P[Y=4 | X=x]$$

$$\sum_{x \in S_X} P[X=x | Y=y] \qquad \sum_{y \in S_Y} P[X=x | Y=y]\, P[Y=y]$$

$$= \sum_{x \in S_X} \frac{P[X=x, Y=y]}{P[Y=y]} \longrightarrow \sum_{x \in S_X} P[X=x | Y=y]\, P[X=x]$$

Conditional Expectation

$E[P_X(X) \mid Y=y]$

Theorem 4.23 Conditional Expected Value of a Function

X and Y are discrete random variables. For any $y \in S_Y$, the conditional expected value of $g(X, Y)$ given $Y = y$ is

$$E[g(X, Y)|Y = y] = \sum_{x \in S_X} g(x, y) P_{X|Y}(x|y).$$

$E_Y\left[E\left[g(X,y) \mid Y \right] \right]$

$P_{X|Y}(x|y)$

- The conditional expectation is a function of which variable(s)?
- E[g(X,Y)] is a function of which variable(s)?

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

$$\longrightarrow \int_{x \in S_X} f_{X|Y}(x|y)\, dx$$

$$\longrightarrow \int_{y \in S_Y} f_{X|Y}(x|y)\, dy$$

Now For the Continuous Case

Definition 4.13 Conditional PDF

For y such that $f_Y(y) > 0$, the conditional PDF of X given $\{Y = y\}$ is

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

Theorem 4.24

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

Problem

- You are given the conditional pdf $f_{Y|X}(y|x)$ and $f_X(x)$

$$P[Y \leq y] = \int_{-\infty}^{y} \int_{-\infty}^{\infty} \underbrace{f_{X,Y}(x,y)}_{f_{Y|X}(y|x) f_X(x)} \, dx \, dy$$

- How would you find $F_Y(y)$?

Expected Values: The Continuous Case

Definition 4.14 Conditional Expected Value of a Function

For continuous random variables X *and* Y *and any* y *such that* $f_Y(y) > 0$, *the* conditional expected value *of* $g(X, Y)$ *given* $Y = y$ *is*

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

$$E[g(X,y)|Y=y] \longrightarrow \int g(x,y) f_{X|Y}(x|y)\, dx$$

$$E[g(X,y)] = \int g(x,y) f_X(x)\, dx$$

Conditional Expected Value

Definition 4.15 *Conditional Expected Value*

The conditional expected value $E[X|Y]$ is a function of random variable Y such that if $Y = y$ then $E[X|Y] = E[X|Y = y]$.

$E[X|Y] = g(Y)$

$E[X|Y=y] = \int_{S_X} x\, f_{X|Y}(x|y)\, dx$

- E[X|Y] returns E[X|Y=y] when Y=y
- Note that E[X|Y] is a random variable
  - It is a function of the RV ?
- E[X|Y=y] is a function of *y*
- E[E[X|Y]] is the expectation of a function of ~~RV Y~~
- E[E[X|Y=1]] = ?
  - E[X|Y=1] ?

$\int_{y \in S_Y} \int_{S_X} x\, f_{X|Y}(x|y)\, dx\, f_Y(y)\, dy$

$f_{X,Y}(x,y)\, dx\, dy$

$$E[X|Y=y] = \int_y^1 x \frac{1}{1-y} dx$$

$$= \frac{1}{1-y}\left(\frac{1-y^2}{2}\right)$$

$$= \frac{1+y}{2}$$

Example 4.20 Problem

For random variables X and Y in Example 4.5, we found in Example 4.19 that the conditional PDF of X given Y is

$$E[X|Y] = \frac{1+Y}{2}$$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} 1/(1-y) & y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.109}$$

Find the conditional expected values $E[X|Y=y]$ and $E[X|Y]$.

- $E[X|Y=y] = (1+y)/2$
- $E[X|Y] = (1+Y)/2$

Independent Random Variables

Definition 4.16 Independent Random Variables

Random variables X and Y are independent *if and only if*

Discrete: $$P_{X,Y}(x,y) = P_X(x)P_Y(y)$$

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

- For independent X and Y, $P_{X|Y}(x|y)$ = ?
  - $P_X(x)$
- What about $F_{X,Y}(x,y)$?

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

$P_X(x)P_Y(y)$

$P[X \le x, Y \le y] = P[X \le x]\,P[Y \le y]$

Independent RVs

Theorem 4.27

For independent random variables X and Y,

$$E[g(X)h(Y)]$$

$$= \int_{S_Y}\int_{S_X} g(x)h(y)\, f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_{S_Y}\int_{S_X} g(x)h(y)\, f_X(x) f_Y(y)\, dx\, dy$$

$$= E[g(X)]\,E[h(Y)]$$

(a) $E[g(X)h(Y)] =$,

(b) $r_{X,Y} = E[XY] =$ $E[X]E[Y]$

(c) $\text{Cov}[X,Y] =$ 0 ,

(d) $\text{Var}[X+Y] =$ $\text{Var}[X] + \text{Var}[Y]$,

(e) $E[X|Y=y] =$ $E[X]$,

(f) $E[Y|X=x] =$.

Theorem 4.27

For independent random variables X and Y,

(a) $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$,

(b) $r_{X,Y} = E[XY] = E[X]E[Y]$,

(c) $\text{Cov}[X, Y] = \rho_{X,Y} = 0$,

(d) $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$,

(e) $E[X|Y = y] = E[X]$ for all $y \in S_Y$,

(f) $E[Y|X = x] = E[Y]$ for all $x \in S_X$.

Example 4.24 Problem

$$f_{U,V}(u,v) = \begin{cases} 24uv & u \geq 0, v \geq 0, u+v \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{4.133}$$

$f_U(u)$ $f_V(v)$

Are U and V independent?

Quiz 4.10(A)

$P[X=x]\ P[Y=y] \qquad P[X=x, Y=y]$

Random variables X and Y in Example 4.1 and random variables Q and G in Quiz 4.2 have joint PMFs:

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=0$	0.01	0	0
$x=1$	0.09	0.09	0
$x=2$	0	0	0.81

$P_{Q,G}(q,g)$	$g=0$	$g=1$	$g=2$	$g=3$
$q=0$	0.06	0.18	0.24	0.12
$q=1$	0.04	0.12	0.16	0.08

$P_X(x) \longrightarrow P[X=1] = 0.18$

(1) Are X and Y independent?

$P_Y(y) \qquad P[X=2] = 0.81$

$P[Y=0] = 0.1 \qquad P[X=0] = 0.01$

(2) Are Q and G independent?

$P[Y=1] = 0.09$

$P[Y=2] = 0.81$

Quiz 4.10(B)

iid

Random variables X_1 and X_2 are independentand identically distributed with probability density function

$$\int_{-\infty}^{x_2}\int_{-\infty}^{x_1} f_{X_1,X_2}(x_1,x_2)\,dx_1\,dx_2$$

$$f_X(x) = \begin{cases} 1 - x/2 & 0 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases} \tag{4.144}$$

$0 \le z_1 \le z_2 \le$

(1) What is the joint PDF $f_{X_1,X_2}(x_1,x_2)$?

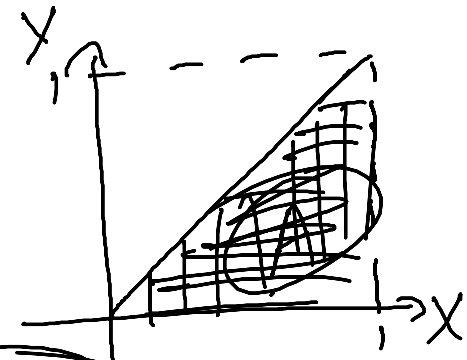
$$f_{X_1,X_2}(x_1,x_2) = \begin{cases} \left(1-\frac{x_1}{2}\right)\left(1-\frac{x_2}{2}\right) & 0 \le x_1 \le 2,\ 0 \le x_2 \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(2) Find the CDF of $Z = \max(X_1, X_2)$.

$$P[Z \le z] = P[\max(X_1,X_2) \le z] \to P[X_1 \le z,\ X_2 \le z]$$

Problem 4.10.13

X and Y are independent random variables with PDFs



$$\frac{P[x \le X < x+\Delta x \mid X>Y]}{\Delta x}$$

$$f_X(x) = \begin{cases} 2x & 0 \le x \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

$$f_Y(y) = \begin{cases} 3y^2 & 0 \le y \le 1, \\ 0 & \text{otherwise,} \end{cases}$$

Let $A = \{X > Y\}$. $\{X>y\}$ $\{X>Y\}$

$$f_{X|A}(x)$$

(a) What are $E[X]$ and $E[Y]$?

(b) What are $E[X|A]$ and $E[Y|A]$?

$$f_{X|\{X>0.5\}} = \begin{cases} \dfrac{f_X(x)}{P(X>0.5)} & 0.5 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_{X|\{X>Y\}}(x) = \int f_{X,Y|A}(x,y)\,dy$$

$$f_{X,Y|A}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{P(X>Y)} & \dfrac{(x,y)\in A}{x>y} \\ 0 & \end{cases}$$

$0 \le y \le x \le 1$

A Shorter Exponential

Let $X \sim Exp(\lambda)$ and $Y \sim Exp(\mu)$. Assume that X and Y are independent random variables. Find $f_{X|X<Y}(x)$.

HW

Extra Slides

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Problem 4.9.14

Suppose you arrive at a bus stop at time 0 and at the end of each minute, with probability p, a bus arrives, or with probability $1-p$, no bus arrives. Whenever a bus arrives, you board that bus with probability q and depart. Let T equal the number of minutes you stand at a bus stop. Let N be the number of buses that arrive while you wait at the bus stop.

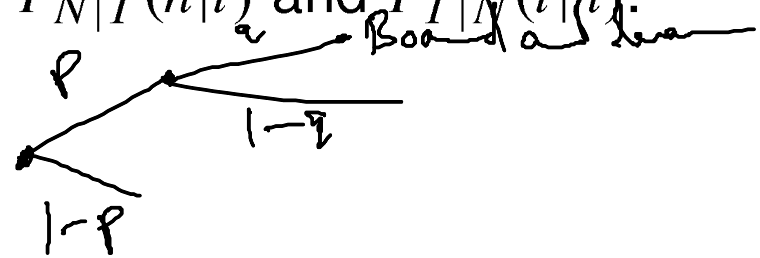
(a) Identify the set of points (n, t) for which $P[N = n, T = t] > 0$.

(b) Find $P_{N,T}(n, t)$.

$[p(1-q)]^{t-1}\, pq$

(c) Find the marginal PMFs $P_N(n)$ and $P_T(t)$.

(d) Find the conditional PMFs $P_{N|T}(n|t)$ and $P_{T|N}(t|n)$.



Problem 4.9.15

Each millisecond at a telephone switch, a call independently arrives with probability p. Each call is either a data call (d) with probability q or a voice call (v). Each data call is a fax call with probability r. Let N equal the number of milliseconds required to observe the first 100 fax calls. Let T equal the number of milliseconds you observe the switch waiting for the first fax call. Find the marginal PMF $P_T(t)$ and the conditional PMF $P_{N|T}(n|t)$. Lastly, find the conditional PMF $P_{T|N}(t|n)$.

Bivariate Gaussian Random Variables

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Joint Density of X and Y is Gaussian

Definition 4.17 Bivariate Gaussian Random Variables

X, Y

$\begin{bmatrix} E[X] \\ E[Y] \end{bmatrix}$

$\begin{bmatrix} Var[X] & Cov[X,Y] \\ Cov[X,Y] & Var[Y] \end{bmatrix}$

2x2

Random variables X and Y have a bivariate Gaussian PDF *with parameters μ_1, σ_1, μ_2, σ_2, and ρ if*

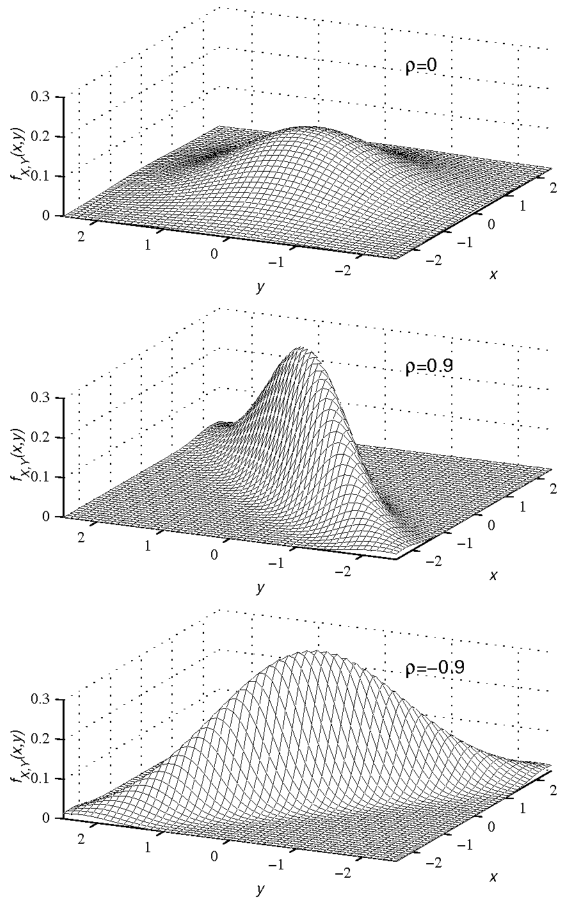
$$f_{X,Y}(x,y) = \frac{\exp\left[-\dfrac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

where μ_1 and μ_2 can be any real numbers, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < \rho < 1$.

- We say that the RV(s) X and Y are $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

Examples of a Joint Gaussian PDF

Figure 4.5



The Joint Gaussian PDF $f_{X,Y}(x, y)$ for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$, and three values of ρ.

If X and Y are Jointly Gaussian then they Marginally Gaussian

Theorem 4.28

If X and Y are the bivariate Gaussian random variables in Definition 4.17, X is the Gaussian (μ_1, σ_1) random variable and Y is the Gaussian (μ_2, σ_2) random variable:

$$f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} \qquad f_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(y-\mu_2)^2/2\sigma_2^2}.$$

- **How do we prove this?**
- Marginally Gaussian may not imply jointly Gaussian

For Jointly Gaussian RVs the Conditional PDF is Gaussian

Theorem 4.29

If X and Y are the bivariate Gaussian random variables in Definition 4.17, the conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-(y-\tilde{\mu}_2(x))^2/2\tilde{\sigma}_2^2},$$

where, given $X = x$, the conditional expected value and variance of Y are

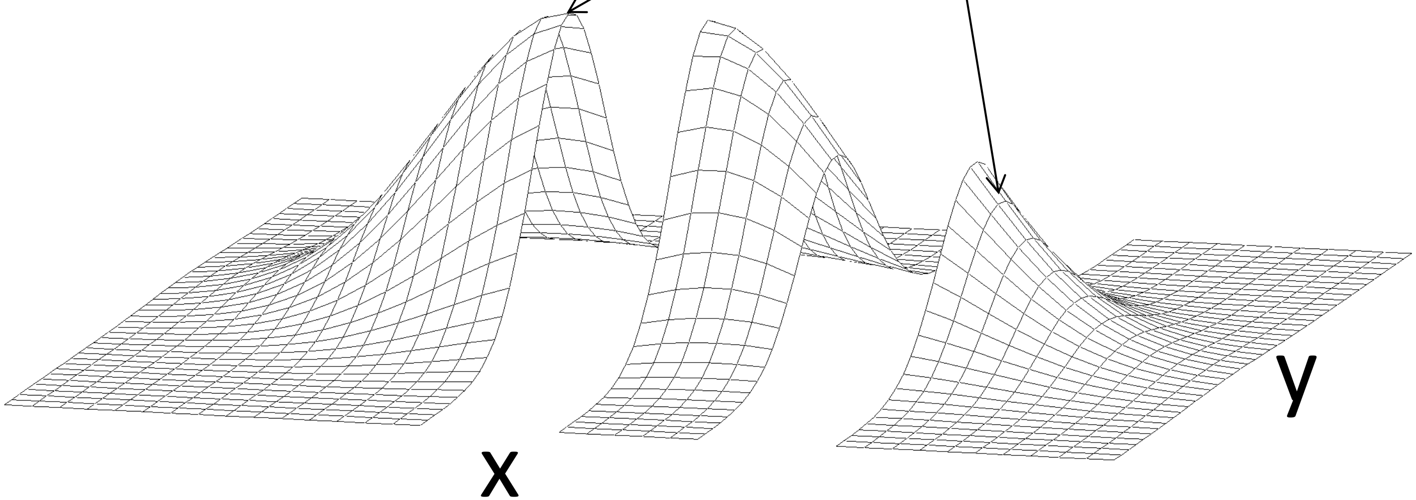
$$\underbrace{\tilde{\mu}_2(x)}_{E[Y|X=x]} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \qquad \underbrace{\tilde{\sigma}_2^2}_{Var[Y|X=x]} = \sigma_2^2(1 - \rho^2).$$

- E[Y|X] is a line that passes through (μ_1,μ_2). Also called regression line for any X and Y.
- (x, E[Y|X=x]) is also the point which maximizes $f_{Y|X}(y|x)$
- Following the values of $f_{X,Y}$ for points along the line implies following the maxima of all $f_{Y|X}(y|x)$

The Conditional Density

Figure 4.6

Each bell shaped curve corresponds to $f_{Y|X}(y|x)$ scaled by $f_X(x)$



Example Joint Gaussian Density

Correlation Coefficient

Theorem 4.31

Bivariate Gaussian random variables X and Y in Definition 4.17 have correlation coefficient

$$\rho_{X,Y} = \rho.$$

- We want to show that $\rho = \rho_{X,Y} = \mathrm{Cov}[XY]/(\sigma_X\ \sigma_Y) = \mathrm{Cov}[XY]/(\sigma_1\ \sigma_2)$
- $\mathrm{Cov}[XY] = E[(X - \mu_X)\ (Y - \mu_Y)] = E[(X - \mu_1)\ (Y - \mu_2)]$
- Using Iterated Expectation we get
  - $E[(X - \mu_1)\ (Y - \mu_2)] = E[(X - \mu_1)\ E[(Y - \mu_2)|X]]$
- Theorem 4.29 gave us $E[Y|X]$

Correlation Coefficient

- We have

$$E[Y-\mu_2|X=x] = E[Y|X=x]-\mu_2$$
$$= \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) - \mu_2$$

- Therefore $E[(X-\mu_1)\ E[(Y-\mu_2)|X]]$ can be written as

$$E[(X-\mu_1)\rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)] = \rho\frac{\sigma_2}{\sigma_1}E[(X-\mu_1)^2]$$

- Therefore, $\rho_{X,Y} = Cov[XY]/(\sigma_1\ \sigma_2) = \rho$

Uncorrelated and Independent

Theorem 4.32

Bivariate Gaussian random variables X and Y are uncorrelated if and only if they are independent.

- If they are independent, we know that they are uncorrelated (Covariance is 0)
- Given that X and Y are jointly Gaussian and that Cov[X,Y] = 0, we have $\rho=0$
- How do we show that X and Y are independent?

Uncorrelated => Independent for Jointly Gaussian

We know that the joint density is given by

Random variables X *and* Y *have a* bivariate Gaussian PDF *with parameters* $\mu_1, \sigma_1, \mu_2, \sigma_2$, *and* ρ *if*

$$f_{X,Y}(x, y) = \frac{\exp\left[-\frac{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

where μ_1 *and* μ_2 *can be any real numbers,* $\sigma_1 > 0$, $\sigma_2 > 0$, *and* $-1 < \rho < 1$.

Substituting $\rho = 0$, we get $f_{X,Y}(x, y) = f_X(x) f_Y(y)$

Example When RVs are Uncorrelated but not Independent (Also see link)

- Consider two coin tosses. Let the outcome of the first toss be denoted by the RV X ~ Bernoulli(0.5) and the outcome of the second toss be modeled as RV Y ~ Bernoulli(0.5). Let the outcomes be independent of each other

- Clearly Cov[X,Y] = 0.
- Let $Z = X - Y$ and $W = X + Y$, $S_Z = \{-1,0,1\}$ and $S_W = \{0,1,2\}$
- $P[Z=0] = ½$
- $P[Z=0|W=2] = P[X=Y|W=2] = 1$
- $P[Z=1] = ¼$ and $P[Z=1|W=2] = 0$
- Clearly W and Z are not independent
- However $Cov[W,Z] = E[(X - Y)(X+Y-2\mu)] = 0$
- W and Z are uncorrelated but NOT independent

Variance of the Conditional Density

- The variance of the conditional density was defined in **Theorem 4.29**

If X and Y are the bivariate Gaussian random variables in Definition 4.17, the conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \frac{1}{\tilde{\sigma}_2\sqrt{2\pi}} e^{-(y-\tilde{\mu}_2(x))^2/2\tilde{\sigma}_2^2},$$

where, given $X = x$, the conditional expected value and variance of Y are

$$\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \qquad \tilde{\sigma}_2^2 = \sigma_2^2(1 - \rho^2).$$

- Var[Y|X] <= Var[Y], since ρ is the corr coeff and hence|ρ| <= 1
- If Y and X are correlated, knowing X is likely to reduce the *uncertainty* in Y

Buffon's Needle!

A fine needle of length $2a$ is dropped at random on a board covered with parallel lines distance $2b$ apart where $b > a$ as in figure. What is the probability that the needle intersects one of the lines?

Linear Combinations of Normal RV(s)

1. Let $Z = aX + bY$, where X and Y are **jointly normal**. Then Z is normally distributed.
   - How will you show this?
   - Find $P[Z <= z]$, then $f_Z(z)$ and use the jointly normal distribution. Do this for at least when X and Y are independent
2. Let $Z = aX + bY$ and $W = cX + dY$. X and Y are **jointly normal.** Then Z and W are jointly normal
   - Proof requires additional machinery
   - Note that 1 follows from 2

Probability and Random Processes

Sums of RV

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Expected Value of Sum is Sum of Expected Values

Theorem 6.1

For any set of random variables $X_1, \ldots, X_n$, the expected value of $W_n = X_1 + \cdots + X_n$ is

$$E[W_n] = E[X_1] + E[X_2] + \cdots + E[X_n].$$

- Proof by induction
- We earlier proved that $E[W_2] = E[X_1] + E[X_2]$
- Assume $E[W_{n-1}] = E[X_1] + E[X_2] + \ldots + E[X_{n-1}]$
- Show $E[W_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$
  - $E[W_n] = E[W_{n-1} + X_n] = E[W_{n-1}] + E[X_n]$ (from the base case). Hence proved

Expected Value of Sum is Sum of Expected Values

- The theorem is valid for any joint distribution of the random variables $X_1, X_2, \ldots, X_n$
- The variables do not need to be independent! Can have ANY joint distribution
- Therefore a very powerful theorem!
- Can you show why the above is true using the chain rule for a pdf?
- The chain rule states that

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = f_{X_2,\ldots,X_n|X_1}(x_2,\ldots,x_n|x_1)f_{X_1}(x_1)$$
$$= f_{X_n|X_{n-1}\ldots X_1}(x_n|x_{n-1},\ldots,x_1)f_{X_{n-1}|X_{n-2}\ldots X_1}(x_{n-1}|x_{n-2},\ldots,x_1)\ldots f_{X_1}(x_1)$$

Variance of a Sum Of Random Variables

Theorem 6.2

The variance of $W_n = X_1 + \cdots + X_n$ is

$$\text{Var}[W_n] = \sum_{i=1}^{n} \text{Var}[X_i] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Cov}\left[X_i, X_j\right].$$

$\text{Cov}[X_i, X_j] = E[(X_i - E[X_i])(X_j - E[X_j])]$

$E[X_i X_j] - E[X_i]E[X_j]$

- This is also the sum of the elements of the Covariance Matrix of the random variables X_1, X_2, …, X_n

$E[X_i X_j] = \int\int x_i x_j f_{X_i X_j}(x_i, x_j) \, dx_i \, dx_j$

$x_j \in S_{X_j}$ $x_i \in S_{X_i}$

$f_{X_i}(x_i) f_{X_j}(x_j)$

Theorem 6.3

If X, Y are independent

then $E[XY] = E[X]E[Y]$

When $X_1, \ldots, X_n$ are uncorrelated,

$$\mathrm{Var}[W_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n].$$

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

$x_1, x_2, \ldots, x_n$

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

$W_n/n =$

$$\frac{X_1 + X_2 + \cdots + X_n}{n}$$

0.5

0.5, 0.5

1 0

PDF of Sum of Two Random Variables

Theorem 6.4

$X \le w - y$

$P[W \le w] = P[X + Y \le w]$

The PDF of $W = X + Y$ is

$$f_W(w) = \int_{-\infty}^{\infty} f_{X,Y}(x, w-x)\, dx = \int_{-\infty}^{\infty} \boxed{f_{X,Y}(w-y, y)}\, dy.$$

$$\frac{d}{dw}\int_{-\infty}^{\infty}\int_{-\infty}^{w-y} f_{X,Y}(u,v)\, du\, dv$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{w-y} f_{X,Y}(u,v)\, du\, dv$$

$$\int_{-\infty}^{\infty} f_{X,Y}(w-y, v)\, dv$$

Theorem 6.5

When X and Y are independent random variables, the PDF of $W = X + Y$ is

$$f_W(w) = \int_{-\infty}^{\infty} f_X(w-y)\, f_Y(y)\, dy = \int_{-\infty}^{\infty} f_X(x)\, f_Y(w-x)\, dx.$$

Moment Generating Function

Moment Generating Function

Definition 6.1 *(MGF)*

$$\frac{d}{ds}\phi_X(s) = \frac{d}{ds}\int_{-\infty}^{\infty} e^{sX} f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} \frac{d}{ds} e^{sX} f_X(x)\,dx$$

For a random variable X*, the* moment generating function (MGF) *of* X *is*

$$\phi_X(s) = E\left[e^{sX}\right]. \longrightarrow \int_{-\infty}^{\infty} e^{sX} f_X(x)\,dx$$

- What about the continuous case?

$$\frac{d}{ds}\phi_X(s) = \int_{-\infty}^{\infty} X e^{sX} f_X(x)\,dx$$

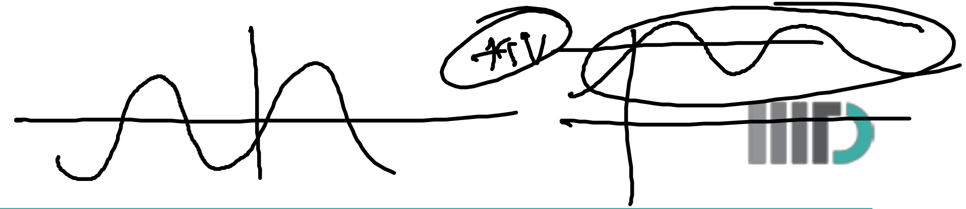
- One-to-one mapping between MGF of a RV and its PDF
  - See table in textbook

Calculating Moments Using the MGF

Theorem 6.6

A random variable X with MGF $\phi_X(s)$ has nth moment

$$\underline{E\left[X^n\right]} = \left.\frac{d^n \phi_X(s)}{ds^n}\right|_{s=0}.$$



Example 6.5 Problem

X is an exponential random variable with MGF $\phi_X(s) = \lambda/(\lambda - s)$. What are the first and second moments of X? Write a general expression for the nth moment.

$$f_X(x) = \lambda e^{-\lambda x} \quad x \ge 0.$$

$$\phi_X(s) = E\left[e^{sX}\right] = \int_0^\infty e^{sx} \lambda e^{-\lambda x} \, dx$$

MGF of a Scaled and Shifted RV

Theorem 6.7

The MGF of $Y = aX + b$ is $\phi_Y(s) = e^{sb}\phi_X(as)$.

$E[e^{sY}]$

Theorem 6.8: MGF of Sum of Independent Random Variables

For a set of independent random variables $X_1, \ldots, X_n$, the moment generating function of $W = X_1 + \ldots + X_n$ is

$$E[e^{sW}] = E[e^{s(X_1 + X_2 + \cdots + X_n)}]$$

$$\phi_W(s) = \phi_{X_1}(s)\phi_{X_2}(s)\ldots\phi_{X_n}(s). \quad = E[e^{sX_1} e^{sX_2} \cdots e^{sX_n}]$$

$$= E[e^{sX_1}] \cdots E[e^{sX_n}]$$

When $X_1, \ldots, X_n$ are iid, each with MGF $\phi_{X_i}(s) = [\phi_X(s)]^n$,

$$\underline{\phi_W(s) = (\phi_X(s))^n.}$$

Example 6.6 Problem

J and K are independent random variables with probability mass functions

$$P_J(j) = \begin{cases} 0.2 & j = 1, \\ 0.6 & j = 2, \\ 0.2 & j = 3, \\ 0 & \text{otherwise,} \end{cases} \qquad P_K(k) = \begin{cases} 0.5 & k = -1, \\ 0.5 & k = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{6.40}$$

Find the MGF of $M = J + K$? What are $E[M^3]$ and $P_M(m)$?

- Note that we are looking at discrete random variables
- Therefore

$$\phi_M(s) = \sum_{m_i \in S_M} e^{sm_i} P_M(m_i)$$

Sum of Independent Poisson Random Variables is Poisson Distributed

Theorem 6.9

If $K_1, \ldots, K_n$ are independent Poisson random variables, $W = K_1 + \cdots + K_n$ is a Poisson random variable.

- The PMF of Poisson RV is

$$P_{K_i}(x) = \begin{cases} \alpha_i^x e^{-\alpha_i}/x! & x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

- The MGF is

$$\phi_{K_i}(s) = e^{(\alpha_i)(e^s - 1)}$$

- Now we use Theorem 6.8

Sum of Independent Gaussian Random Variables is Gaussian Distributed

Theorem 6.10

The sum of n independent Gaussian random variables $W = X_1 + \cdots + X_n$ is a Gaussian random variable.

- The MGF of a Gaussian is

$$\phi_{X_i}(s) = e^{s\mu_i + s^2\sigma_i^2/2}$$

- Can you show the above?

Sum of Independent Exponentials

Theorem 6.11

If $X_1, \ldots, X_n$ are iid exponential (λ) random variables, then $W = X_1 + \cdots + X_n$ has the Erlang PDF

$$f_W(w) = \begin{cases} \frac{\lambda^n w^{n-1} e^{-\lambda w}}{(n-1)!} & w \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

- The MGF of an exponential is

$$\phi_{X_i}(s) = \left(\frac{\lambda}{\lambda - s}\right)$$

Quiz 6.4(B)

Let $X_1, \ldots, X_n$ be independent Gaussian random variables with $E[X_i = 0]$ and $\mathrm{Var}[X_i] = i$. Find the PDF of

$$W = \alpha X_1 + \alpha^2 X_2 + \cdots + \alpha^n X_n. \tag{6.54}$$

e^{sW}

$\uparrow$

Problem 6.8.4

In a subway station, there are exactly enough customers on the platform to fill three trains. The arrival time of the nth train is $X_1 + \cdots + X_n$ where $X_1, X_2, \ldots$ are iid exponential random variables with $E[X_i] = 2$ minutes. Let W equal the time required to serve the waiting customers.

MGF of W $\longleftrightarrow$ CDF of W

Find $P[W > 20]$.

Central Limit Theorem

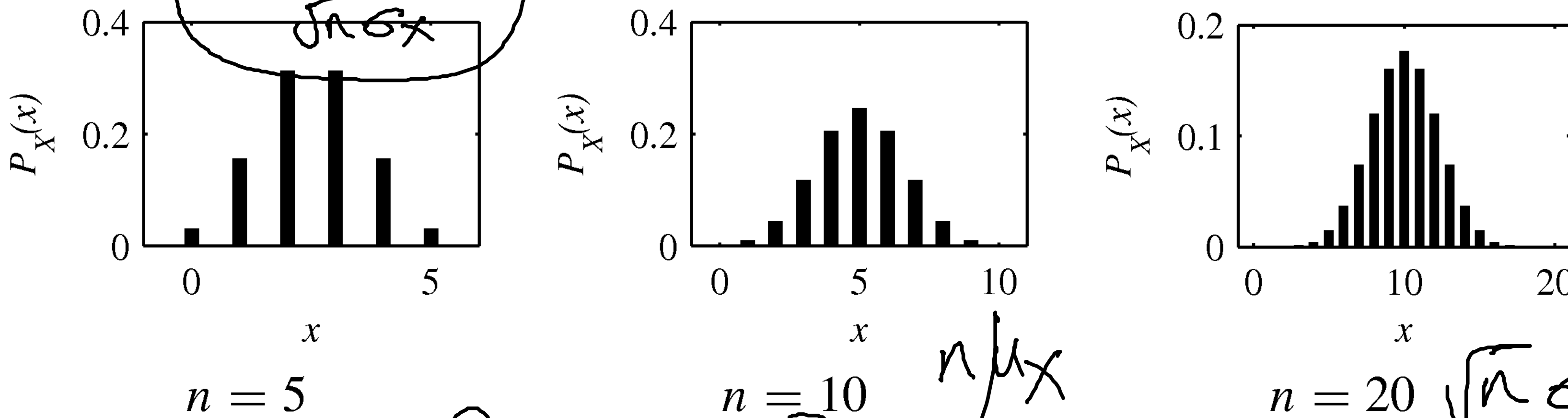
INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Central Limit Theorem

Figure 6.1

Binomial: Sum of which RV(s)? Bernoulli (p) iid

$\frac{W_n}{\sqrt{n}} \rightarrow \frac{W_n - n\mu_X}{\sqrt{n}\sigma_X} \xrightarrow{n \to \infty}$



$n\mu_X$ $\quad \sqrt{n}\,\sigma_X^2$

$Var[X_1 + \cdots + X_n] = n\sigma_X^2$

The PMF of the X, the number of heads in n coin flips for $n = 5, 10, 20$. As n increases, the PMF more closely resembles a bell-shaped curve.

- Which PMF are we talking about? Mean, Variance

$W_n = X_1 + X_2 + \cdots + X_n$

$X_i \sim X_j \quad E[X] = \mu_X, \; Var[X] = \sigma_X^2$

Central Limit Theorem

Theorem 6.14 Central Limit Theorem

Given $X_1, X_2, \ldots$, a sequence of iid random variables with expected value μ_X and variance σ_X^2, the CDF of $Z_n = (\sum_{i=1}^n X_i - n\mu_X)/\sqrt{n\sigma_X^2}$ has the property

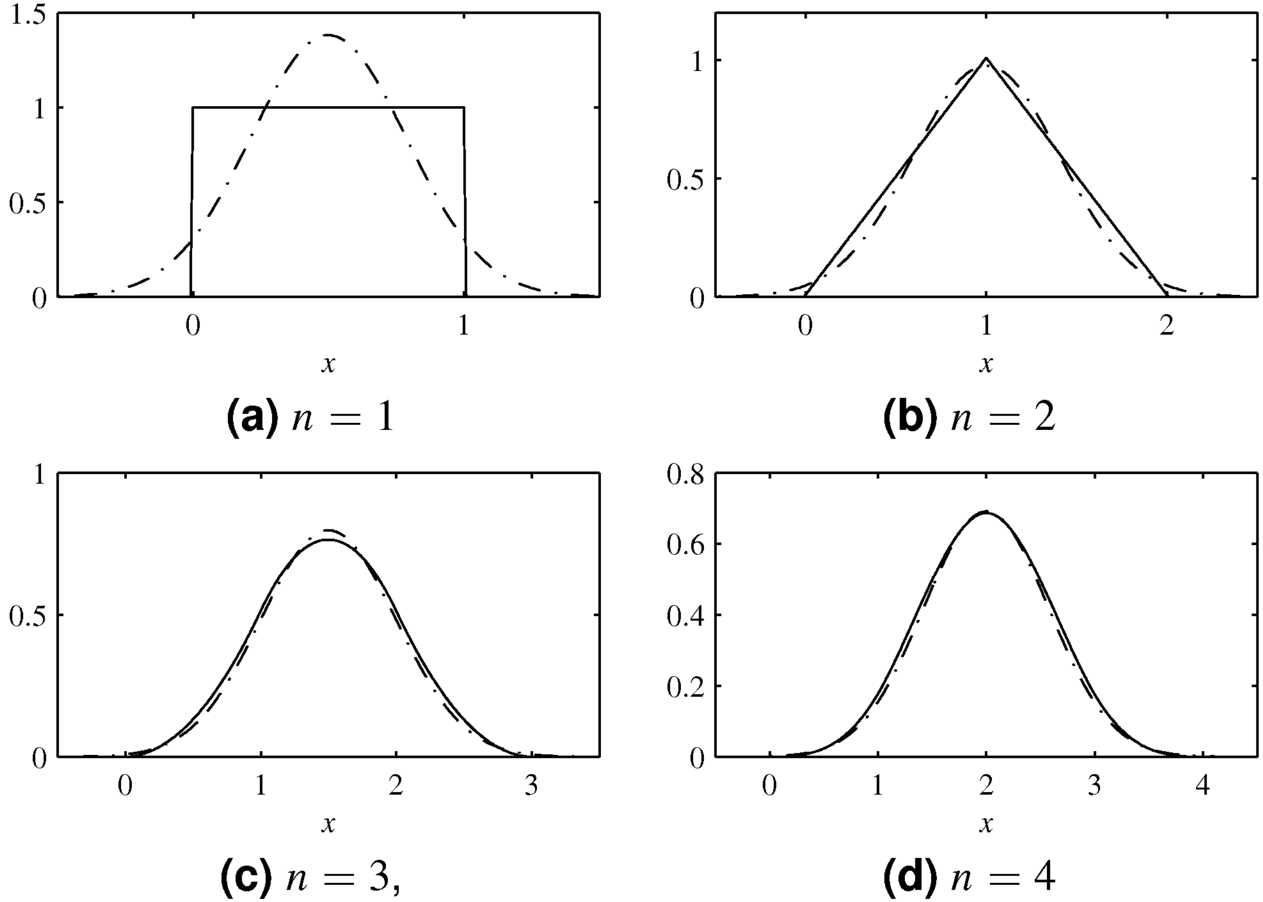
$$\lim_{n\to\infty} F_{Z_n}(z) = \Phi(z).$$

- Equivalently, we can write

$$\frac{S_n - nE[X]}{\sqrt{n\sigma_X^2}} \xrightarrow{D} \Phi(z) \text{ as } n \to \infty.$$

- That is Z_n *converges in distribution* to the standard normal distribution

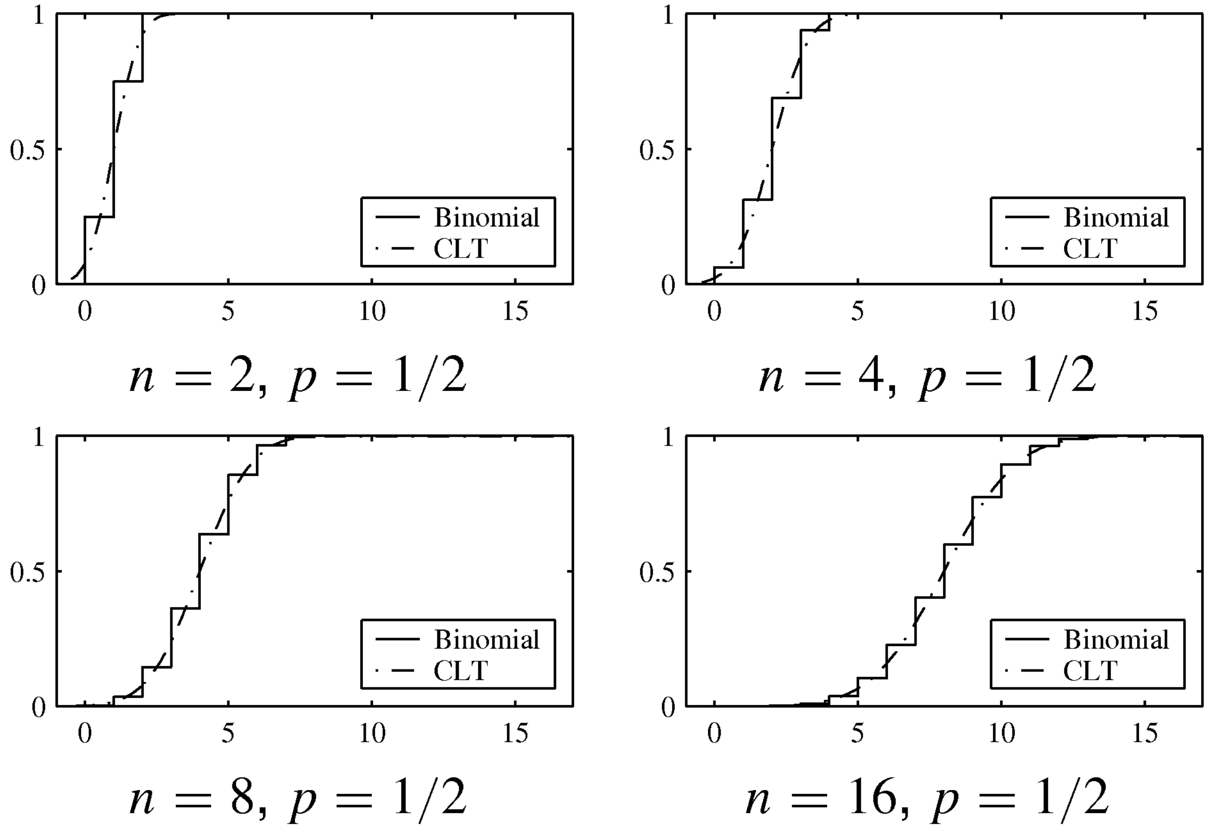
Figure 6.2



The PDF of W_n, the sum of n uniform $(0,1)$ random variables, and the corresponding central limit theorem approximation for $n = 1, 2, 3, 4$. The solid — line denotes the PDF $f_{W_n}(w)$, while the $- \cdot -$ line denotes the Gaussian approximation.

Note that here n=4 is large enough!

Figure 6.3



The binomial (n, p) CDF and the corresponding central limit theorem approximation for $n = 4, 8, 16, 32$, and $p = 1/2$.

Central Limit Theorem

Definition 6.2 Approximation

Let $W_n = X_1 + \cdots + X_n$ be the sum of n iid random variables, each with $E[X] = \mu_X$ and $\mathrm{Var}[X] = \sigma_X^2$. The central limit theorem approximation to the CDF of W_n is

$$F_{W_n}(w) \approx \Phi\left(\frac{w - n\mu_X}{\sqrt{n\sigma_X^2}}\right).$$

Quiz 6.6

The random variable X milliseconds is the total access time (waiting time + read time) to get one block of information from a computer disk. X is uniformly distributed between 0 and 12 milliseconds. Before performing a certain task, the computer must access 12 different blocks of information from the disk. (Access times for different blocks are independent of one another.) The total access time for all the information is a random variable A milliseconds.

(1) What is $E[X]$, the expected value of the access time?

(2) What is $\mathrm{Var}[X]$, the variance of the access time?

(3) What is $E[A]$, the expected value of the total access time? $\rightarrow 12 \times 6$

(4) What is σ_A, the standard deviation of the total access time? $\rightarrow n\left(\frac{12^2}{12}\right) = 12n$

(5) Use the central limit theorem to estimate $P[A > 75 \text{ ms}]$, the probability that the total access time exceeds 75 ms.

$P[A - 72 > 3] = P\left(\frac{A-72}{\sqrt{144}} > \frac{3}{\sqrt{144}}\right)$

(6) Use the central limit theorem to estimate $P[A < 48 \text{ ms}]$, the probability that the total access time is less than 48 ms.

$P[\text{Bit is } 1] = p$

$P[499000 \le \text{No. of ones} \le 501000]$

$= P[\text{No of ones} \le 501000]$

$- P[\quad]$

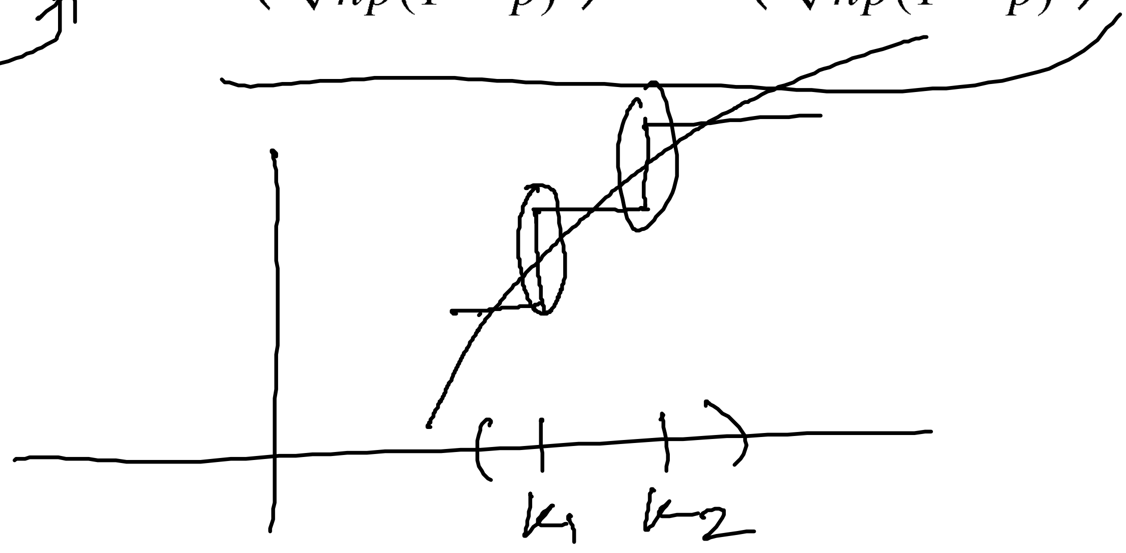
Example 6.15 Problem

Transmit one million bits. Let A denote the event that there are at least 499,000 ones but no more than 501,000 ones. What is $P[A]$?

Definition 6.3 De Moivre–Laplace Formula

For a binomial (n, p) *random variable* K,

$$P\left[k_1 \le K \le k_2\right] \approx \Phi\left(\frac{k_2 + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - 0.5 - np}{\sqrt{np(1-p)}}\right).$$



Quiz 6.7

Telephone calls can be classified as voice (V) if someone is speaking or data (D) if there is a modem or fax transmission. Based on a lot of observations taken by the telephone company, we have the following probability model: $P[V] = 3/4$, $P[D] = 1/4$. Data calls and voice calls occur independently of one another. The random variable K_n is the number of voice calls in a collection of n phone calls.

(1) What is $E[K_{48}]$, the expected number of voice calls in a set of 48 calls?

(2) What is $\sigma_{K_{48}}$, the standard deviation of the number of voice calls in a set of 48 calls?

(3) Use the central limit theorem to estimate $P[30 \leq K_{48} \leq 42]$, the probability of between 30 and 42 voice calls in a set of 48 calls.

(4) Use the De Moivre–Laplace formula to estimate $P[30 \leq K_{48} \leq 42]$.

Probability and Random Processes

Sanjit Kaul

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Parameter Estimation

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Sampling From a Finite Population

$\underline{n}$ $\quad \underbrace{k=2}$

$X_1, X_2, \ldots, X_n$

$X_i \sim X$

$X^{(1)}$

$X^{(2)}$

Are $X^{(1)}$ and $X^{(2)}$ independent.

Sampling w/o replacement
with

Definition 7.1 ***Sample Mean***

For iid random variables $X_1, \ldots, X_n$ *with PDF* $f_X(x)$*, the* sample mean *of* X *is the random variable*

$$M_n(X) = \frac{X_1 + \cdots + X_n}{n}.$$

Theorem 7.1

The sample mean $M_n(X)$ has expected value and variance

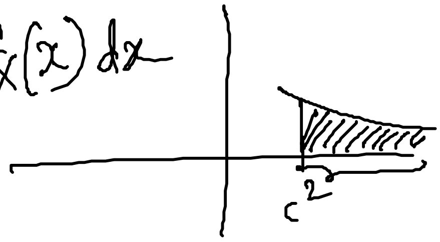
$$E\left[M_n(X)\right] = E\left[X\right], \qquad \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}.$$

Quiz 7.1

Let X be an exponential random variable with expected value 1. Let $M_n(X)$ denote the sample mean of n independent samples of X. How many samples n are needed to guarantee that the variance of the sample mean $M_n(X)$ is no more than 0.01?

$$E[X] = \int_0^\infty x f_X(x)\,dx = \int_0^{c^2} x f_X(x)\,dx + \int_{c^2}^\infty x f_X(x)\,dx$$

$$\geq \int_{c^2}^\infty x f_X(x)\,dx \geq \int_{c^2}^\infty c^2 f_X(x)\,dx$$



Theorem 7.2 Markov Inequality

For a random variable X such that $P[X < 0] = 0$ and a constant c,

$$\int_{c^2}^\infty f_X(x)\,dx \qquad P\left[X \geq c^2\right] \leq \frac{E[X]}{c^2} \qquad Z = (Y - E[Y])^2$$

$$P\left[|Y - E[Y]| \leq c\right] \geq \frac{Var[Y]}{c^2} \qquad P\left[Z \geq c^2\right] \leq \frac{Var[Y]}{c^2}$$

$$P\left[(Y - E[Y])^2 \geq c^2\right] \leq \frac{Var[Y]}{c^2}$$

Theorem 7.3 Chebyshev Inequality

For an arbitrary random variable Y and constant $c > 0$,

$$P\left[|Y - \mu_Y| \geq c\right] \leq \frac{\text{Var}[Y]}{c^2}.$$

$$|Y - \mu_Y| \geq c \iff (Y - \mu_Y)^2 \geq c^2$$

$$-(Y - \mu_Y) \geq c$$

Quiz 7.2

Elevators arrive randomly at the ground floor of an office building. Because of a large crowd, a person will wait for time W in order to board the third arriving elevator. Let X_1 denote the time (in seconds) until the first elevator arrives and let X_i denote the time between the arrival of elevator $i-1$ and i. Suppose X_1, X_2, X_3 are independent uniform $(0, 30)$ random variables. Find upper bounds to the probability W exceeds 75 seconds using

$E[W]=45$ 0   75 $P[W>\underline{75}] \leq \dfrac{E[W]}{75}$

(1) the Markov inequality,

$|W-45|>30$ $-(W-45)>30$ $P[W-E[W] > 75-E[W]]$

(2) the ~~Chebyshev ine~~quality. $W-45$

<-30

$W<-75$

$\boxed{P[|W-\mu_W| \geq c^2] \leq \dfrac{\mathrm{Var}[W]}{c^2}}$

$P[|W-45|>30]$

$W-45>30$

$p = \frac{1}{36}$ $\quad P[R = r] = {}^{r-1}C_2 \, p^2 (1-p)^{r-3} p \quad P[R \geq r] = \sum$

Problem 7.2.4

$R = X_1 + X_2 + X_3 \qquad E[R] = 3E[X] = \frac{3}{p}$

In a game with two dice, the event *snake eyes* refers to both dice showing one spot. Let R denote the number of dice rolls needed to observe the third occurrence of *snake eyes*. Find

$P[R \geq (\sqrt{250})^2] \leq \frac{3/p}{250}$

(a) the upper bound to $P[R \geq 250]$ based on the Markov inequality,

(b) the upper bound to $P[R \geq 250]$ based on the Chebyshev inequality,

(c) the exact value of $P[R \geq 250]$.

$P[|R - \mu_R| \geq c] \leq \frac{\mathrm{Var}[R]}{c^2}$

$P[|R - 3(36)| \geq c] \leq \frac{\mathrm{Var}[R]}{c^2}$

$\{|R - 108| \geq 142\} = \{-(R-108) \geq 142\} \cup \{R - 108 \geq 142\}$

$P[R \geq 250]$

$= P[R - 108 \geq 142]$

Section 7.3

Point Estimates of Model Parameters

$\hat{R}_1 = \frac{X_1}{1}$ $\hat{R}_2 = \frac{X_1 + X_2}{2}$ $\hat{R}_3 = \frac{X_1 + X_2 + X_3}{3}$ if estimator is sample mean

$r = E[X]$

Definition 7.2 Consistent Estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *of the parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n\to\infty} P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

$X \leftarrow \pi$

$$E\left[\frac{X_1+X_2+X_3}{3}\right] = E[X]$$

Definition 7.3 Unbiased Estimator

An estimate, $\hat{R}$*, of parameter* r *is* unbiased *if* $E[\hat{R}] = r$*; otherwise,* $\hat{R}$ *is* biased*.*

Definition 7.4 Asymptotically Unbiased Estimator

The sequence of estimators $\hat{R}_n$ *of parameter* r *is* asymptotically unbiased *if*

$$\lim_{n \to \infty} E[\hat{R}_n] = r.$$

Definition 7.5 **Mean Square Error**

The mean square error *of estimator* $\hat{R}$ *of parameter* r *is*

$$e = E\left[(\hat{R} - r)^2\right].$$

$$e_n = E\left[\underbrace{(\hat{R}_n - r)^2}\right]$$

For unbiased $\hat{R}_n$

$$e_n = E\left[(\hat{R}_n - E[\hat{R}_n])^2\right] = \mathrm{Var}[\hat{R}_n]$$

Theorem 7.4

Fact given that estimates are unbiased

If a sequence of unbiased estimates $\hat{R}_1, \hat{R}_2, \ldots$ of parameter r has mean square error $e_n = \mathrm{Var}[\hat{R}_n]$ satisfying $\lim_{n\to\infty} e_n = 0,$ then the sequence $\hat{R}_n$ is consistent.

Given

$$\lim_{n\to\infty} P\left[|\hat{R}_n - r| \geq \epsilon\right] = 0$$

To prove

$$P\left[|\hat{R}_n - r| \geq \epsilon\right] = P\left[|\hat{R}_n - \underbrace{E[\hat{R}_n]}| \geq \epsilon\right] \leq \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2}$$

$$E[\hat{R}_k] = E\left[\frac{N_k}{k}\right] = \frac{1}{k} E[N_k] = r$$

Example 7.4 Problem

In any interval of k seconds, the number N_k of packets passing through an Internet router is a Poisson random variable with expected value $E[N_k] = kr$ packets. Let $\hat{R}_k = N_k/k$ denote an estimate of r. Is each estimate $\hat{R}_k$ an unbiased estimate of r? What is the mean square error e_k of the estimate $\hat{R}_k$? Is the sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ consistent?

$$e_k = E\left[(\hat{R}_k - r)^2\right] = E\left[(\hat{R}_k - E[\hat{R}_k])^2\right] = \mathrm{Var}[\hat{R}_k]$$
$$= \frac{\mathrm{Var}[N_k]}{k^2}$$

Theorem 7.5

The sample mean $M_n(X)$ is an unbiased estimate of $E[X]$.

Theorem 7.6

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = E\left[(M_n(X) - E[X])^2\right] = \text{Var}[M_n(X)] = \frac{\text{Var}[X]}{n}.$$

This is also the variance of the sample mean estimator. Square-root of this is called the standard error.
Given the mean and variance of the estimator, what can we say about the sample mean that results from n IID trials?

INEQUALITIES

MARKOV INEQ

Let X be a non-negative R.V.

We want to find an upper bound for the tail probability.

$$E[X] = \int_0^\infty x f_X(x)\,dx$$
$$= \int_0^{c^2} x f_X(x)\,dx + \int_{c^2}^\infty x f_X(x)\,dx$$
$$\geq \int_0^{c^2} x f_X(x)\,dx + \int_{c^2}^\infty c^2 f_X(x)\,dx$$
$$\geq c^2 \int_{c^2}^\infty f_X(x)\,dx$$
$$\Rightarrow \boxed{P[X > c^2] \leq \frac{E[X]}{c^2}}$$

Let $X = (Y-\mu_Y)^2$ where $E[Y] = \mu_Y$.

$$\therefore P[(Y-\mu_Y)^2 > c^2] \leq \frac{E[(Y-\mu_Y)^2]}{c^2}$$
$$= \frac{Var[Y]}{c^2}$$

2nd order statistic

We get
$$\boxed{P[|Y-\mu_Y| > c] \leq \frac{Var[Y]}{c^2}}$$

This is called the Chebyshev Ineq.

Obs: The smaller the $Var[Y]$, the more unlikely that a value drawn from Y is far from the $E[Y]$.

Obs: Chebyshev is true for any R.V. RV does not need to be non-negative.

Problem 7.2.4

$P[\text{Occurrence of a snake eye}] = \left(\frac{1}{6}\right)^2 \triangleq p.$

Let N be the number of dice rolls needed to observe the third occurrence of snake eyes.

$$P[N=n] = p \; {}^{N-1}C_2 \; p^2 (1-p)^{N-1-2}$$
$$= \frac{(N-1)(N-2)}{2} p^3 (1-p)^{N-3}$$

$E[N] = E[X_1 + X_2 + X_3]$

where $X_i \sim \text{Geom}(p)$. Let X identically dist. as X_i

$$\therefore E[N] = \frac{3}{p}. \quad \text{——— (1)}$$

$$Var[N] = E[N^2] - (E[N])^2$$
$$= E[(X_1+X_2+X_3)^2] - \frac{9}{p^2}$$
$$= E[X_1^2 + X_2^2 + X_3^2 + 2X_1X_2 + 2X_2X_3 + 2X_3X_1] - \frac{9}{p^2}$$
$$= 3E[X^2] + 6E[X_1X_2] - \frac{9}{p^2}$$
$$= 3E[X^2] + 6(E[X])^2 - 9/p^2$$
$$E[X^2] = \frac{1-p}{p^2} + \frac{1}{p^2} = \frac{2-p}{p^2}$$
$$\therefore Var[N] = \frac{3(2-p)}{p^2} + \frac{6}{p^2} - \frac{9}{p^2} = \frac{6-3p+6}{p^2} - \frac{9}{p^2}$$
$$= \frac{3(1-p)}{p^2}$$

(1) & (2) help give bounds based of Markov & Chebyshev

$$P[R \geq 250] \leq \frac{E[R]}{250} = \frac{3/p}{250}$$
$$= \left(\frac{3}{250}\right) 36$$
$$= \frac{108}{250} = \frac{54}{125}$$

From Chebyshev:

$$P[R \geq 250]$$
$$= P[R - \mu_R \geq 250 - \mu_R]$$
$$= P[R - 108 \geq 142]$$
$$= P[|R-108| \geq 142] \quad \because R \geq 0$$
$$\leq \frac{Var[R]}{(142)^2} = \frac{3(1-p)}{(142)^2 p^2} = 0.1875 \quad \because p = \frac{1}{36}.$$

$|R-108| \geq 142$

If $R - 108 < 0$,

$-(R-108) \geq 142$

$(R-108) \leq -142$

$R \leq -34$

If $R - 108 \geq 0$

$\Rightarrow R - 108 \geq 142$

$R \geq 108 + 142.$

$\therefore P[|R-108| \geq 142]$

$= P[R \leq -34]$

$+ P[R \geq 108 + 142]$

$= P[R - 108 \geq 142]$

Theorem 7.4

The sequence $\hat{R}_n$ of estimators is consistent if, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P[|\hat{R}_n - r| > \epsilon] = 0.$$

Mean squared error $e_n = E[(\hat{R}_n - r)^2]$

Given that $\hat{R}_n$ are unbiased estimators

$\Rightarrow E[\hat{R}_n] = r.$

$\therefore e_n = E[(\hat{R}_n - r)^2] = Var[\hat{R}_n].$

Consider
$$P[|\hat{R}_n - r| > \epsilon] = P[|\hat{R}_n - E[\hat{R}_n]| > \epsilon]$$
$$\leq \frac{Var[\hat{R}_n]}{\epsilon^2} = \frac{e_n}{\epsilon^2}$$

Also, given $\lim_{n\to\infty} e_n = 0.$

$$\therefore \lim_{n\to\infty} P[|\hat{R}_n - r| > \epsilon] = 0 \quad \text{// End of Proof.}$$

Example 7.4

$$\hat{R}_k = \frac{N_k}{k} \Rightarrow E[\hat{R}_k] = \frac{E[N_k]}{k} = \frac{kr}{k} = r.$$

$\therefore \hat{R}_k, \forall k$, is an unbiased estimator of parameter r.

$$e_k = E[(\hat{R}_k - r)^2]$$
$$= E\left[\left(\frac{N_k}{k} - r\right)^2\right] = \frac{E[N_k^2]}{k^2} + r^2 - \frac{2r E[N_k]}{k}$$
$$= \frac{E[N_k^2]}{k^2} - r^2$$
$$= \frac{(kr + k^2r^2)}{k^2} - r^2 = \frac{kr}{k^2} = \frac{r}{k}$$

$\lim_{k\to\infty} e_k = 0. \quad \therefore$ From Thm 7.4, the sequence of estimators $\hat{R}_k$ is consistent.

We could also simply calculate the bound on $P[|\hat{R}_k - r| > \epsilon]$, that is

$$P[|\hat{R}_k - r| > \epsilon] = P[|\hat{R}_k - E[\hat{R}_k]| > \epsilon]$$
$$\leq \frac{Var[\hat{R}_k]}{\epsilon^2} = \frac{r}{\epsilon^2 k}.$$
$$\therefore \lim_{k\to\infty} P[|\hat{R}_k - r| > \epsilon] = 0 \text{ //}$$

Example 7.5

$\hat{P}_n(A)$ is the relative freq estimate of $P[A]$

Standard error = ?

$$e_n = E[(\hat{P}_n(A) - P[A])^2]$$
$$\hat{P}_n(A) = \frac{X_1 + \dots + X_n}{n}$$
$$E[X_i] = P[A]$$

$\therefore E[\hat{P}_n(A)] = P[A]$, that is $\hat{P}_n(A)$ is an unbiased estimator of $P[A]$.

$\therefore e_n = Var[\hat{P}_n(A)]$

For satisfying

Std error < 0.1,

$e_n < 0.01.$

$$Var[\hat{P}_n(A)] = E[(\hat{P}_n(A))^2] - (P[A])^2 \quad \because \text{Unbiased.}$$
$$= \frac{1}{n} Var[X_i]$$

$$\boxed{\begin{aligned} X_i &= 1 \quad \text{w.p } P[A] \\ &= 0 \quad \text{otherwise.} \end{aligned}}$$

$$E[X_i^2] = P[A] \quad E[X_i] = P[A]$$
$$Var[X_i] = P[A] - (P[A])^2$$
$$= P[A](1 - P[A]).$$
$$\therefore Var[\hat{P}_n(A)] = \frac{1}{n} P[A](1 - P[A])$$

For $Var[\hat{P}_n(A)] < 0.01$,

we need
$$n > \left\lceil \frac{P[A](1-P[A])}{0.01} \right\rceil$$

Note that since $\hat{P}_n(A)$ is unbiased

$$P[|\hat{P}_n(A) - P[A]| > \epsilon] \leq \frac{Var[\hat{P}_n(A)]}{\epsilon^2}$$

So for any given ϵ picking a smaller $Var[\hat{P}_n(A)] \Rightarrow$ Value drawn from $\hat{P}_n(A)$ is more likely to be within $\pm\epsilon$ of $P[A]$!

Obs: $Var[\hat{P}_n(A)] = \frac{1}{n} Var[X_i]$

If $Var[X_i]$ is finite, in the limit as $n \to \infty$,

$Var[\hat{P}_n(A)] \to 0.$

The sequence $\hat{P}_n(A)$ is a consistent estimator of $P[A]$.

Theorem 7.7

Sample mean $M_n(X) = \frac{X_1 + \dots + X_n}{n}$

Note, it is an unbiased estimator of $E[X]$.

From Chebyshev: (for any $\epsilon > 0$)

$$P[|M_n(X) - E[M_n(X)]| > \epsilon] \leq \frac{Var[M_n(X)]}{\epsilon^2}$$
$$\therefore P[|M_n(X) - E[X]| > \epsilon] \leq \frac{Var[M_n(X)]}{\epsilon^2}$$
$$Var[M_n(X)] = \frac{1}{n} Var[X].$$

$\therefore$ As $n \to \infty$

$$P[|M_n(X) - E[X]| > \epsilon] = 0 \quad \left[\text{As long as } Var[X] \text{ is finite}\right]$$

Hence Proved //.

Problem 7.3.1

Given X is Gaussian.

$M_n(X)$ is the sample mean.

$$P[|M_n(X) - \mu_X| \geq \sqrt{e_n}] = ?$$

Since $M_n(X)$ is an unbiased estimator the mean squared error $e_n = Var[M_n(X)]$

$\therefore$ We want to calculate

$$P[|M_n(X) - \mu_X| \geq \sigma_{M_n(X)}]$$
$$= P\left[\frac{|M_n(X) - \mu_X|}{\sigma_{M_n(X)}} \geq 1\right]$$
$$= 1 - P[-1 \leq Z \leq 1]$$
$$= 1 - 0.68 \text{ //.}$$

The question asks for $P[|M_n(X) - \mu_X| \leq \sqrt{e_n}]$

$$\therefore \text{We get } P[|M_n(X) - \mu_X| \leq \sqrt{e_n}] = \underline{\underline{0.68}}.$$

If n is large enough for the CLT approx, then $M_n(X)$ can be approximated by a Gaussian with mean μ_X and variance e_n. Therefore, the above conclusion may hold as long as CLT can be used to approximate. Note $e_n = \frac{Var[X]}{n}$

Point Estimates of the Variance

Two cases:

(a) $\mu_X = E[X]$ is known.

(b) μ_X is not known.

If known:

We want to estimate $\mathrm{Var}[X]$

$$= E\left[(X-\mu_X)^2\right]$$

Let $Y = (X-\mu_X)^2$

$\therefore E[Y] = \mathrm{Var}[X].$

n sub-experiments $X_1, X_2, \ldots, X_n$.

The Sample mean $M_n(Y)$ of Y is therefore an unbiased and consistent estimator of $\mathrm{Var}[X]$.

If Unknown:

The mean μ_X also needs to be estimated.

Define sample variance:

$$V_n(X) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - M_n(X)\right)^2$$

What kind of an estimator is sample variance?

1) Is it unbiased?

$$E[V_n(X)] = \frac{1}{n}\sum_{i=1}^{n} E\left[(X_i - M_n(X))^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left[\left(X_i - \frac{(X_1+X_2+\cdots+X_n)}{n}\right)^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\left[X_i^2 + \frac{1}{n^2}(X_1+\cdots+X_n)^2 - \frac{2}{n}X_i(X_1+\cdots+X_n)\right]$$

$$= \frac{1}{n^2}E\left[(X_1+\cdots+X_n)^2\right] + \frac{1}{n}E\left[X_1^2+\cdots+X_n^2\right] - \frac{2}{n^2}E\left[(X_1+\cdots+X_n)^2\right]$$

$$= \frac{1}{n}E\left[X_1^2+\cdots+X_n^2\right] - \frac{1}{n^2}E\left[(X_1+X_2+\cdots+X_n)^2\right]$$

$$= \frac{1}{n}nE[X^2] - \frac{1}{n^2}E\left[X_1^2+\cdots+X_n^2 + 2X_1X_2 + \cdots + 2X_{n-1}X_n\right]$$

$$= E[X^2] - \frac{1}{n^2}\left[nE[X^2] - n(n-1)E[X_1X_2]\right]$$

$$= E[X^2] - \frac{1}{n}E[X^2] - \frac{n(n-1)}{n^2}E^2[X]$$

$$= \left(\frac{n-1}{n}\right)E[X^2] - \left(\frac{n-1}{n}\right)E^2[X]$$

$$= \left(\frac{n-1}{n}\right)\mathrm{Var}[X].$$

$\therefore$ The Sample Var is biased

This leads us to an unbiased estimate given by $\left(\frac{n}{n-1}\right)V_n(X)$.

Problem 7.33

The parameter we want to estimate is $r = E[X_1X_2]$.

The estimate used is

$$\hat{R}_n = \frac{1}{n}\sum_{i=1}^{n} X_1(i)X_2(i).$$

Given $\mathrm{Var}[X_1X_2]$ is finite.

To show: $\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_n, \ldots$ is an unbiased consistent sequence of estimates of r.

Proof:

$$E[\hat{R}_n] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_1(i)X_2(i)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_1(i)X_2(i)] = \frac{1}{n}(nr) = r.$$

(Unbiased)

From Chebyshev:

$$P\left[\left|\hat{R}_n - E[\hat{R}_n]\right| > \epsilon\right] \le \frac{\mathrm{Var}[\hat{R}_n]}{\epsilon^2}$$

To show consistency, we need to show that

$$\lim_{n\to\infty} P\left[|\hat{R}_n - r| > \epsilon\right] = 0$$

same as (because $\hat{R}_n$ is an unbiased est)

$$\lim_{n\to\infty} P\left[|\hat{R}_n - E[\hat{R}_n]| > \epsilon\right] = 0$$

If we set $Y_i = X_1(i)X_2(i)$, and note that Y_i and Y_j are independent and identically distributed, $\hat{R}_n$ is the sample mean of Y_i.

$$\therefore \mathrm{Var}[\hat{R}_n] = \mathrm{Var}\left[\frac{Y_1+\cdots+Y_n}{n}\right]$$

$$= \frac{n}{n^2}\mathrm{Var}[Y]$$

$$= \frac{n}{n^2}\mathrm{Var}[X_1X_2]$$

$$= \boxed{\frac{\mathrm{Var}[X_1X_2]}{n}}.$$

~~If $\mathrm{Var}[X_1X_2] \neq 0$,~~

$$\lim_{n\to\infty}\mathrm{Var}[\hat{R}_n] = 0$$

$$\Rightarrow \lim_{n\to\infty} P\left[|\hat{R}_n - E[X_1X_2]| > \epsilon\right] = 0$$

for any $\epsilon > 0$.

$\therefore \hat{R}_1, \hat{R}_2, \ldots, \hat{R}_n$ are consistent.

Problem 7.3.4

(a) Straight forward.

(b) To show

$$\lim_{n\to\infty} P\left[\max_{j=1,\ldots,k}|M_j(n) - \mu_j| \ge c\right] = 0$$

$$P\left[\text{At least one of } |M_j(n)-\mu_j| \ge c,\ 1\le j\le k\right]$$

$$= P\left[|M_1(n)-\mu_1| \ge c \cup \cdots \cup |M_k(n)-\mu_k| \ge c\right]$$

$$\le P\left[|M_1(n)-\mu_1| \ge c\right] + \cdots + P\left[|M_k(n)-\mu_k| \ge c\right]$$

$$= P\left[|M_1(n)-E[M_1(n)]| \ge c\right] + \cdots + P\left[|M_k(n)-E[M_k(n)]| \ge c\right]$$

($\because M_j(n)$ is an unbiased estimate of μ_j)

Taking the limit $\lim_{n\to\infty}$ we get the desired result.

Problem 7.37

$$\mathrm{Var}[\hat{R}_{ij}(n)] = \mathrm{Var}\left[\frac{1}{n}\sum_{m=1}^{n} X_i(m)X_j(m)\right]$$

$$= \frac{1}{n^2}n\,\mathrm{Var}[X_iX_j]$$

$\because X_i(m)X_j(m)$ is independent of $X_i(k)X_j(k)$, $k \ne m$.

$$\mathrm{Var}[X_iX_j] = E[X_i^2X_j^2] - E^2[X_iX_j]$$

$$E[X_i^2X_j^2] = E\left[E[X_i^2X_j^2 \mid X_j]\right]$$

$$= E\left[X_j^2\,\underbrace{E[X_i^2 \mid X_j]}_{\text{Conditional pdf of Gaussian}}\right]$$

(Didn't teach iterated Exp in class)

$$E[X_iX_j] = \mathrm{Cov}[X_i, X_j] + E[X_i]E[X_j]$$

$$= \rho_{X_i,X_j}\sigma_{X_i}\sigma_{X_j} + \mu_{X_i}\mu_{X_j}$$

And so on... to show that $\mathrm{Var}[X_iX_j]$ is finite.

Probability and Random Processes

Sanjit Kaul

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Parameter Estimation

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Sampling From a Finite Population

$\underline{n}$ $\underbrace{k=2}$

$X_1, X_2, \ldots, X_n$

$X_i \sim X$

$X^{(1)}$

$X^{(2)}$

Are $X^{(1)}$ and $X^{(2)}$ independent.

Sampling w/o replacement
with

Definition 7.1 Sample Mean

For iid random variables $X_1, \ldots, X_n$ with PDF $f_X(x)$, the sample mean *of X is the random variable*

$$M_n(X) = \frac{X_1 + \cdots + X_n}{n}.$$

Theorem 7.1

The sample mean $M_n(X)$ has expected value and variance

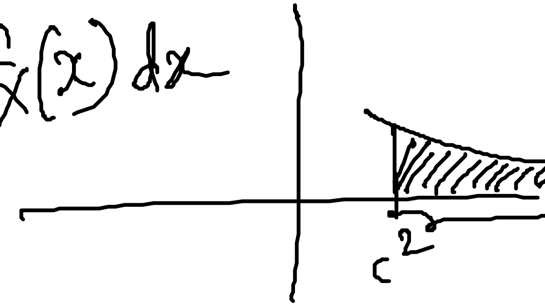
$$E\left[M_n(X)\right] = E\left[X\right], \qquad \mathrm{Var}[M_n(X)] = \frac{\mathrm{Var}[X]}{n}.$$

Quiz 7.1

Let X be an exponential random variable with expected value 1. Let $M_n(X)$ denote the sample mean of n independent samples of X. How many samples n are needed to guarantee that the variance of the sample mean $M_n(X)$ is no more than 0.01?

$$E[X] = \int_0^\infty x f_X(x)\,dx = \int_0^{c^2} x f_X(x)\,dx + \int_{c^2}^\infty x f_X(x)\,dx$$

$$\geq \int_{c^2}^\infty x f_X(x)\,dx \geq \int_{c^2}^\infty c^2 f_X(x)\,dx$$



Theorem 7.2 Markov Inequality

For a random variable X such that $P[X < 0] = 0$ and a constant c,

$$\int_{c^2}^\infty f_X(x)\,dx \qquad P\left[X \geq c^2\right] \leq \frac{E[X]}{c^2}. \qquad Z = (Y - E[Y])^2$$

$$P\left[|Y - E[Y]| \leq c\right] \geq \frac{Var[Y]}{c^2} \qquad P\left[Z \geq c^2\right] \leq \frac{Var[Y]}{c^2}$$

$$P\left[(Y - E[Y])^2 \geq c^2\right] \leq \frac{Var[Y]}{c^2}$$

Theorem 7.3 Chebyshev Inequality

For an arbitrary random variable Y and constant $c > 0$,

$$P\left[|Y - \mu_Y| \geq c\right] \leq \frac{\mathrm{Var}[Y]}{c^2}.$$

$$|Y - \mu_Y| \geq c \iff (Y - \mu_Y)^2 \geq c^2$$

$$-(Y - \mu_Y) \geq c$$

Quiz 7.2

Elevators arrive randomly at the ground floor of an office building. Because of a large crowd, a person will wait for time W in order to board the third arriving elevator. Let X_1 denote the time (in seconds) until the first elevator arrives and let X_i denote the time between the arrival of elevator $i-1$ and i. Suppose X_1, X_2, X_3 are independent uniform $(0, 30)$ random variables. Find upper bounds to the probability W exceeds 75 seconds using

$E[W] = 45$ 0 75 $P[W > 75] \le \frac{E[W]}{75}$

(1) the Markov inequality,

$|W-45| > 30$ $-(W-45) > 30$ $P[W - E[W] > 75 - E[W]]$

(2) the Chebyshev inequality.

$W-45 < -30$

$W < -75$

$\boxed{P[|W-\mu_W| \ge c^2] \le \frac{\mathrm{Var}[W]}{c^2}}$

$P[|W-45| > 30]$

$W-45 > 30$

$p = \frac{1}{36}$ $P[R=r] = {}_{r-1}C_2\, p^2 (1-p)^{r-3} p$ $P[R \geq r] = \sum$

Problem 7.2.4

$R = X_1 + X_2 + X_3$ $E[R] = 3E[X] = \frac{3}{p}$

In a game with two dice, the event *snake eyes* refers to both dice showing one spot. Let R denote the number of dice rolls needed to observe the third occurrence of *snake eyes*. Find

$P[R \geq (\sqrt{250})^2] \leq \frac{3/p}{250}$

(a) the upper bound to $P[R \geq 250]$ based on the Markov inequality,

(b) the upper bound to $P[R \geq 250]$ based on the Chebyshev inequality,

(c) the exact value of $P[R \geq 250]$.

$P[|R - \mu_R| \geq c] \leq \frac{Var[R]}{c^2}$

$P[|R - 3(36)| \geq c] \leq \frac{Var[R]}{c^2}$

$P[R \geq 250]$

$= P[R - 108 \geq 142]$

$\{|R - 108| \geq 142\} = \{-(R-108) \geq 142\} \cup \{R - 108 \geq 142\}$

Section 7.3

Point Estimates of Model Parameters

$\hat{R}_1 = \frac{X_1}{1}$ $\quad \hat{R}_2 = \frac{X_1 + X_2}{2}$ $\quad \hat{R}_3 = \frac{X_1 + X_2 + X_3}{3}$

if estimator is sample mean

$r = E[X]$

Definition 7.2 Consistent Estimator

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ *of the parameter* r *is* consistent *if for any* $\epsilon > 0$,

$$\lim_{n\to\infty} P\left[\left|\hat{R}_n - r\right| \geq \epsilon\right] = 0.$$

$X \leftarrow \pi$

$$E\left[\frac{X_1+X_2+X_3}{3}\right] = E[X]$$

Definition 7.3 Unbiased Estimator

An estimate, $\hat{R}$, of parameter r is unbiased *if $E[\hat{R}] = r$; otherwise, $\hat{R}$ is* biased.

Definition 7.4 Estimator

The sequence of estimators $\hat{R}_n$ *of parameter* r *is* asymptotically unbiased *if*

$$\lim_{n \to \infty} E[\hat{R}_n] = r.$$

Definition 7.5 Mean Square Error

The mean square error *of estimator* $\hat{R}$ *of parameter* r *is*

$$e = E\left[(\hat{R} - r)^2\right].$$

$$e_n = \underbrace{E\left[(\hat{R}_n - r)^2\right]}$$

For unbiased $\hat{R}_n$

$$e_n = E\left[(\hat{R}_n - E[\hat{R}_n])^2\right] = \mathrm{Var}[\hat{R}_n]$$

Theorem 7.4

Fact given that estimates are unbiased

If a sequence of unbiased estimates $\hat{R}_1, \hat{R}_2, \ldots$ of parameter r has mean square error $e_n = \mathrm{Var}[\hat{R}_n]$ satisfying $\lim_{n\to\infty} e_n = 0$, then the sequence $\hat{R}_n$ is consistent.

Given

$$\lim_{n\to\infty} P\left[|\hat{R}_n - r| \geq \epsilon\right] = 0$$

To prove

$$P\left[|\hat{R}_n - r| \geq \epsilon\right] = P\left[|\hat{R}_n - \underbrace{E[\hat{R}_n]}| \geq \epsilon\right] \leq \frac{\mathrm{Var}(\hat{R}_n)}{\epsilon^2}$$

$$E[\hat{R}_k] = E\left[\frac{N_k}{k}\right] = \frac{1}{k}E[N_k] = r$$

Example 7.4 Problem

In any interval of k seconds, the number N_k of packets passing through an Internet router is a Poisson random variable with expected value $E[N_k] = kr$ packets. Let $\hat{R}_k = N_k/k$ denote an estimate of r. Is each estimate $\hat{R}_k$ an unbiased estimate of r? What is the mean square error e_k of the estimate $\hat{R}_k$? Is the sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ consistent?

$$e_k = E\left[(\hat{R}_k - r)^2\right] = E\left[(\hat{R}_k - E[\hat{R}_k])^2\right] = \mathrm{Var}[\hat{R}_k]$$

$$= \frac{\mathrm{Var}[N_k]}{k^2}$$

Theorem 7.5

The sample mean $M_n(X)$ is an unbiased estimate of $E[X]$.

Theorem 7.6

The sample mean estimator $M_n(X)$ has mean square error

$$e_n = E\left[(M_n(X) - E[X])^2\right] = \text{Var}[M_n(X)] = \frac{\text{Var}[X]}{n}.$$

This is also the variance of the sample mean estimator. Square-root of this is called the standard error.
Given the mean and variance of the estimator, what can we say about the sample mean that results from n IID trials?

$$\text{Bern}(P[A]) \sim X \qquad \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Example 7.5 Problem

How many independent trials n are needed to guarantee that $\hat{P}_n(A)$, the relative frequency estimate of $P[A]$, has standard error less than 0.1?

$$E\left[\left(\hat{P}_n(A) - P[A]\right)^2\right] = \text{Var}\left[\hat{P}_n(A)\right] \leq 0.01$$

$$\text{Var}[X] = P[A](1 - P[A])$$

$$\frac{\text{Var}[X]}{n} \leq 0.01$$

$$n \geq \frac{\text{Var}[X]}{0.01} = \frac{P[A](1-P[A])}{0.01}$$

$\hat{R}_1 \quad \hat{R}_2 \quad \cdots \quad \hat{R}_n$

$\hat{R}_1 = X_1$

$\hat{R}_2 = \frac{X_1 + X_2}{2} \quad \cdots$

Theorem 7.7

If X has finite variance, then the sample mean $M_n(X)$ is a sequence of consistent estimates of $E[X]$.

$\lim_{n \to \infty} \mathrm{Var}[\hat{R}_n] = 0$

Theorem 7.8 Weak Law of Large Numbers

If X has finite variance, then for any constant $c > 0$,

(a) $\lim_{n\to\infty} P[|M_n(X) - \mu_X| \geq c] = 0$,

(b) $\lim_{n\to\infty} P[|M_n(X) - \mu_X| < c] = 1$.

Theorem 7.9

As $n \to \infty$, the relative frequency $\hat{P}_n(A)$ converges to $P[A]$; for any constant $c > 0$,

$$\lim_{n\to\infty} P\left[\left|\hat{P}_n(A) - P\left[A\right]\right| \geq c\right] = 0.$$

Definition 7.6 ***Convergence in Probability***

The random sequence Y_n *converges in probability to a constant* y *if for any* $\epsilon > 0$,

$$\lim_{n\to\infty} P\left[|Y_n - y| \geq \epsilon\right] = 0.$$

$$E\left[\left(M_n(X) - E(X)\right)^2\right]$$

$$= Var\left(M_n(X)\right)$$

Problem 7.3.1

When X is Gaussian, verify the claim of Equation (7.16) that the sample mean is within one standard error of the expected value with probability 0.68.

Estimating the Variance

- Mean is known
- Mean Unknown

$X \quad \mu_X$ ✓

$$Var[X]$$
$$= E\left[(X-\mu_X)^2\right]$$
$$= E[Y]$$
$$Y = (X-\mu_X)^2$$

μ

Definition 7.7 Sample Variance

The sample variance of a set of n independent observations of random variable X is

$$V_n(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n(X))^2 .$$

Is

$E[V_n(X)] = Var[X]$?

$E[(X_i - M_n(X))^2]$

$= E[X$

$E[V_n(X)]$

$= \frac{n-1}{n} Var[X]$

Theorem 7.10

$$E\left[V_n(X)\right] = \frac{n-1}{n}\,\mathrm{Var}[X].$$

Theorem 7.11

The estimate

$$V'_n(X) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M_n(X))^2$$

is an unbiased estimate of $\mathrm{Var}[X]$.

$$\mathrm{Var}[\hat{R}_n] = \frac{\mathrm{Var}[Y]}{n} = \frac{\mathrm{Var}[X_1 X_2]}{n}$$

$$Y_i = X_1(i) X_2(i)$$

Problem 7.3.3

An experimental trial produces random variables X_1 and X_2 with correlation $r = E[X_1 X_2]$. To estimate r, we perform n independent trials and form the estimate

$$\hat{R}_n = \frac{1}{n} \sum_{i=1}^{n} X_1(i) X_2(i) = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

where $X_1(i)$ and $X_2(i)$ are samples of X_1 and X_2 on trial i. Show that if $\mathrm{Var}[X_1 X_2]$ is finite, then $\hat{R}_1, \hat{R}_2, \ldots$ is an unbiased, consistent sequence of estimates of r.

$$Y \sim Y_i \sim X_1(i)\, X_2(i)$$

$$X_1(i) \xrightarrow{d} X_1, \quad X_2(i) \xrightarrow{d} X_2$$

Problem 7.3.4

An experiment produces random vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_k \end{bmatrix}'$ with expected value $\boldsymbol{\mu}_{\mathbf{X}} = \begin{bmatrix} \mu_1 & \cdots & \mu_k \end{bmatrix}'$. The ith component of $\mathbf{X}$ has variance $\mathrm{Var}[X_i] = \sigma_i^2$. To estimate $\boldsymbol{\mu}_{\mathbf{X}}$, we perform n independent trials such that $\mathbf{X}(i)$ is the sample of $\mathbf{X}$ on trial i, and we form the vector mean

$$\mathbf{M}(n) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}(i).$$

(a) Show $\mathbf{M}(n)$ is unbiased by showing $E[\mathbf{M}(n)] = \boldsymbol{\mu}_{\mathbf{X}}$.

(b) Show that the sequence of estimates $\mathbf{M}_n$ is consistent by showing that for any constant $c > 0$,

$$\lim_{n\to\infty} P\left[\max_{j=1,\ldots,k} \left|M_j(n) - \mu_j\right| \geq c\right] = 0.$$

Hint: Let $A_i = \{|M_i(n) - \mu_i| \geq c\}$ and apply the union bound (see Problem 1.4.5) to upper bound $P[A_1 \cup A_2 \cup \cdots \cup A_k]$. Then apply the Chebyshev inequality.

Problem 7.3.7

An experiment produces a zero mean Gaussian random vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_k \end{bmatrix}'$ with correlation matrix $\mathbf{R} = E[\mathbf{X}\mathbf{X}']$. To estimate $\mathbf{R}$, we perform n independent trials, yielding the iid sample vectors $\mathbf{X}(1), \mathbf{X}(2), \ldots, \mathbf{X}(n)$, and form the sample correlation matrix

$$\hat{\mathbf{R}}(n) = \frac{1}{n} \sum_{m=1}^{n} \mathbf{X}(m)\mathbf{X}'(m).$$

(a) Show $\hat{\mathbf{R}}(n)$ is unbiased by showing $E[\hat{\mathbf{R}}(n)] = \mathbf{R}$.

(b) Show that the sequence of estimates $\hat{\mathbf{R}}(n)$ is consistent by showing that every element $\hat{R}_{ij}(n)$ of the matrix $\hat{\mathbf{R}}$ converges to R_{ij}. That is, show that for any $c > 0$,

$$\lim_{n\to\infty} P\left[\max_{i,j} \left|\hat{R}_{ij} - R_{ij}\right| \geq c\right] = 0.$$

Hint: Extend the technique used in Problem 7.3.4. You will need to use the result of Problem 4.11.8 to show that $\mathrm{Var}[X_i X_j]$ is finite.

Section 7.4

μ $(\mu - c, \mu + c)$ Estimator $\triangleq f(X_1, X_2, \ldots, X_n)$

p $M_n(X)$ $X_i \sim X$

Confidence Intervals

Suppose $M_n(X) = 10$ P

$$P\left[|M_n(X) - \mu| \le c\right] = P\left[M_n(X) \in (\mu - c, \mu + c)\right] \ge 1 - \alpha$$

$2c$

Theorem 7.12

For any constant $c > 0$,

(a) $P[|M_n(X) - \mu_X| \geq c] \leq \dfrac{\text{Var}[X]}{nc^2} = \alpha,$

(b) $P[|M_n(X) - \mu_X| < c] \geq 1 - \dfrac{\text{Var}[X]}{nc^2} = 1 - \alpha.$

Confidence *Coefficient* of your estimate being in an interval of length 2c

Example 7.6 Problem

Suppose we perform n independent trials of an experiment and we use the relative frequency $\hat{P}_n(A)$ to estimate $P[A]$. Use the Chebyshev inequality to calculate the smallest n such that $\hat{P}_n(A)$ is in a confidence interval of length 0.02 with confidence 0.999.

Example 7.7 Problem

$X_1 \; X_2 \cdots X_n \quad X_i \sim X$

Suppose we perform n independent trials of an experiment. For an event A of the experiment, use the Chebyshev inequality to calculate the number of trials needed to guarantee that the probability the relative frequency of A differs from $P[A]$ by more than 10% is less than 0.001.

$r = P[A]$ $\quad 2c \quad c = 0.1\,P[A]$

$P[M_n(X) \in (r-c, r+c)] \geq 1 - 0.001$

$0.1r$

$\rightarrow P[|M_n(X) - r| \leq 0.1r] \geq 1 - 0.001$

$P[|M_n(X) - E[M_n(X)]| \leq 0.1E[M_n(X)]] \geq 1 - \frac{\mathrm{Var}[X]}{n(0.1r)^2}$

$n > \frac{10^5}{P[A]}$

Note $E[M_n(X)] = r$; $\quad P[A](1-P[A])$

Example 7.8 Problem

Based on a sample of 1103 potential voters, the percentage of people supporting Candidate Jones is 58% with an accuracy of plus or minus 3 percentage points.

$0.58 = \hat{P}_n(A)$ $n = 1103$

$c = 0.03$

With what minimum probability does your method for estimating support for Jones satisfy the accuracy requirement?

$P\left[\,|\hat{P}_n(A) - P(A)| \geq 0.03\right]$

$\leq \dfrac{Var[X]}{(1103)(0.03)^2}$

0.25

In other words, what is the confidence coefficient?

Interval Estimates

- The interval is random
- A probability is associated with the parameter lying within the random interval
  - You estimate a random interval
  - With some probability your parameter of interest lies within your estimated interval

$P[A \le \pi \le B] \ge 1-\alpha$ $\qquad \pi = E[X]$

(a, b)

$(\hat{\pi} - c, \hat{\pi} + c)$

$P[M_n(X) - c \le \pi \le M_n(X) + c]$

$= P[-c \le M_n(X) - \pi \le c]$

$= P[-c \le M_n(X) - E[X] \le c]$

$\ge 1 - \frac{Var(X)}{nc^2}$

Example 7.9 Problem

Suppose X_i is the ith independent measurement of the length (in cm) of a board whose actual length is b cm. Each measurement X_i has the form

$$X_i = b + Z_i, \qquad E(X_i) = b \tag{7.43}$$

where the measurement error Z_i is a random variable with expected value zero and standard deviation $\sigma_Z = 1$ cm. Since each measurement is fairly inaccurate, we would like to use $M_n(X)$ to get an accurate confidence interval estimate of the exact board length. How many measurements are needed for a confidence interval estimate of b of length $2c = 0.2$ cm to have confidence coefficient $1 - \alpha = 0.99$?

$$P[M_n(X) - c \le E[X] \le M_n(X) + c] \ge 1 - \frac{\operatorname{Var}(X)}{nc^2} = 1 - \frac{1}{n(0.1)^2}$$

$$n \ge 10^4$$

$$P[-c \le M_n(X) - E[X] \le c]$$

$$= P[-c \le M_n(X) - E[M_n(X)] \le c]$$

$$= P\left[\frac{-c}{\sigma_{M_n(X)}} \le \frac{M_n(X) - E[M_n(X)]}{\sigma_{M_n(X)}} \le \frac{c}{\sigma_{M_n(X)}}\right] = 1 - 2Q\left(\frac{c\sqrt{n}}{\sigma_X}\right) \ge 0.99$$

$$Q(0.1\sqrt{n}) \le 0.005 \Rightarrow n \ge 666$$

Example 7.10 Problem

In Example 7.9, suppose we know that the measurement errors Z_i are iid Gaussian random variables. How many measurements are needed to guarantee that our confidence interval estimate of length $2c = 0.2$ has confidence coefficient $1 - \alpha \geq 0.99$?

Theorem 7.13

$\sigma_{M_n(X)} = \sigma/\sqrt{n}$

$\mu = E[M_n(X)]$

Let X be a Gaussian (μ, σ) random variable. A confidence interval estimate of μ of the form

$$P[M_n(X) - c \le \mu \le M_n(X) + c] =$$

$P[-c \le \mu - M_n(X) \le c]$

has confidence coefficient $1 - \alpha$ where

$$\alpha/2 = Q\left(\frac{c\sqrt{n}}{\sigma}\right) = 1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right).$$

$= P[c \ge M_n(X) - \mu \ge -c]$

Z

$= P\left[\frac{-c}{(\sigma/\sqrt{n})} \le \boxed{\frac{M_n(X) - \mu}{(\sigma/\sqrt{n})}} \le \frac{c}{(\sigma/\sqrt{n})}\right] = P\left[\frac{-c}{\sigma/\sqrt{n}} \le Z \le \frac{c}{\sigma/\sqrt{n}}\right]$

$= 1 - \alpha$

$P\left[Z \le \frac{c}{\sigma/\sqrt{n}}\right] - P\left[Z \le \frac{-c}{\sigma/\sqrt{n}}\right]$

$\hookrightarrow 1 - P\left[Z \le -\frac{c}{\sigma/\sqrt{n}}\right] - P\left[Z \ge \frac{c}{\sigma/\sqrt{n}}\right]$

$= 1 - 2P\left[Z \ge \frac{c}{\sigma/\sqrt{n}}\right]$

Example 7.11 Problem

$E(Y)$: trend of interest.

Y is a Gaussian random variable with unknown expected value μ but known variance σ_Y^2. Use $M_n(Y)$ to find a confidence interval estimate of μ_Y with confidence 0.99. If $\sigma_Y^2 = 10$ and $M_{100}(Y) = 33.2$, what is our interval estimate of μ formed from 100 independent samples?

$n = 100$ ✓

$1 - \alpha = 0.99$

$c = 10$

$(-10, 10)$

$(-10\sqrt{10}, 10\sqrt{10})$

$(-10 + 33.2, 10 + 33.2)$

$1 - \alpha$

We want to compare the mean family income in two states. For state 1, we had a random sample of $n_1 = 100$ families with a sample mean of $\bar{x}_1 = 35000$. For state 2, we had a random sample of $n_2 = 144$ families with a sample mean of $\bar{x}_2 = 36000$. Past studies have shown that for both states $\sigma = 4000$. Estimate $\mu_1 - \mu_2$ and place a two-standard-error bound on the error of estimation. How much confidence do we have in this interval?

X_1

X_2

$c = 2e$

http://old.stat.duke.edu/courses/Spring02/sta103/lec/ch8b_4.pdf

$M_n(X_1) - M_n(X_2)$

$P\left[M_n(X_1) - M_n(X_2) - c \le \mu_1 - \mu_2 \le M_n(X_1) - M_n(X_2) + c\right]$

$e^2 = E\left[\left(M_n(X_1) - M_n(X_2) - (\mu_1 - \mu_2)\right)^2\right] = \mathrm{Var}\left[M_n(X_1) - M_n(X_2)\right]$

$$\mathrm{Var}[M_n(X_1) - M_n(X_2)] = \mathrm{Var}[M_n(X_1)] + \mathrm{Var}[M_n(X_2)]$$

$$\geq 1 - \frac{\mathrm{Var}(X_1 - X_2)}{n 4 e^2}$$

$$(-1000 - c, 1000 + c)$$

$$2e$$

EXAMPLE 7.3a Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance 4. That is, if μ is sent, then the value received is $\mu + N$ where N, representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, let us construct a 95 percent confidence interval for μ.

$$M_n(X) = \hat{\mu} \qquad \frac{1}{8}\left((5-\hat{\mu})^2 + (8.5-\hat{\mu})^2 + \cdots\right)$$

α $\qquad$ α

$$P\left[\,|M_n(X) - \mu| \le c\,\right] \ge 1 - \alpha$$

$$P\left[\frac{|M_n(X) - \mu|}{\sigma_{M_n(X)}} \le \frac{c}{\sigma_{M_n(X)}}\right]$$

$\sigma_{M_n(X)} \to \sigma/\sqrt{n}$ $\qquad$ $\sigma_{M_n(X)} \to \sigma/\sqrt{n}$

S^2

Unbiased estimate of σ^2

$$P\left[\frac{|M_n(X)-\mu|}{(S/\sqrt{n})} \leq \frac{c}{(S/\sqrt{n})}\right]$$

$\rightarrow$ Student-t distribution $\rightarrow$ Symmetric about the mean

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - M_n(X)\right)^2$$

$$P\left[\frac{-c}{S/\sqrt{n}} \leq \frac{M_n(X)-\mu}{S/\sqrt{n}} \leq \frac{c}{S/\sqrt{n}}\right] \geq \underbrace{1-\alpha}_{0.95}$$

$$P\left[-t_{\alpha/2,(n-1)} \leq \frac{M_n(X)-\mu}{S/\sqrt{n}} \leq t_{\alpha/2,(n-1)}\right]$$

$$\left(M_n(X) - \frac{S}{\sqrt{n}}\,t_{\alpha/2,(n-1)},\ M_n(X) + \frac{S}{\sqrt{n}}\,t_{\alpha/2,(n-1)}\right) = 1-\alpha$$

EXAMPLE 7.3e Let us again consider Example 7.3a but let us now suppose that when the value μ is transmitted at location A then the value received at location B is normal with mean μ and variance σ^2 but with σ^2 being unknown. If 9 successive values are, as in Example 7.3a, 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, compute a 95 percent confidence interval for μ.

$t_{0.025,8} = 2.306$

$1-\alpha = 0.95$

$\alpha = 0.05$

$\alpha/2 = 0.025$

Hypothesis Testing

Sanjit Kaul
(and Slides from RY)

Hypothesis Testing

- Perform Experiment
- Observe Outcome
- Make a conclusion/inference with high accuracy

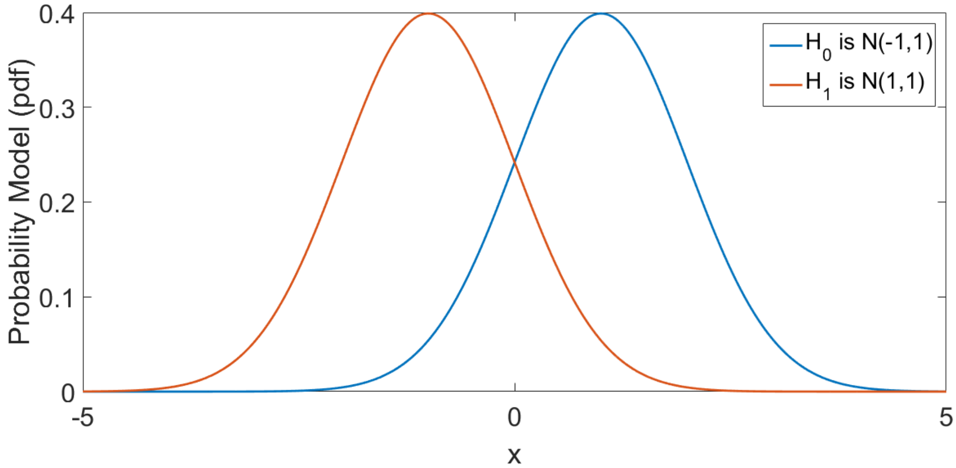
Two Categories of Statistical Inference

- Goal is to assign a probability model to the outcome of your experiment as accurately as possible
- Start with one or more possible models (one or more hypotheses) a priori
- Significance Testing
  - Start with a probability model as your hypothesis
    - Often named H_0 and called the null hypothesis
  - Conclusion: Accept or reject the hypothesis that outcome is from the probability model
  - Accuracy Measure: Probability of rejecting a true hypothesis

Two Categories of Statistical Inference

- Hypothesis Testing
  - Start with multiple probability models.
    - They are your hypotheses H_0, H_1, H_2, …, H_{M-1}
  - Conclusion: Observation from the experiment is from one of the starting Hypotheses
  - Accuracy Measure: Probability you conclude that the observation is from H_i when it is truly from H_k

The Communications Question



- The receiver makes an observation resulting from a transmitter sending a bit (either +1V or -1V)
- Receiver knows that two probability models N(-1,1) and N(1,1) are possible
- Receiver receives a voltage 0.00001 V.
- The receiver must conclude either N(-1,1) or N(1,1)
  - Equivalent to picking -1 or +1

Significance Testing

- Split sample space into two mutually exclusive and collectively exhaustive sets A (accept) and R (reject).
- If outcome belongs to A, then say H_0 occurred, null hypothesis is accepted
  - Else, the null hypothesis is rejected
- Accuracy is a function of how A (and R) are created

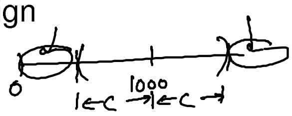
Significance Level

- $P[s \in R | H_0]$
  - Probability that the hypothesis is rejected when the observation is from it
- You may start with this value as a design parameter
- Then decide on sets R and A
- The error is called a Type I error or False Rejection
  - Reject H0 when H0 is true
- Another possible error is Type II or False Acceptance
  - Accept H0 when H0 is false
  - Can't calculate as we don't know the *other* probability model

Example 8.2 Problem

Suppose that on Thursdays between 9:00 and 9:30 at night, the number of call attempts N at a telephone switching office is a Poisson random variable with expected value 1000. Next Thursday, the President will deliver a speech at 9:00 that will be broadcast by all radio and television networks. The null hypothesis, H_0, is that the speech does not affect the probability model of telephone calls. In other words, H_0 states that on the night of the speech, N is a Poisson random variable with expected value 1000. Design a significance test for hypothesis H_0 at a significance level of $\alpha = 0.05$.

$P[S_{ER}|H_0] = 0.05$



$P[S_{ER}|H_0] = P[|N-1000| \geq c] = 0.05$

$P[R|H_0] = 0.01$

Type-I error

$P[M_n(X) < 190 - c | H_0]$

$P[M_n(X) - 190 < -c | H_0]$ $Q\left(\frac{c}{3}\right) = 0.01$

$c > 0$

Example 8.3 Problem

Before releasing a diet pill to the public, a drug company runs a test on a group of 64 people. Before testing the pill, the probability model for the weight of the people measured in pounds, is a Gaussian $(190, 24)$ random variable W. Design a test based on the sample mean of the weight of the population to determine whether the pill has a significant effect. The significance level is $\alpha = 0.01$.

$$P[X \le x \mid H_0] = \frac{P[X \le x, H_0]}{P[H_0]}$$

$$P[X \le x]$$

$0 \quad c$

$$P[X \in (0, c) \mid H_0] = 0.01$$

Quiz 8.1

$$(1 - e^{-c})^{15} = 0.01$$

Under hypothesis H_0, the interarrival times between phone calls are independent and identically distributed exponential (1) random variables. Given X, the maximum among 15 independent interarrival time samples $X_1, \ldots, X_{15}$, design a significance test for hypothesis H_0 at a level of $\alpha = 0.01$.

$c = 1.33$

$$P[X \le x] = P[\mathrm{Max}(X_1, X_2, \ldots, X_{15}) \le x] = P[X_1 \le x, X_2 \le x, \ldots, X_{15} \le x]$$

$$= (1 - e^{-x})^{15}$$

$E[L|H] = 2$

$P(L > c|H) = 0.05$

$(0.5)^c = 0.05$ $c = 4.32$

Problem 8.1.1

Let L equal the number of flips of a coin up to and including the first flip of heads. Devise a significance test for L at level $\alpha = 0.05$ to test the hypothesis H that the coin is fair. What are the limitations of the test?

Problem 8.1.3

When a chip fabrication facility is operating normally, the lifetime of a microchip operated at temperature T, measured in degrees Celsius, is given by an exponential (λ) random variable X with expected value $E[X] = 1/\lambda = (200/T)^2$ years. Occasionally, the chip fabrication plant has contamination problems and the chips tend to fail much more rapidly. To test for contamination problems, each day m chips are subjected to a one-day test at $T = 100°C$. Based on the number N of chips that fail in one day, design a significance test for the null hypothesis test H_0 that the plant is operating normally.

(a) Suppose the rejection set of the test is $R = \{N > 0\}$. Find the significance level of the test as a function of m, the number of chips tested.

(b) How many chips must be tested so that the significance level is $\alpha = 0.01$.

(c) If we raise the temperature of the test, does the number of chips we need to test increase or decrease?

$P[N>0]$
$= 1 - P[N=0]$
$= 1 - (1 - e^{(\ )})^m$

$m = \left(\frac{200}{T}\right)^2 \times 365 \log\left(\frac{1}{1-\alpha}\right)$

$P[\text{A chip passes} \mid H_0]$
$= P\left(X > \frac{1}{365}\right)$
$= e^{-\left(\frac{T}{200}\right)^2 \frac{1}{365}}$

Problem 8.1.5

When a pacemaker factory is operating normally (the null hypothesis H_0), a randomly selected pacemaker fails a test with probability $q_0 = 10^{-4}$. Each day, an inspector randomly tests pacemakers. Design a significance test for the null hypothesis with significance level $\alpha = 0.01$. Note that testing pacemakers is expensive because the pacemakers that are tested must be discarded. Thus the significance test should try to minimize the number of pacemakers tested.

$$\frac{\ln(1-0.01)}{\ln(1-10^{-4})} = M$$

$R = \{N > 0\}$

$P[N>0] = 0.01$

$1 - P[N=0]$

$P[N=0] = (1-10^{-4})^m$

m

Problem 8.1.5

When a pacemaker factory is operating normally (the null hypothesis H_0), a randomly selected pacemaker fails a test with probability $q_0 = 10^{-4}$. Each day, an inspector randomly tests pacemakers. Design a significance test for the null hypothesis with significance level $\alpha = 0.01$. Note that testing pacemakers is expensive because the pacemakers that are tested must be discarded. Thus the significance test should try to minimize the number of pacemakers tested.

$$\frac{\ln(1-0.01)}{\ln(1-10^{-4})} = M$$

$R = \{N > 0\}$

M

$P[N>0] = 0.01$

$1 - P[N=0]$

$P[N=0] = (1-10^{-4})^M$

Binary Hypothesis Testing

$P[R|H_0]$

A_0, A_1

$H_0 \quad H_1$

$P[A_0|H_1]$

$P[A_1|H_0]$

- Start with two probability models.
  - They are your hypotheses H_0, H_1
    - What are the PDF(s)/CDF(s)?
      - Also called likelihood functions
- Conclusion: Observation from the experiment is from one of the starting Hypotheses
- Outcome is governed by H_0 when there is no target
- Else by H_1
- Accuracy Measure:
  - Miss Detection (Type II): Probability you conclude that the observation is from H_0 when it is truly from H_1
    - Target missed by RADAR when it is present
  - False Alarm (Type I): Probability you conclude that the observation is from H_1 when it is truly from H_0
    - Target detected by RADAR when none exists

Acceptance Sets

- A_0, A_1
- What is False alarm probability? $\rightarrow P[A_0|H_1]$
- What is Miss Detection probability? $\rightarrow P[A_1|H_0]$

Receiver Operating

ROC

- Plot a graph (Pfa (x-axis) vs. Pmiss)

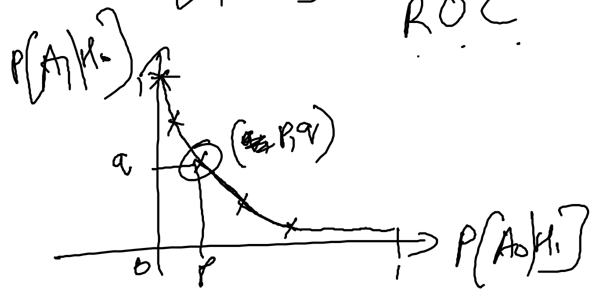
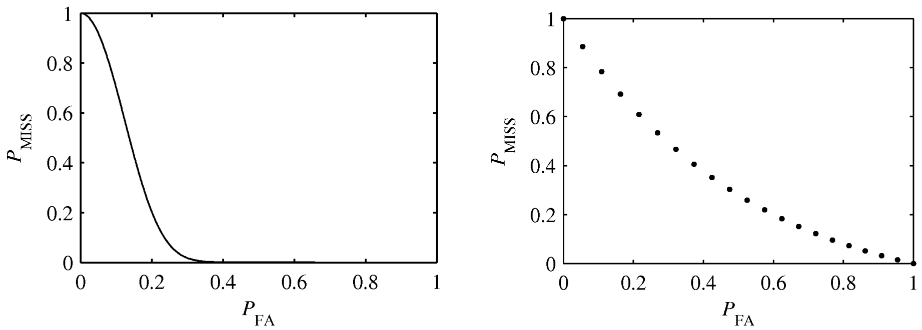


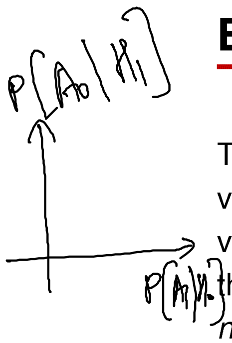
Figure 8.1



ROC for continuous X **ROC for discrete** X

Continuous and discrete examples of a receiver operating curve (ROC).

Example 8.4 Problem



$H_1 \sim N(v,1)$

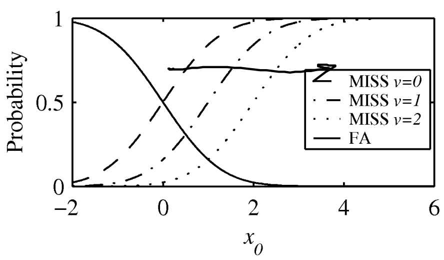
$H_0 \sim N(0,1)$

The noise voltage in a radar detection system is a Gaussian $(0, 1)$ random variable, N. When a target is present, the received signal is $X = v + N$ volts with $v \geq 0$. Otherwise the received signal is $X = N$ volts. Periodically, the detector performs a binary hypothesis test with H_0 as the hypothesis *no target* and H_1 as the hypothesis *target present.* The acceptance sets for the test are $A_0 = \{X \leq x_0\}$ and $A_1 = \{X > x_0\}$. Draw the receiver operating curves of the radar system for the three target voltages $v = 0, 1, 2$ volts.

$\rightarrow P[A_0|H_1] = P[X \leq x_0|H_1] =$

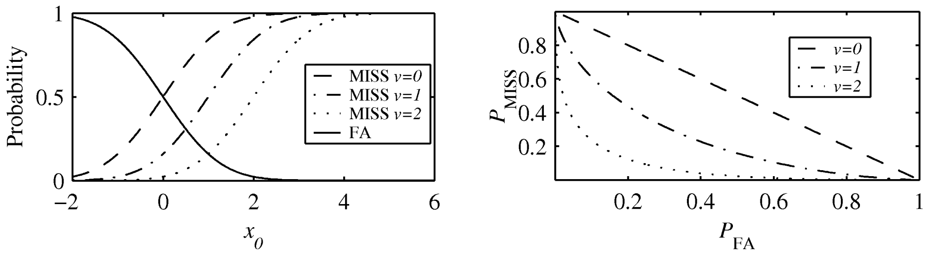
$\rightarrow P(A_1|H_0) = P(X > x_0|H_0)$

Figure 8.2



(a)

Figure 8.2



(a) **(b)**

How Do We Select the Set A_0

Theorem 8.1 Maximum A posteriori Probability (MAP) Binary Hypothesis Test

$$P[H_0|s] = \frac{P[s|H_0]\,P[H_0]}{P[s]}$$

$$P[s|H_0]P[H_0] + P[s|H_1]P[H_1]$$

$P[H_0]$

$P[H_1]$

Assign s to A_0 $\longrightarrow$ $P[s|H_1]\,P[H_1] \le P[s|H_0]\,P[H_0]$

$$P[\text{Error}] = P[A_0|H_1]\,P[H_1] + P[A_1|H_0]\,P[H_0]$$

Given a binary hypothesis testing experiment with outcome s, the following rule leads to the lowest possible value of P_{ERR}:

$$s \in A_0 \text{ if } P[H_0|s] \geq P[H_1|s]; \qquad s \in A_1 \text{ otherwise.}$$

$$\frac{1}{P[H_1]}\,P[H_1|s]\,\cancel{P[s]}\,P[H_1] \le P[H_0|s]\,\cancel{P[s]}\,P[H_0]\,\frac{1}{P[H_0]}$$

$$P[H_1|s] \le P[H_0|s]$$

$$\frac{P[s|H_0]\,P[H_0]}{P[s]} \geq \frac{P[s|H_1]\,P[H_1]}{P[s]}$$

MAP Rule: Accept the hypothesis with the larger a posteriori probability

$$\frac{P[s|H_0]}{P[s|H_1]} \geq \frac{P[H_1]}{P[H_0]}$$

Theorem 8.2

For an experiment that produces a random vector $\mathbf{X}$, the MAP hypothesis test is

Discrete: $\mathbf{x} \in A_0$ if $\dfrac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]}{P[H_0]}$; $\mathbf{x} \in A_1$ otherwise

Continuous: $\mathbf{x} \in A_0$ if $\dfrac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]}{P[H_0]}$; $\mathbf{x} \in A_1$ otherwise.

Evidence in favor of H_1 prior to observation

The likelihood ratio. Evidence in favor of H_0 after an observation

$$\frac{f_{X|H_0}(x)}{f_{X|H_1}(x)} \ge \frac{P[H_1]}{P[H_0]}$$

Note that $f_{X|H_0}(x) \approx \dfrac{P[x \le X < x+\Delta x]}{\Delta x}$

$H_0: N(-v, \sigma)$ $\quad P[H_0] = p$

$H_1: N(v, \sigma)$ $\quad P[H_1] = (1-p)$

Assign ~~x~~ to A_0:

$$\frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left[(x+v)^2/(2\sigma^2)\right]}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-v)^2/2\sigma^2}} \ge \frac{(1-p)}{p}$$

Example 8.6 Problem

With probability p, a digital communications system transmits a 0. It transmits a 1 with probability $1-p$. The received signal is either $X = -v + N$ volts, if the transmitted bit is 0; or $v + N$ volts, if the transmitted bit is 1. The voltage $\pm v$ is the information component of the received signal, and N, a Gaussian $(0, \sigma)$ random variable, is the noise component. Given the received signal X, what is the minimum probability of error rule for deciding whether 0 or 1 was sent?

Is the threshold is a function of the prior probabilities? How?
What else matters?

Select $x \in A_0$ if $x \; \boxed{} \; (\text{Expr})$

Ineq

$$x \le \frac{\sigma^2}{2v} \ln\left(\frac{p}{1-p}\right) \; x^*$$

$A_0 = \{X \leq x^*\}$
$A_1 = \{X > x^*\}$

Example 8.7 Problem

Find the error probability of the communications system of Example 8.6.

$$P[E] = P[A_0|H_1]P[H_1] + P[A_1|H_0]P[H_0]$$
$$= P[X \leq x^*|H_1]P[H_1] + P[X > x^*|H_0]P[H_0]$$

Example 8.8 Problem

$H_0 : \text{Geom}(q_0)$

$H_1 : \text{Geom}(q_1)$

Assign n to A_0

$$\frac{P(N=n \mid H_0)}{P(N=n \mid H_1)} \geq P[H_1]/P[H_0]$$

At a computer disk drive factory, the manufacturing failure rate is the probability that a randomly chosen new drive fails the first time it is powered up. Normally the production of drives is very reliable, with a failure rate $q_0 = 10^{-4}$. However, from time to time there is a production problem that causes the failure rate to jump to $q_1 = 10^{-1}$. Let H_i denote the hypothesis that the failure rate is q_i.

$$\frac{(1-q_0)^{n-1} q_0}{(1-q_1)^{n-1} q_1} \geq \frac{1}{9}$$

Assign n to A_0:

$$n \geq 1 + \frac{\ln(\quad)}{\ln(\quad)}$$

Every morning, an inspector chooses drives at random from the previous day's production and tests them. If a failure occurs too soon, the company stops production and checks the critical part of the process. Production problems occur at random once every ten days, so that $P[H_1] = 0.1 = 1 - P[H_0]$. Based on N, the number of drives tested up to and including the first failure, design a MAP hypothesis test. Calculate the conditional error probabilities P_{FA} and P_{MISS} and the total error probability P_{ERR}.

Theorem 8.3 Minimum Cost Binary Hypothesis Test

For an experiment that produces a random vector $\mathbf{X}$, the minimum cost hypothesis test is

Discrete: $\mathbf{x} \in A_0$ if $\dfrac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]C_{01}}{P[H_0]C_{10}}$; $\quad \mathbf{x} \in A_1$ otherwise

Continuous: $\mathbf{x} \in A_0$ if $\dfrac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq \dfrac{P[H_1]C_{01}}{P[H_0]C_{10}}$; $\quad \mathbf{x} \in A_1$ otherwise.

Example 8.9 Problem

Continuing the disk drive test of Example 8.8, the factory produces 1,000 disk drives per hour and 10,000 disk drives per day. The manufacturer sells each drive for \$100. However, each defective drive is returned to the factory and replaced by a new drive. The cost of replacing a drive is \$200, consisting of \$100 for the replacement drive and an additional \$100 for shipping, customer support, and claims processing. Further note that remedying a production problem results in 30 minutes of lost production. Based on the decision statistic N, the number of drives tested up to and including the first failure, what is the minimum cost test?

Definition 8.1 Maximum Likelihood Decision Rule

For a binary hypothesis test based on the experimental outcome $s \in S$, the maximum likelihood (ML) decision rule is

$$s \in A_0 \text{ if } P\left[s|H_0\right] \geq P\left[s|H_1\right]; \qquad s \in A_1 \text{ otherwise.}$$

Theorem 8.6

If an experiment produces a random vector $\mathbf{X}$, the ML decision rule states

$$\text{Discrete:} \quad \mathbf{x} \in A_0 \text{ if } \frac{P_{\mathbf{X}|H_0}(\mathbf{x})}{P_{\mathbf{X}|H_1}(\mathbf{x})} \geq 1; \qquad \mathbf{x} \in A_1 \text{ otherwise}$$

$$\text{Continuous:} \quad \mathbf{x} \in A_0 \text{ if } \frac{f_{\mathbf{X}|H_0}(\mathbf{x})}{f_{\mathbf{X}|H_1}(\mathbf{x})} \geq 1; \qquad \mathbf{x} \in A_1 \text{ otherwise.}$$