Lecture Notes for EE230
Probability and Random Variables

Department of Electrical and Electronics Engineering
Middle East Technical University

Elif Uysal-Biyikoglu

May 24, 2012

Chapter 1

Elementary Concepts

1.1 Introduction

Applied probability is an extremely useful tool in engineering as well as other fields. Specific examples from our field, electrical engineering, where probability is heavily used are communication theory, networking, detection and estimation. Examples from other disciplines that rely on probabilistic models are statistics, operations research, finance, genetics, games of chance, etc. A working knowledge of applied probability is useful in understanding and interpreting many phenomena in everyday life.

In applied probability, we learn to construct and analyze probabilistic models, using which we can solve interesting problems. It is important to distinguish probability from statistics: probabilistic models that we construct do not belong to the "real world". Rather, they live inside a *probability space*, which is a mathematical construction. Probability Theory is a mathematical theory, based on axioms. Generally, the *three axioms* we will introduce in Section 1.3 are used to define probability theory (due to Kolmogorov in his 1933 book). Probability theory is heavily based on the theory of sets, so we will start by reviewing them.

1.2 Set Theory

Definition 1 *A set is a collection of objects, which are called the elements of the set.*

Ex: $A = \{1, 2, 3, \ldots, \}$, $B = \{$Monday, Wednesday, Friday$\}$, $C = \{$real numbers $(x, y) : \min(x, y) \leq 2\}$. (Finite, Countably Infinite, Uncountably Infinite)

Null set=empty set=$\emptyset$=$\{\}$.

The universal set (Ω): The set which contains all the elements under investigation

Some relations

- A is a subset of B ($A \subset B$) if every element of A is also an element of B.
- A and B are equal ($A = B$) if they have exactly the same elements.

1.2.1 Set Operations

1. UNION

2. INTERSECTION

3. COMPLEMENT of a set

4. DIFFERENCE

Two sets are called disjoint or mutually exclusive if $A \cap B = \emptyset$.

A collection of sets is said to be a partition of a set S if the sets in the collection are disjoint and their union is S.

1.2.2 Properties of Sets and Operations

- Commutative: $A \cup B = B \cup A$

- Associative: $A \cup (B \cup C) = (A \cup B) \cup C$

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

- Distributive: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

- $A \cap \emptyset =$ $A \cup \emptyset =$ $\emptyset^c =$ $\Omega^c =$

- $A \cup A^c =$ $A \cap A^c =$ $A \cup \Omega =$ $A \cap \Omega =$

- De Morgan's Laws
  - $(A \cup B)^c = A^c \cap B^c$
  - $(A \cap B)^c = A^c \cup B^c$

The cartesian product of two sets A and B is the set of all ordered pairs such that $A \times B = \{(a, b) | a \in A, b \in B\}$.

Ex:

$|A|=$ Cardinality of set A (The number of elements in A)

The power set $\mathcal{P}(A)$ of a set A: the set of all subsets of A $|\mathcal{P}(A)| =?$

1.3 Probabilistic Models

A probabilistic model is a mathematical description of an uncertain situation. A probability model consists of an experiment, a sample space, and a probability law.

1.3.1 Experiment

Every probabilistic model involves an underlying process called the experiment.

Ex: Consider the underlying experiments in the two classic probability puzzles: The girl's sibling, and the 3-door problem.

1.3.2 Sample Space

The set of all possible results (OUTCOMES) of an experiment is called the SAMPLE SPACE (Ω) of the experiment.

Ex: List the sample spaces corresponding to the following experiments:

- Experiment 1: Toss a coin and look at the outcome.
  $\Omega =$
- Experiment 2: Toss a coin until you get "Heads".
  $\Omega =$
- Experiment 3: Throw a dart into a circular region of radius r, and check how far it fell from the center.
- Experiment 4: Pick a point (x, y) on the unit square.
- Experiment 5: A family has two children.
- Experiment 6: I select a door, one of the three doors is concealing a prize.

Definition 2 *An event is a subset of the sample space* Ω*.*

- Ω: certain event, $\emptyset$: impossible event
- TRIAL: single performance of an experiment
- An event A is said to have OCCURRED if the outcome of the trial is in A.
- A given physical situation may be modeled in many different ways. The sample space should be chosen appropriately with regard to the intended goal of modeling.
- Sequential models: tree-based sequental description
  Ex: Consider two rounds of the double-and-quarter game and list all possible outcomes. Consider three tosses of a coin and write all possible outcomes.

1.3.3 Probability Law

The probability law assigns to every event A a nonnegative number $P(A)$ called the probability of event A.

$P(A)$ reflects our knowledge or belief about A. It is often intended as a model for the frequency with which the experiment produces a value in A when repeated many times independently.

Ex:

Probability Axioms

1. (Nonnegativity) $P(A) \geq 0$ for every event A
2. (Additivity) If A and B are two disjoint events, then

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

More generally, if the sample space has an infinite number of elements and $A_1, A_2, \ldots$ is a sequence of disjoint events, then

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots.$$

3. (Normalization) $P(\Omega) = 1$

1.3.4 Properties of Probability Laws

(a) $P(\emptyset) = 0$

(b) $P(A^c) = 1 - P(A)$

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

(d) $A \subset B \Rightarrow P(A) \leq P(B)$

(e) $P(A \cup B) \leq P(A) + P(B)$ $(P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i))$

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

1.3.5 Discrete Probability Models

The sample space is a countable (finite or infinite) set in discrete models.

Ex: An experiment involving a single coin toss. We say that the coin is "fair", equal probabilities are assigned to the possible outcomes. That is, $P(H) = P(T) = 1/2$.

In a discrete probability model,

- The probability of any event $\{s_1, s_2, \ldots, s_k\}$ is the sum of the probabilities of its elements. (Recall "additivity".)

$$\begin{aligned} P(\{s_1, s_2, \ldots, s_k\}) &= P(\{s_1\}) + P(\{s_2\}) + \ldots + P(\{s_k\}) \\ &= P(s_1) + P(s_2) + \ldots + P(s_k) \end{aligned}$$

  Ex: Throw a 6-sided die. Express the probability that the outcome is 1 or 6.

- Discrete uniform probability law: If the sample space consists of n possible outcomes which are equally likely, then

$$P(A) = \frac{|A|}{n}.$$

Ex: Throw a fair 6-sided die. Find the probability that the outcome is 1 or 6.

Ex: A fair coin is tossed until a tails is observed. Determine the probabilities of each outcome in the sample space.

Ex: A file contains 1Kbytes. The probability that there exists at least one corrupted byte is 0.01. The probability that at least two bytes are corrupted is 0.005. Let the outcome of the experiment be the number of bytes in error.

(a) Define the sample space.

(b) Find P(no errors).

(c) P(exactly one byte in error)=?

(d) P(at most one byte is in error)=?

1.3.6 Continuous Models

The sample space is an uncountable set in continuous models. We compute the probability by measuring the probability "weight" of the desired event relative to the sample space.

Ex: I start driving to work in the morning at some time uniformly chosen in the interval $[8:30, 9:00]$.

- What is the probability that I start driving before $8:45$?

- My favorite radio program comes on at $8:30$, and may last anywhere between 5 to 15 minutes, with equal probability. What is the probability that I catch at least part of the program?

1.4 Conditional Probability

$P(A|B) =$ probability of A, given that B occurred

Definition 3 *Let A and B two events with $P(B) \neq 0$. The conditional probability $P(A|B)$ of A given B is defined as*

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

Example: Consider two rolls of a tetrahedral die. Let B be the event that the minimum of the two rolls is 2. Let M be the maximum of the rolls.

- $\mathrm{P}(M = 1|B) =$
- $\mathrm{P}(M = 2|B) =$

1.4.1 Properties of Conditional Probability

Conditional probability is a probability law, where B is the new universe.

Theorem 1 *For a fixed event B with $P(B) > 0$, the conditional probabilities $P(A|B)$ form a probability law satisfying all three axioms.*

Proof 1

- If A and B are disjoint, $P(A|B) =$.
- If $B \subset A$, $P(A|B) =$.
- When all outcomes are equally likely,

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

Ex: A girl I met told me she has 1 sibling. What is the probability that her sibling is a boy? (Assumption: each birth results in a boy or girl with equal probability.)

1.4.2 Chain (Multiplication) Rule

Assuming that all of the conditioning events have positive probability, the following expression holds

$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\dots P(A_n|\bigcap_{i=1}^{n-1} A_i).$$

Ex: There are two balls in an urn numbered with 0 and 1. A ball is drawn. If it is 0, the ball is simply put back. If it is 1, another ball numbered with 1 is put in the urn along with the drawn ball. The same operation is performed once more. A ball is drawn in the third time. What is the probability that the drawn balls are all labeled 1?

1.4.3 Total Probability Theorem

- This is the "divide and conquer" idea. Very useful in modelling and solving problems.
- Partition set B into $A_1, A_2, \dots, A_n$. The A_i's should be disjoint inside B. That is, $A_i \cap A_j \cap B = \emptyset$ for all i, j.

- One way of computing $\mathrm{P}(B)$:

$$\mathrm{P}(B) = \mathrm{P}(A_1)P(B|A_1) + \mathrm{P}(A_2)\mathrm{P}(B|A_2) + \ldots \mathrm{P}(A_n)\mathrm{P}(B|A_n)$$

- An often used partition is A and A^c, where A is any event in the sample space, not disjoint with B.

Ex: The "Monty Hall Problem" (Example 1.12 in textbook.) There is a prize behind one of three identical doors. You are told to pick a door. The game show host then opens one of the remaining doors with no prize behind it. At this point, you have the option to switch to the unopened door, or stick to your original choice. What is the better strategy- to stick or to switch? (Hint: Examine each strategy separately. In each, let B be the event of winning, and A the event that the initially chosen door has the prize behind it.) Show that, when you adopt a randomized strategy (you decide whether to switch or not by tossing a fair coin) the probability of winning is 1/2.

1.4.4 Bayes's Rule

This is a rule for combining "evidence".

$$\begin{aligned} \mathrm{P}(A|B) &= \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \\ &= \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B)} \text{ Note how } A \text{ and } B \text{ changed places} \end{aligned}$$

Ex: Criminal X and Criminal Y are both 20 percent likely to commit a certain crime, and they are both 50 percent likely to be near the site of the crime at a given time. As a result of the investigation, it is revealed that Criminal X was near the site at the time of the crime, but Y was not. What are the posterior probabilities of committing the crime for X and Y?

Bayes's rule is often applied to events A_i that form a partition of the given event B.

$$\begin{aligned} \mathrm{P}(A_i|B) &= \frac{\mathrm{P}(A_i \cap B)}{\mathrm{P}(B)} \\ &= \frac{\mathrm{P}(B|A_i)\mathrm{P}(A_i)}{\mathrm{P}(B)} \\ &= \frac{\mathrm{P}(A_i)\mathrm{P}(B|A_i)}{\sum_j \mathrm{P}(A_j)\mathrm{P}(B|A_j)} \end{aligned}$$

Where, going from the second line to the third, we applied the Total Probability Theorem.

Ex: Let B be the event that the sum of the numbers obtained on two tosses of a die is seven. Given that B happened, find the probability that the first toss resulted in a 3.

1.5 Modelling using conditional probability

Ex: If an aircraft is present in a certain area, a radar detects it and generates and alarm signal with probability 0.99. If an aircraft is not present, the radar generates a (false) alarm with probability 0.10. We assume that an aircraft is present with probability 0.05.
Event A: Airplane is flying above
Event B: Something registers on the radar screen

(a) $\mathrm{P}(A \cap B) =$

(b) $\mathrm{P}(B) =$

(c) $\mathrm{P}(A|B) =?$ (Discuss how to improve this probability.)

Ex: The "false positive puzzle" (Example 1.18 in textbook.): A test for a certain disease is assumed to be correct 95% of the time: if a person has the disease, the test results are positive with probability 0.95, and if the person is not sick, the test results are negative with probability 0.95. A person randomly drawn from the population has the disease with probability 0.01. Given that a person is tested positive, what is the probability that the person is actually sick?

1.6 Independence

Definition: $\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$

- If $\mathrm{P}(A) > 0$, independence implies $\mathrm{P}(B|A) = \mathrm{P}(B)$.
- Symmetrically, if $\mathrm{P}(B) > 0$, independence implies $\mathrm{P}(A|B) = \mathrm{P}(A)$.
- Show that, if A and B are independent, so are A and B^c. (If A is independent of B, the occurrence (or non-occurrence) of B does not convey any information about A.)

- Show that, if A and B are disjoint, they are always *dependent*.

Ex: Consider two independent rolls of a tetrahedral die.

(a) Let $A_i = \{\text{the first outcome is } i\}$. Let $B_j = \{\text{the second outcome is } j\}$. "Independent rolls" implies A_i and B_j are independent for any i and j. Find $P(A_i, B_j)$.

(b) Let $A = \{\text{the max of the two rolls is } 2\}$. Let $B = \{\text{the min of the two rolls is } 2\}$. Are A and B independent?

(c) Note that independence can be counter-intuitive. For example, let $A_2 = \{\text{the first roll is } 2\}$. Let $S_5 = \{\text{the sum of the two rolls is } 5\}$. Show that A_2, S_5 are independent, although the sum of the two rolls and the first roll are dependent in general (try A_2, S_6 as a counterexample.)

1.6.1 Conditional Independence

Recall that conditional probabilities form a legitimate probability law. So, A and B are independent, conditional on C, if

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

Show that this implies

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

(assuming $\mathrm{P}(B|C) \neq 0$, $\mathrm{P}(C) \neq 0$.)

Conditioning may affect independence.

Ex: Assume A and B are independent, but $A \cap B \cap C = \emptyset$. If we are told that C occurred, are A and B independent? (draw Venn Diagram exhibiting a counterexample.)

Ex: Two unfair coins, A and B.

$$\mathrm{P}(H|\text{coinA}) = 0.9, \ \mathrm{P}(\mathrm{H}|\text{coinB}) = 0.1$$

Choose either coin with equal probability.

- Once we know it is coin A, are future tosses independent.
- If we don't know which coin it is, are future tosses independent?
- Compare

  $$\mathrm{P}(\text{5th toss is a T})$$

  and P(5th toss is a T|first 4 tosses are T).

Independence of a collection of events: Information on some of the events tells us nothing about the occurrence of the others.

- Events A_i, $i = 1, 2, \ldots, n$ are independent iff $\mathrm{P}(\cap_{i \in S} A_i) = \pi_{i \in S} \mathrm{P}(A_i)$ for any $S \subset \{1, 2, \ldots, n\}$
- Note that

$$\mathrm{P}(A_5 \cup A_2 \cap (A_1 \cup A_4)^c | A_3 \cup A_6^c) = \mathrm{P}(A_5 \cup A_2 \cap (A_1 \cup A_4)^c)$$

- Pairwise independence does not imply independence! (Checking $\mathrm{P}(A_i \cap A_j) = \mathrm{P}(A_i)\mathrm{P}(A_j)$ for all i and j is not sufficient for confirming independence.)
- For three events, checking $\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\mathrm{P}(A_2)\mathrm{P}(A_3)$ is not enough for confirming independence.

Ex: Consider two independent tosses of a fair coin. $A =$ First toss is H.
$B =$ Second toss is H.
$C =$ First and second toss have the same outcome.
Are these events pairwise independent?

$$\begin{aligned} \mathrm{P}(C) &= \\ \mathrm{P}(C \cap A) &= \\ \mathrm{P}(C \cap A \cap B) &= \\ \mathrm{P}(C | B \cap A) &= \end{aligned}$$

Ex: Network Connectivity: In the electrical network in Fig. 1.2, each circuit element is "on" with probability p, independently of all others. What is the probability that there is a connection between points A and B?

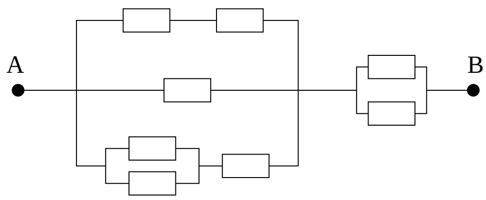


Figure 1.1: Electrical network with randomly operational elements.

1.6.2 Independent Trials and Binomial Probabilities

Consider n tosses of a coin with bias p. P(k H's in an n-toss sequence) $=$ $\binom{n}{k}p^k(1-p)^{n-k}$

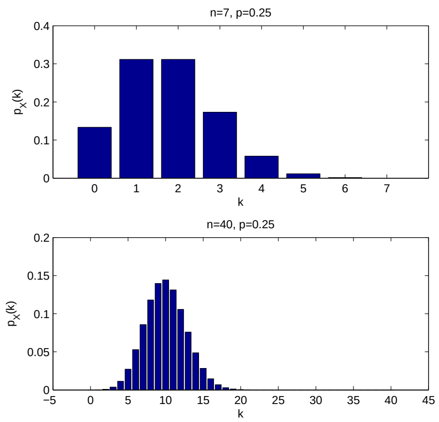


Figure 1.2: Binomial probability law.

Ex: Binary symmetric channel: Fig. 1.3 depicts a binary symmetric channel, where each symbol ("0" or "1") sent is inverted (turned to "1" or "0", respectively)with probability p_o, independently of all other symbols. (First consider the case where 1s and 0s are equiprobable, then the case when they are not.)

(a) What is the probability that a string of length n is received correctly?

(b) Given that a "110" is received, what is the probability that actually a "100" was sent?

(c) In an effort to improve reliability, each symbol is repeated 3 times and the received string is decoded by majority rule. What is the probability that a transmitted "1" is correctly decoded?

(d) Can you think about a better coding scheme than the one in (c)?

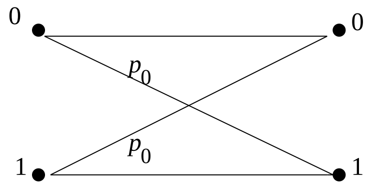


Figure 1.3: Binary symmetric channel.

1.7 Counting

A special case: all outcomes are equally likely.

$$\begin{aligned}
\Omega &= \{s_1, s_2, \ldots, s_n\} \\
\mathrm{P}(\{s_j\}) &= \frac{1}{n}, \text{for all } j \\
A &= \{s_{j_1}, s_{j_2}, \ldots, s_{j_k}\}, j_k \in \{1, 2, \ldots, n\} \\
\mathrm{P}(A) &=
\end{aligned}$$

The problem of finding $\mathrm{P}(A)$ reduces to counting its elements.

Ex: 6 balls in an urn, $\Omega = \{1, 2, \ldots, 6\}$.
$A = \{$the number on the ball drawn is divisible by 3$\}$.

Ex: A. (Permutations) The number of different ways of picking an ordered set of k out of n distinct objects

Ex: B. (Combinations) The number of different ways of choosing a group of k out of n distinct objects

Ex: C. (Partitions) How many different ways can a set of size n be partitioned into r disjoint subsets, with the i^{th} subset having size n_i? For example, pick a captain, a goalie and five players from among 7 friends.

Ex: D. (Distributing n identical objects into to r boxes) Consider n identical balls, to be colored red, black or white. How many possible configurations for the numbers of red, black and white balls? (Think about putting dividers between objects and shuffling objects and dividers.)

Ex: Categorize the following examples with respect to the following two criteria: Is the sampling with or without replacement? Does ordering matter or not?

(a) How many distinct words can you form by shuffling the letters of PROBABILITY?

(b) As a result of a race with 100 entrants, how many possibilities for the gold, silver and bronze medalists?

(c) Choose a captain, goalie and 5 players from a group of 9 friends.

(d) Choose a team of 7 from among 9 friends.

(e) How many possible car plate numbers are there in Ankara (assume two or three letters, and two or three digits are used on a car plate, chosen out of 23 letters and 9 numerals)?

(f) I can use the numbers 0, 1, and 9 arbitrarily many times to form a sequence of length 8. How many possibilities are there for the total weight of my sequence (sum of all numbers in the sequence)?

(g) Find the number of solutions of the equation $x_1 + x_2 + \ldots + x_r = n$, where $n \geq 1$ and $x_i \geq 0$'s are integers. (Hint: note that this is an example of "type-D" as well. Also think of the case where $x_i > 0$. In that case, there has to be at least one ball in each bin.)

Chapter 2

Discrete Random Variables

2.1 Preliminaries

Definition 4 *A random variable is a mapping (a function) from the sample space into real numbers.*

- We can define an arbitrary number of different random variables on the same sample space.

Ex: Toss a fair 6-sided die. Let the random variable X take on the value 1 if the outcome is 6, and 0 otherwise. Let the random variable Y be equal to the outcome of the die. Illustrate the mappings from the sample space associated with X and Y. (Note that $\{X = 1\} = \{\text{outcome is } 6\} = A$, and $\{X = 0\} = A^c$.)

Definition 5 *A discrete random variable takes a discrete set of values. The Probability Mass Function (PMF) of a discrete random variable is defined as*

$$p_X(x) = \mathrm{P}(X = x)$$

Ex: Find and plot the PMFs of X and Y defined in the previous example.

- A discrete random variable is completely characterized by its PMF.

Ex: Let M be the maximum of the two rolls of a fair die. Find $p_M(m)$ for all m. (Think of the sample space description and the sets of outcomes where M takes on the value m.)

2.2 Some Discrete Random Variables

2.2.1 The Bernoulli Random Variable

In the rest of this course, we shall define the Bernoulli random variable with parameter p as the following:

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

In shorthand we say $X \sim$Ber(p).

Ex: Express and sketch the PMF of a Bernoulli(p) random variable.

Despite its simplicity, the Bernoulli r.v. is very important since it can model generic probabilistic situations with just two outcomes (often referred to as binary r.v.).

Examples:

- Indicator function: Consider the random variable X defined previously. $X(w) = 1$ if outcome $w \in A$, and $X(w) = 0$ otherwise. So, X indicates whether the outcome is in set A or A^c. X, a Bernoulli random variable, is sometimes called the "indicator function" of the event A. This is sometimes denoted as $X(w) = I_A(w)$.

- Consider n tosses of a coin. Let $X_i = 1$ if the i^{th} roll comes up H, and $X_i = 0$ if it comes up T. Each of the X_i's are *independent* Bernoulli random variables. The X_i's, $i = 1, 2, \ldots$ are a sequence of independent "Bernoulli Trials".

- Let Z be the total number of successes in n independent Bernoulli trials. Express Z in terms of n independent Bernoulli random variables.

2.2.2 The Geometric Random Variable

Consider a sequence of independent Bernoulli trials where the probability of success in each trial is p (We will later call this a "Bernoulli Process".) Let Y be the number of trials up to and including the first success. Y is a *Geometric* random variable with parameter p.

$\mathrm{P}(Y = k) =$ for $k =$

Sketch $p_Y(k)$ for all k.

Check that this is a legitimate PMF.

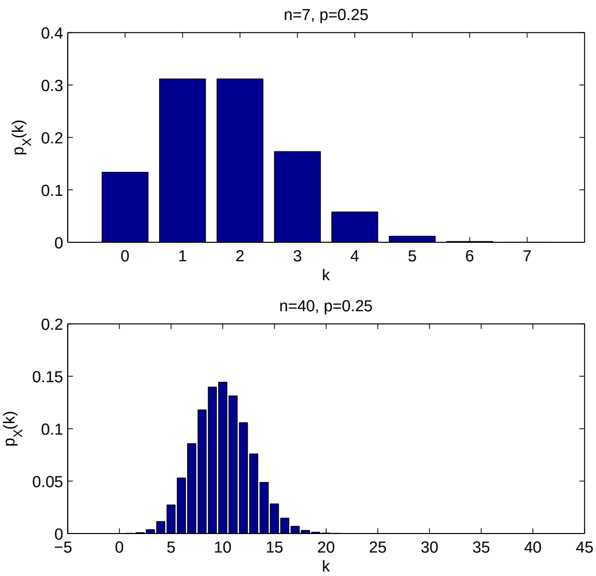
Ex: Let Z be the number of trials up to (but not including) the first success. Find and sketch $p_Z(z)$.

2.2.3 The Binomial Random Variable

Consider n independent Bernoulli Trials each with probability of success p, and let B be the number of successes in the n trials. B is Binomial with parameters (n, p).

$\mathrm{P}(B = k) =$ for $k =$

Ex: Let R be the number of Heads in n independent tosses of a coin with bias p.



2.2.4 The Poisson Random Variable

A Poisson random variable X with parameter λ has the PMF

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

Ex: Show that $\sum_k p_X(k) = 1$ (Hint: use the Taylor series expansion of e^{λ}.

- The Binomial is a good approximation for the Poisson with $\lambda = np$ when n is very large and p is small, for small values of k. That is, if $k \ll n$

$$\frac{\lambda^k e^{-\lambda}}{k!} \approx \frac{n!}{k!(n-k)!} p^k (1-p)^{(n-k)}$$

2.2.5 The Discrete Uniform R.V.

The discrete uniform random variable takes consecutive integer values within a finite range with equal probability. That is, X is Discrete Uniform in $[a, b]$, $b > a$ if and only if

$$p_X(k) = 1/(b - a + 1) \text{ for } k = a, a+1, a+2, \ldots, b$$

Ex: A four-sided die is rolled. Let X be equal to the outcome, Y be equal to the outcome divided by three, and Z be equal to the square of the outcome.

(Note that Y and Z both take four equally likely values, however they do not have the discrete uniform distribution.)

2.3 Functions of Random Variables

$$Y = f(X)$$

Ex: Let X be the temperature in Celsius, and Y be the temperature in Fahrenheit. Clearly, Y can be obtained if you know X.

$$Y = 1.8X + 32$$

Ex: $\mathrm{P}(Y \geq 14) = \mathrm{P}(X \geq ?)$

Ex: A uniform r.v. X whose range is the integers in $[-2, 2]$. It is passed through a transformation $Y = |X|$.

To obtain $p_Y(y)$ for any y, we add the probabilities of the values x that results in $g(x) = y$:

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x).$$

Ex: A uniform r.v. whose range is the integers in $[-3, 3]$. It is passed through a transformation $Y = u(X)$ where $u(\cdot)$ is the discrete unit step function.

2.4 Expectation and Variance

We are sometimes interested in a summary of certain properties of a random variable.

Ex: Instead of comparing your grade with each of the other grades in class, as a first approximation you could compare it with the class average.

Ex: A fair die is thrown in a casino. If 1 or 2 shows, the casino will pay you a net amount of $30,000$ TL (so they will give you your money back plus 30,000), if $3, 4, 5$ or 6 shows you they will take the money you put down. Up to how much would you pay to play this game?

Ex: Alternatively, suppose they give you a total of $30,000$ if you win (regardless of how much you put down), and nothing if you lose. How much would you pay to play this game?

(Answer: the value of the first game (the break-even point) is 15,000, and for the second game, it is 10,000. In the second game, you expect to get 30,000 with probability 1/3, so you expect to get 10,000 on average.)

Definition 6 *The* ***expected value*** *or* ***mean*** *of a discrete r.v. X is defined as*

$$E[X] = \sum_x xP(X = x) = \sum_x xp_X(x).$$

The intuition for the definition is a weighted some of the values the r.v. takes, where the weights are the probability masses of these values.

The mean of X is a representative value, which lies somewhere in the middle of its range. The definition above tells us that the mean corresponds to the *center of gravity* of the PMF.

Ex: Let X be your net earnings in the (first) Casino problem above, where you put down $12,000$ TL to play the game. Find $E(X)$.

Answer: $E(X) = 2000$ (You expect to make money, and the Casino expects to lose money. A more realistic Casino would charge you something strictly more than 15,000, so that they can expect to make a profit.)

2.4.1 Variance, Moments, and the Expected Value Rule

A very important quantity that provides a measure of the spread of X around its mean is variance.

$$var(X) = E[(X - E[X])^2] \tag{2.1}$$

The variance is always nonnegative. One way to calculate $var(X)$ is to use the PMF of $(X - E[X])^2$.

Ex: Find the variance of the random variables X with the following PMFs.

(a) $p_X(15) = p_X(20) = p_X(25) = 1/3$.

(b) $p_X(15) = p_X(25) = 1/2$.

The standard deviation of X is also a measure of the spread of X around its mean: $\sigma_X = \sqrt{var(X)}$. It is usually simpler to interpret since it has the same units as X.

Another way to evaluate $var(X)$ is by using the following result.

Theorem 2 *Let X be a r.v. with PMF $p_X(x)$ and $g(X)$ be a function of X. Then,*

$$E[g(X)] = \sum_x g(x) p_X(x).$$

Proof: Exercise.

Note: Unless $g(x)$ is a linear function, $E[g(X)]$ is in general not equal to $g(E[X])$.

Ex: When I listen to Radyo ODTU in the morning, I drive at a speed of 50 km per hour, and otherwise I drive at 70 km per hour. Suppose I listen to Radyo ODTU with probability 0.3 on any given day. What is the average duration of my 5 km trip to work?
Answer: 4.8 minutes.

Notes: The trip duration T is a nonlinear function $T = D/V$ of the speed V. In fact it is a convex function, which means $E[g(X)] > g(E[X])$. So it would be **wrong** to calculate the expected speed, which is 0.3*50+0.7*70=64 km/hour, and find the expected duration as 5/64*60=4.68 min.

2.4.2 Properties of Expectation and Variance

Expectation is always linear: $E(aX + b) = aE(X) + b$, which follows from the definition (note that the definition is a linear sum.)

Evaluating variance in terms of moments is sometimes more convenient.

$$var(X) = E[X^2] - (E[X])^2 \tag{2.2}$$

Proof:

Variance is NOT linear: $\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$.

Proof:

Consequently,

- adding a constant to a random variable does not change its variance,
- scaling a random variable by a scales the variance by a^2,
- the variance of a constant is 0 (and conversely, a random variable with

zero variance is a deterministic constant.)

Ex: As exercise, derive the mean and variance of the

- Bernoulli(p) random variable
- Discrete Uniform[a,b] random variable.
- Poisson(λ) random variable.

2.5 Joint PMFs of multiple random variables

Often, we need to be able to think about more than one random variable defined on the same probability space. They may or may not contain information about each other. Consider:

- two signals received as a result of two radar measurements
- the current workload at each of a group of network routers
- your grades received from three consecutive exams

Let X and Y be random variables defined on the same probability space. Their joint PMF is defined as the following.

$$p_{X,Y}(x,y) = P(X = x, Y = y)$$

More precise notations for $\mathrm{P}(X = x, Y = y)$: $\mathrm{P}(X = x$ and $Y = y)$, $\mathrm{P}(\{X = x\} \cap \{Y = y\})$, $\mathrm{P}(\{X = x, Y = y\})$.

$$\mathrm{P}((X, Y) \in A) =$$

The term **marginal PMF** is used for $p_X(x)$ and $p_Y(y)$ to distinguish them from the joint PMF. Can one find marginal PMFs from the joint PMF?

Note that the event $\{X = x\}$ is the union of the disjoint sets $\{X = x, Y = y\}$ as y ranges over all the different values of Y. Then,

$$\begin{aligned} p_X(x) &= \mathrm{P}(X = x) \\ &= \mathrm{P}(\{X = x\}) = \mathrm{P}(\bigcup_y \{X = x, Y = y\}) \\ &= \sum_y \mathrm{P}(\{X = x, Y = y\}) = \sum_y \mathrm{P}(X = x, Y = y) \\ &= \sum_y p_{X,Y}(x, y). \end{aligned}$$

Similarly, $p_Y(y) = \sum_x p_{X,Y}(x, y)$.

The tabular method can be utilized to obtain the marginal PMFs from the joint PMF.

Ex: Two r.v.s X and Y have the joint PMF given in the 2-D table.

	x=1	2	3
y=1	0	1/10	2/10
2	1/10	1/15	1/30
3	2/10	2/10	1/10

Ex: For the joint PMF in the previous example, please compute the following:

1. $\mathrm{P}(X < Y) =$

2. $p_X(x) =$

Ex: Consider the joint PMF of random variables X and Y which take positive integer values:

$$p_{X,Y}(x,y) = \begin{cases} c & 1 < x + y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

1. $c =?$

2. Find the marginals.

2.6 Functions of Multiple Random Variables

Let $Z = g(X,Y)$.

$$p_Z(z) = \sum_{\{(x,y)|g(x,y)=z\}} p_{X,Y}(x,y)$$

The expected value rule naturally extends to functions of more than one random variable:

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

Special case when g is linear: $g(X,Y) = aX + bY + c$

$$E[aX + bY + c] =$$

Ex: Expectation of the Binomial r.v.

Ex: The hat problem: The hats of n people are shuffled and randomly redistributed to them. What is the expected number of people getting their own hat? (Alternatively, consider a string of length n randomly formed by shuffling the string 123..n without any regard to the integers. Note that the probability of integer i occuring in the ith location is $1/n$ for each i, by symmetry. Now, compute the expected number of integers staying in their original locations.)

Ex: Multi-sensor laser communication: On/off signaling can be used to

transmit bits in laser communication. In an on period of 1μsec, the number of photons detected at a sensor is a Poisson with parameter $\lambda = 100$. In a high-quality system, five such sensors are used to enhance communication. Find the expected value of the total number of photons detected in the system in 1μsec.

2.7 Conditioning

The conditional PMF of the random variable X, conditioned on the event A with $\mathrm{P}(A) > 0$ is defined by:

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

Show that $p_{X|A}$ is a legitimate PMF. (Expand $\mathrm{P}(A)$ using the total probability theorem)

Ex: Let X be the outcome of one roll of a tetrahedral die, and A be the event that we did not get 1.

Ex: Ali will take the motorcycle test again and again until he passes; however, he is only allowed n chances to take the test. Suppose each time Ali takes the test, his probability of passing is p, irrespective of what happened in the previous attempts. What is the PMF of the number of attempts, given that he passes?

Ex: Consider an optical communications receiver that uses a photodetector that counts the number of photons received within a constant time unit. The sender conveys information to the receiver by transmitting or not transmitting photons. There is shot noise at the receiver, and consequently even if nothing is transmitted during that time unit, there may be a positive count of photons. If the sender transmits (which happens with probability 1/2), the number of photons counted (including the noise) is Poisson with parameter $a + n$. If nothing is transmitted, the number of photons counted by the detector is again Poisson with parameter n. Given that the detector counted k photons, what is the probability that a signal was sent? Examine the behavior of this probability with a, n and k.

2.7.1 Conditioning one random variable on another

$$p_{X|Y}(x|y) = \mathrm{P}(X = x|Y = y)$$

Show that

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

The function $p_{X|Y}(x|y)$ has the same shape as $p_{X,Y}(x,y)$ (a slice through the joint pmf at a fixed value of y), and because of the normalization (division by $p_Y(y)$), it is a legitimate PMF.

Ex: The joint PMF of two r.v.s X and Y that share the same range of values $\{0, 1, 2, 3\}$ is given by

$$p_{X,Y}(x,y) = \begin{cases} 1/7 & 1 < x+y \leq 3 \\ 0 & \text{otherwise} \end{cases}.$$

Find $p_{X|Y}(x|y)$ and $p_{Y|X}(y|x)$.

One can obtain the following sequential expressions directly from the definition:

$$\begin{aligned} p_{X,Y}(x,y) &= p_X(x)p_{Y|X}(y|x) \\ &= p_Y(y)p_{X|Y}(x|y). \end{aligned}$$

Ex: A die is tossed and the number on the face is denoted by X. A fair coin is tossed X times and the total number of heads is recorded as Y.

(a) Find $p_{Y|X}(y|x)$.

(b) Find $p_Y(y)$.

2.7.2 Conditional Expectation

Let X and Y be random variables defined in the same probability space, and let A be an event such that $\mathrm{P}(A) > 0$.

$E(X|A) =$

$E(g(X)|A) =$

$E(X|Y = y) =$

Furthermore, let A_i, $i = 1, \ldots, n$ be a disjoint partition of the sample space. Then,

$E(X) = \sum_i E(X|A_i)P(A_i)$

$E(X|B) = \sum_i E(X|B \cap A_i)P(A_i|B)$

$E(X) = \sum_y p_Y(y)E(X|Y = y)$

The above three are statements of the "Total Expectation Theorem".

Ex: Data flows entering a router are low rate with probability 0.7, and high rate with probability 0.3. Low rate sessions have a mean rate of 10 Kbps, and high rate ones have a rate of 200 Kbps. What is the mean rate of flow entering the router?

Ex: Find the mean and variance of the Geometric random variable (with parameter p) using the Total Expectation Theorem. (Hint: condition on the events $\{X = 1\}$ and $\{X > 1\}$.

Ex: X and Y have the following joint distribution:

$$p_{XY}(x,y) = \begin{cases} 1/27 & x \in \{4,5,6\}, y \in \{4,5,6\} \\ 2/27 & x \in \{1,2,3\}, y \in \{1,2,3\} \end{cases}$$

Find $E(X)$ using the total expectation theorem.

Ex: Consider three rolls of a fair die. Let X be the total number of 6's, and Y be the total number of 1's. Find the joint PMF of X and Y, $E(X|Y)$ and $E(X)$.

Reading assignment: Example 2.18: The two envelopes paradox, and Problem 2.34: The spider and the fly problem.

2.7.3 Iterated expectation

Using the total expectation theorem lets us compute the expectation of a random variable iteratively: To compute $E(X)$, first determine $E(X|Y)$, then use:

$$E(X) = E[E(X|Y)]$$

The outer expectation is over the marginal distribution of Y. This follows from the total expectation theorem, because it is simply a restatement of:

$$E(X) = \sum_y E(X|Y = y)p_Y(y) == E[E(X|Y)]$$

(recall that $E(X|Y)$ is a random variable, taking values $E(X|Y = y)$ with probability $p_Y(y)$.)

Ex: The joint PMF of the random variables X and Y takes the values $[3/12, 1/12, 1/6, 1/6, 1/6, 1/6]$ at the points $[(-1, 2), (1, 2), (1, 1), (2, 1), (-1, -1), (1, -1)]$, respectively. Compute $E(X)$ using iterated expectations.

Ex: Consider three rolls of a fair die. Let X be the total number of 6's, and Y be the total number of 1's. Note that $E(X) = 1/2$. Confirm this result by computing $E(X|Y)$ and then $E(X)$ using iterated expectations.

2.8 Independence

The results developed here will be based on the independence of events we covered in before. Two events A and B are independent if $\mathrm{P}(A \cap B) = \mathrm{P}(A, B) = \mathrm{P}(A)\mathrm{P}(B)$.

2.9 Independence of a R.V. from an Event

Definition 7 *The random variable X is independent of the event A if*

$$\mathrm{P}(\{X = x\} \cap A) = \mathrm{P}(X = x)\mathrm{P}(A) = p_X(x)\mathrm{P}(A)$$

for all x.

Ex: Consider two tosses of a coin. Let X be the number of heads and let A be the event that the number of heads is even. Show that X is NOT independent of A.

2.10 Independence of Random Variables

Consider two events $\{X = x\}$ and $\{Y = y\}$.

Definition 8 *Two random variables X and Y are independent if*

$$P(\{X = x\} \cap \{Y = y\}) = P(\{X = x\})P(\{Y = y\})$$

for all x, y.

Intuitively speaking, knowledge on Y conveys no information on X, and vice versa.

Independence of two random variables conditioned on an event A: $p_{X,Y|A}(x, y) = p_{X|A}(x)p_{Y|A}(y)$ for all x, y.

When X and Y are independent, $E[XY] = E[X]E[Y]$.

If X and Y are independent, so are $g(X)$ and $h(Y)$.

The independence definition given above can be extended to multiple random variables in a straightforward way. For example, three random variables X, Y, Z are independent if:

2.10.1 Variance of the Sum of Independent Random Variables

Let us calculate the variance of the sum $X + Y$ of two independent random variables X, Y.

If one repetitively uses the above result, the general formula for the sum of independent random variables is obtained:

Ex: During April in Ankara, it rains with probability p each day, independently of every other day. Compute the variance of the number of rainy days in the month. Consider how the variance changes with p.

Ex: Show that, when $E(XY) = E(X)E(Y)$ is satisfied, then the variance of the sum $X + Y$ is equal to the sum of the variances, that is:

$$E(XY) = E(X)E(Y) \rightarrow \mathrm{var}(X + Y) = \mathrm{var}X + \mathrm{var}Y$$

- Note that $E(XY) = E(X)E(Y)$ always holds when X and Y are independent. In general, when $E(XY) = E(X)E(Y)$ holds, the random variables are said to be "uncorrelated", they are not necessarily independent.

- Also note that in contrast, expectation is always linear, expectation of the sum is equal to the sum of expectations:

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

  This is true whether the random variables are dependent or not.

Ex: The number of e-mail messages I get every day is Poisson distributed with mean 10. Let L be the total number of e-mail messages I receive in a week. Compute the mean and variance of R.

Ex: (Mean and variance of the sample mean) An opinion poll is conducted to determine the average public opinion on an issue. It is modelled that a person randomly selected from the society will vote in favour of the issue with probability p, and against it with probability $1 - p$, independently of everyone else. The goal of the survey is to estimate p. To keep the cost of the poll at a minimum, we are interested in surveying the smallest number of people such that the variance of the result is below 0.001. (Hint: Note that an upperbound on the variance of a Bernoulli random variable is 1/4.)

Chapter 3

General Random Variables

3.1 Continuous Random Variables

Definition 9 *A random variable X is continuous if there is a nonnegative function f_X, called the probability density function (PDF) such that*

$$P(X \in B) = \int_{x \in B} f_X(x)dx$$

for every subset B of the real line.

The probability that the value of X falls within an interval is

$$\mathrm{P}(a \leq X \leq b) = \int_a^b f_X(x)dx,$$

which can be interpreted as the area under the graph of the PDF.

3.1.1 Properties of PDF

If $f_X(x)$ is a PDF , the following hold.

1. Nonnegativity: $f_X(x) \geq 0$
2. Normalization property:
3. (for small δ) $\mathrm{P}(x < X \leq x+\delta) = \int_x^{x+\delta} f_X(a)da \approx$

By the last item, $f_X(x)$ can be viewed as the "probability mass per unit length near x". Although it is used to evaluate probabilities of some events, $f_X(x)$ is not itself an event's probability. It tells us the relative concentration of probability around the point x.

Ex: ($f_X(x)$ may be larger than 1)

$$f_X(x) = \begin{cases} cx^2 & , 0 \leq x \leq 1 \\ 0 & , o.w. \end{cases}$$

1. Find c.
2. Find $P(|X|^2 \leq 0.5)$.

Ex: (A PDF can take arbitrarily large values) Sketch the following PDF.

$$f_X(x) = \begin{cases} c/(\sqrt{x}) & , |x| \leq 2 \\ 0 & , o.w. \end{cases}$$

3.1.2 Some Continuous Random Variables and Their PDFs

Continuous Uniform R.V.

We sometimes have information only about the interval of a random variable and nothing else. A PDF used very commonly in such a case is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & , a < x < b \\ 0 & , o.w. \end{cases} .$$

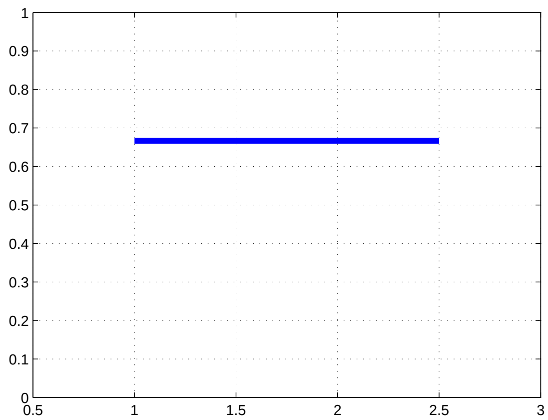


Figure 3.1: Uniform PDF

Gaussian R.V.

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

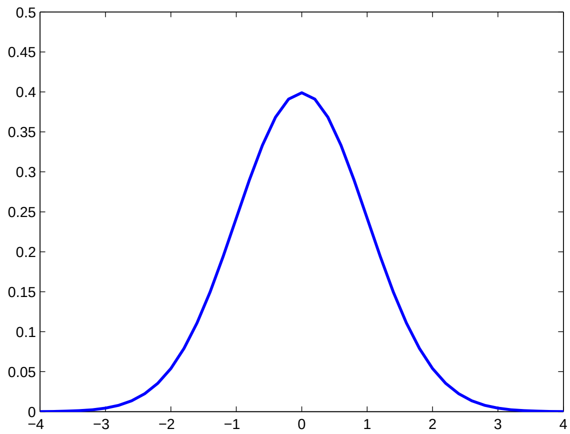


Figure 3.2: Gaussian (normal) PDF

Exponential R.V.

An exponential r.v. has the following PDF

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

where λ is a positive parameter.

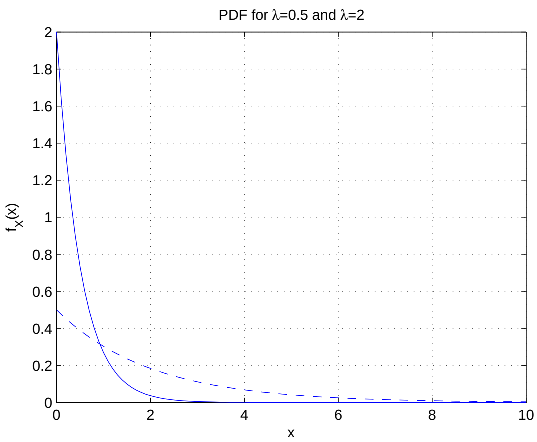


Figure 3.3: Exponential PDF

3.2 Expectation

The expected value (or the *mean*) of a continuous random variable X with PDF $f_X(x)$ is defined as

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

whenever the integral converges[1].

The following results can be show similarly as in the discrete case:

- Note that $E[X]$ is the center of gravity of the PDF.

- $E[g(X)] = \dots$

- $\text{var}X = \dots$

- If $Y = aX + b$, $E[Y] = \dots$ $\qquad$ $\text{var}Y = \dots$

Ex: Compute expectations and variances for Uniform and Exponential random variables.

[1] If the integral is not finite, X is said to "not have a well-defined expectation".

3.3 Cumulative Distribution Functions

Definition 10 *The cumulative distribution function (CDF) of a random variable X is defined as*

$$F_X(x) = \mathrm{P}(X \leq x).$$

In particular,

$$F_X(x) = \begin{cases} \sum_{k \leq x} p_X(k) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{x} f_X(\tau) d\tau, & \text{if } X \text{ is continuous} \end{cases}.$$

The CDF $F_X(x)$ accumulates probability "up to (and including)" the value x.

3.3.1 Properties of Cumulative Distribution Functions (CDFs)

(a) $0 \leq F_X(x) \leq 1$

(b) $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) =$, $F_X(\infty) =$

(c) $\mathrm{P}(X > x) = 1 - F_X(x)$

(d) $F_X(x)$ is a monotonically nondecreasing function: if $x \leq y$, then $F_X(x) \leq F_X(y)$.

Proof:

(e) If X is discrete, CDF is a piecewise constant function of x.

(f) If X is continuous, CDF is a continuous function of x. (The definition of a continuous r.v. is based on this.)

(g) $\mathrm{P}(a < X \leq b) = F_X(b) - F_X(a)$.
Proof:

(h) $F_X(x)$ is continuous from the right. That is, for $\delta > 0$

$$\lim_{\delta \to 0} F_X(x+\delta) = F_X(x^+) = F_X(x).$$

Proof:

(i) $\mathrm{P}(X = x) = F_X(x) - F_X(x^-)$, where $F_X(x^-) = \lim_{\delta \to 0} F_X(x - \delta)$.
Proof:

(j) When X is discrete with integer values,

$$F_X(x) = \sum_{k \le x} p_X(k),$$
$$p_X(x) = \mathrm{P}(X = x) = F_X(x) - F_X(x^-) = F_X(x) - F_X(x-1).$$

(k) If X is continuous,

$$F_X(x) = \int_{-\infty}^{x} f_X(\tau)d\tau,$$
$$f_X(x) = \frac{dF_X(x)}{dx}.$$

The second equality is valid for values of x where $F_X(x)$ is differentiable.

Ex: Let X be exponentially distributed with parameter λ. Derive and sketch $F_X(x)$, the CDF of X.

3.3.2 CDFs of Discrete Random Variables

Discrete random variables have piecewise constant CDFs.

Ex: Find and sketch the CDF of a geometric random variable with parameter p. Compare and contrast this with the CDF of an exponential random variable with rate λ, when p is selected as $1 - e^{-\lambda\delta}$, as $\delta > 0$ becomes arbitrarily small.

Ex: You are allowed to take a certain exam n times, and let X_1, X_2, ..., X_n be the grades you get from each of these trials. Assume that the grades are independent and identically distributed. The maximum will be your final score. Find the CDF of your final score, in terms of the CDF of one exam grade. What happens to the CDF as n increases?

Ex: CDF of a r.v. which is neither continuous nor discrete

3.3.3 The Gaussian CDF

The random variable X is Gaussian, in other words, Normal, with parameters (μ, σ^2) if it has the PDF:

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

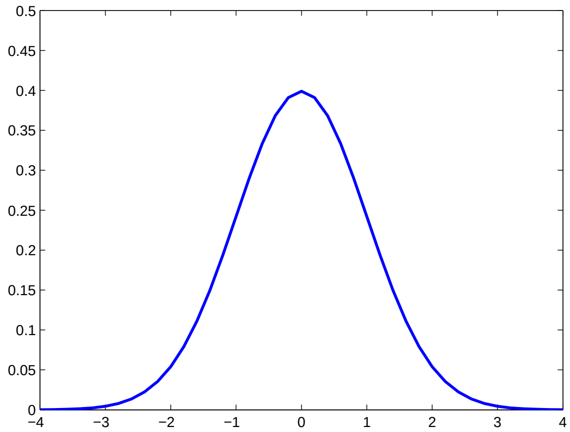


Figure 3.4: Gaussian (normal) PDF

X is said to be a Standard Normal if it's Normal (i.e. Gaussian) with mean 0 and variance 1. That is,

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

The CDF of the standard Gaussian is defined as follows:

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

Note that this is the area under the standard Gaussian curve, up to point x. We often use the function $\Phi(x)$ to make calculation involving general Gaussian random variables.

Normality is preserved by linear transformations: If X is Normal(μ, σ^2) and $Y = aX + b$, then Y is Normal$(a\mu + b, a^2\sigma^2)$ (We can prove this after we learn about Transforms of PDFs.) So, we can obtain any Gaussian by making a linear transformation on a standard Gaussian. That is, letting X be a standard Gaussian, if we let $Y = \sigma X + \mu$, then Y is Normal with mean μ and variance σ^2.

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Then we have:

$$P(Y \leq y) = P((Y-\mu)/\sigma \leq (y-\mu)/\sigma) = P(X \leq (y-\mu)/\sigma) = \Phi((y-\mu)/\sigma)$$

Ex: (Adapted from Ex 3.7 from the textbook.) The annual snowfall at Elmadağ̃ is modeled as a normal random variable with a mean of $\mu = 150$ cm and a standard deviation of 50 cm. What is the probability that next year's snowfall will be at least 200 cm? (Note that from the standard normal table, $\Phi(1) = 0.8413$.)

Ex: Signal Detection (Adapted from Ex 3.7 from the textbook.) A binary message is transmitted as a signal S, which is either $+1$ or -1 with equal probability. The communication channel corrupts the transmission with additive Gaussian noise with mean $\mu = 0$ and variance σ^2. The receiver concludes that the signal $+1$ (or -1) was transmitted if the value received is not negative (or negative, respectively). Find the probability of error in terms of σ and study its behaviour with increasing σ.

3.4 Conditional PDFs

The conditional PDF of a continuous random variable X, given an event $\{X = A\}$, with $\mathrm{P}(\{X = A\}) > 0$ is defined by

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{\mathrm{P}(\{X=A\})}, & \text{if } x \in A \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, for any set B,

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

Ex: The exponential random variable is memoryless: Let X be the lifetime of a lightbulb, exponential with parameter λ. Given $X > t$, find $\mathrm{P}(X > t + x)$.)

Conditional Expectation:

$E(X|A) =$

$E(g(X)|A) =$

Ex: The voltage across the terminals of a power source is known to be between 4.8 to 5.2 Volts, uniformly distributed. The DC current supplied to the system is nonzero only when the source branch voltage exceeds 5 V, and then, it is linearly proportional to the voltage with a constant of proportionality a . Given that the system is working, compute the expected value of the power generated by the source.

Divide and Conquer Principle: Let A_1, A_2, ... A_n be disjoint partition of the sample space, with $\mathrm{P}(\{X = A_i\}) > 0$ for each i. We can find $f_X(x)$ by

$f_X(x) =$

and

$E(X) =$

as well as

$E(g(X)) =$

Ex: The metro train arrives at the station near your home every quarter hour starting at 6:00 a.m. You walk into the station every morning between 7:10 and 7:30, with your arrival time being random and uniformly distributed in this interval. What is the PDF of the time that you have to wait for the first train to arrive? Also find the expectation and variance of your waiting time.

3.5 Multiple Continuous Random Variables

Two random variables defined for the same sample space are said to be *jointly continuous* if there is a **joint probability density function** $f_{XY}(x, y)$ such that for any subset B of the two-dimensional plane,

$$\mathrm{P}((X, Y) \in B) = \int \int_{(x,y)\in B} f_{XY}(x, y)dxdy$$

When B is a rectangle:

When B is the entire real plane:

The joint pdf at a point can be approximately interpreted as the "probability per unit area" near the vicinity of that point. Just like the joint

PMF, the joint PDF contains all possible information about the individual random variables in consideration, and their dependencies. For example, the marginals are found as:

$f_X(x) =$

and
$f_Y(y) =$

Ex: Finding marginals from given two dimensional PDF: The joint PDF of the random variables X and Y is equal to a constant on the set S sketched on the board. Find the value of the constant c and the marginals of X and Y. Also compute the expectation of $X + 2Y$.

Ex: X and Y are "jointly uniform" in a circular region of radius r centered at the origin. Compute the marginal PDFs of X and Y, their expectations, and the expectation of the product XY.

3.5.1 Conditioning One R.V. on Another

Let X, Y be two r.v.s with joint PDF $f_{X,Y}(x, y)$. For any y with $f_Y(y) > 0$, the conditional PDF of X given that $Y = y$ is defined by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

It is best to view y as a fixed number and consider the conditional CDF $f_{X|Y}(x|y)$ as a function of the single variable x.

It may seem that the conditioning is on an event with zero probability which would contradict the definition of conditional probability. However, the PDF is not a probability. The conditional PDF $f_{X|Y}(a|b)$ describes the concentration of probability when X is within a small neighbourhood of a, given that Y is within a small neighbourhood of b.

$$\begin{aligned}
\mathrm{P}(a \leq X \leq a+\delta_1 | b \leq Y \leq b+\delta_2) & \\
&= \frac{\mathrm{P}(a \leq X \leq a+\delta_1, b \leq Y \leq b+\delta_2)}{\mathrm{P}(b \leq Y \leq b+\delta_2)} \\
&\approx \\
&=
\end{aligned}$$

Ex: In the example above where X and Y are jointly uniform on a circular region, compute the conditional PDF $f_{X|Y}(x|y)$.

Ex: The speed of Iron-Man flying past an air traffic radar is modelled as an exponentially distributed random variable with mean 200 km per hour. The radar sensor will measure the speed of any target with an additive error. The error is modelled as a zero-mean normal with a standard deviation equal to one tenth of the speed of the target. What is the joint PDF of the actual speed and the error?

Ex: Harry's magic wand breaks at a random point (location of the point is uniform along the length of the stick, which is 40 cm long). Suppose the piece of the stick that Harry is left with is X cm long. Unfortunately, the next day part of the stick is accidentally burnt while casting a spell. After this accident, the length of the stick is reduced to D, where D is uniformly distributed between $[0, X]$. Find $f_D(d)$.

3.5.2 Computing Conditional Expectation with PDFs

The conditional expectation of a continuous r.v. X is defined similar to its expectation. The only difference is that the conditional PDFs are used.

$$E[X|A] = \int \qquad\qquad E[X|Y=y] = \int$$

$$E[g(X)|A] = \int \qquad\qquad E[g(X)|Y=y] = \int$$

Ex: Consider a r.v. U which is uniform in $[0, 100]$. Find $E[U|B]$ where $B = \{U > 60\}$. Compare it to $E[U]$.

Total expectation theorem: the divide-and-conquer principle

$E[X] =$

$E[g(X)] =$

Ex: A coin is tossed 5 times. Knowing that the probability of heads is a r.v. P uniformly distributed in $[0.4, 0.7]$, find the expected value of the number of heads to be observed.

Independence for Continuous Random Variables

Definition 11 *In general, two random variables X and Y are called independent if for any events A and B,*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \tag{3.1}$$

where A and B are two arbitrary subsets real numbers.

- Suppose X and Y are continuous random variables, with $f_{XY}(x,y) = f_X(x)f_Y(y)$. Show that this condition implies the independence of X and Y.

Sometimes, observing that random variables are NOT independent is obvious from the region in which the joint PDF exists (recall the earlier joint PMF example.) However, proving that they ARE independent usually requires analytical operation (again, recall the discrete examples.)

Example:

$$f_{XY}(x,y) = \begin{cases} c & , 0 < x \leq 1, 0 < y \leq x \\ 0 & , o.w. \end{cases} \tag{3.2}$$

Ex: Show that each of the following conditions are equivalent to the definition of independence for continuous random variables X and Y:

$$f_{X|Y}(x|y) = f_X(x)$$
$$f_{Y|X}(y|x) = f_Y(y)$$

Ex: If two random variables X and Y are independent, then

$E(g(X)h(Y)) = E(g(X))E(h(Y))$ for any two functions $g(\cdot)$ and $h(\cdot)$.

Ex: As a consequence of the property above, $\text{var}(X+Y)= \text{var}(X)+ \text{var}(Y)$ when X and Y are independent. (Proof is identical to the proof of the discrete version, which we did before.)

Inference and Continuous Bayes's Rule

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)}$$

Ex: Consider the lightbulb with exponential lifetime Y, with parameter λ (failures per day). However, this time, λ is also random, and known to be uniformly distributed in $[1/4, 1]$. Having recorded the actual lifetime of a particular bulb as 4 days, what can we say about the distribution of λ?

Now, consider inference of the probability of an event based on the observation of the value of a random variable:

$$\begin{aligned} \mathrm{P}(A|Y = y) &= \\ &= \ldots \\ &= \frac{\mathrm{P}(A) f_{Y|A}(y)}{f_Y(y)} \end{aligned}$$

where the denominator can be expressed as

$$f_Y(y) = \mathrm{P}(A) f_{Y|A}(y) + \mathrm{P}(A^c) f_{Y|A^c}(y)$$

Ex: *Antipodal signaling under additive zero-mean unit-variance Gaussian noise.* Suppose that probabilities of sending a "1" or "-1" are p and $1-p$, respectively. What is the posterior probability that a "1" was sent, given that the noisy signal is measured as y. (Also consider the usual case where $p = 1/2$).

Chapter 4

Further Topics on Random Variables

4.1 Derived Distributions

Let $Y = g(X)$ be a function of a continuous random variable X. The general procedure for deriving the distribution of Y is as follows:

1. Calculate the CDF of Y:
$$F_Y(y) = \mathrm{P}(g(x) \leq y) = \int_{\{x:g(x)\leq y\}} f_X(x)dx$$

2. Differentiate to obtain the PDF of Y:
$$f_Y(y) = \frac{dF_Y}{dy}(y)$$

Ex: Find the distribution of $g(X) = \frac{180}{X}$ when $X \sim U[30, 60]$.

Ex: Find the PDF of $Y = g(X) = X^2$ in terms of the PDF of X, $f_X(x)$.

Ex: Show that, if $Y = aX + b$, where X has PDF $f_X(x)$, and $a \neq 0$ and b are scalars, $f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a})$. Note the following special case: $f_{-X}(x) = f_X(-x)$.

Ex: Show that a linear function of a Normal random variable is Normal. (exercise.)

4.1.1 Functions of Two Random Variables

Ex: Let X and Y be both uniformly distributed in $[0, 1]$ and independent. Let $Z = XY$. Find the PDF of Z.

Ex: Let X and Y be two independent discrete random variables. Express the PMF of $Z = X + Y$ in terms of the PMFs $p_X(x)$ and $p_Y(y)$ of X and Y. Do you recognize this expression?

Ex: Let X and Y be two independent continuous random variables. Show that, similarly to the discrete case, the PDF of $Z = X + Y$ is given by the "convolution" of the PDFs $f_X(x)$ and $f_Y(y)$ of X and Y.

Ex: Let X_1, X_2, X_3, ..., be a sequence of IID (independent, identically distributed) random variables, whose distribution is uniform in $[0, 1]$. Using convolution, compute and sketch the PDF of $X_1 + X_2$. As exercise, also compute and sketch the PDF of $X_1 + X_2 + \ldots + X_n$ for $n = 3, 4$, and observe the trend. As we add more and more random variables, the pdf of the sum is getting smoother and smoother. It turns out that, in the limit the shape of the density around the center will be converge to the Gaussian PDF. (It turns out that the Gaussian PDF is a fixed point for convolution: convolving two Gaussian PDFs results in another Gaussian PDF.)

4.2 Covariance and Correlation

Covariance of two random variables, X and Y, is defined as:

$$cov(X,Y) = E[(X - E(X))(Y - E(Y))]$$

It is a quantitative measure of the relationship between the two random variables: the magnitude reflects the strength of the relationship, the sign conveys the direction of the relationship. When $cov(X,Y) = 0$, the two random variables are said to be "uncorrelated". A positive correlation implies, roughly speaking, that they tend to increase or decrease together. A negative correlation, on the other hand, implies that when X increases, Y tends to decrease, and vice versa.

Ex: Show that $cov(X,Y) = E(XY) - E(X)E(Y)$ (as exercise.)

It follows from the alternative definition proven in the above exercise that independence implies uncorrelatedness.

Ex: Exhibit a counterexample showing that uncorrelatedness does not necessarily imply independence. (Consider X and Y uniformly distributed in an area shaped like a rhombus.)

The correlation coefficient is a normalized form of covariance:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

What is the possible range of values for $\rho(X,Y)$? The positive and negative extremes of this range correspond to full positive and full negative correlation, respectively. They are only attained when $X = aY$, where a

is a positive, or negative scalar, respectively.

Ex: Show that, for any two random variables X and Y

$$var(X+Y) = varX + varY + 2cov(X,Y)$$

Ex: Variance in the hat problem.

Ex: Let X_1, X_2, X_3, ..., be a sequence of IID Bernoulli(p) random variables (this "random sequence" is also called a "Bernoulli Process"). For concreteness, suppose that X_i stands for the result of the i^{th} trial in a sequence of independent trials, such that $X_i = 1$ if the trial is a success, and $X_i = 0$ if the trial does not result in a success.

1. Find $cov(X_i, X_j)$ for any i and j.

2. For every $i = 1, 2, \ldots$ if trial i is successful, we toss fair a coin. If the coin comes up H, we let $Y_i = 1$. Otherwise, the value of Y_i is set to zero. Find $cov(Y_i, Y_j)$ and $cov(X_i, Y_j)$ for all i and j.

3. Now, let $Z_i = 1$ whenever $X_i = 1$ and the coin toss comes up T. Find $cov(Y_i, Z_i)$. What is sign of the correlation coefficient for Y and Z?

4.3 Transforms

Transforms often provide us convenient ways with regard to certain mathematical manipulations.

The **transform**, in other words, the **moment generating function** (MGF), of a r.v. X is defined as

$$M_X(s) = E[e^{sX}].$$

Discrete case MGF: $M_X(s) = \sum_x e^{sx} p_X(x)$
Continuous case MGF: $M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x) dx$

Ex: MGF of a Poisson r.v.

Ex: MGF of an exponential r.v.

Ex: MGF of a linear function of a r.v.

Ex: MGF of a Gaussian random variable with mean μ and variance σ^2

Ex: Using the MGF of the Gaussian, and the property we just proved about the transform of a linear function of a random variable, show that a linear function of a Gaussian r.v. is also Gaussian.

4.3.1 Computation of Moments from Transforms

The name moment generating function follows from the following property.

$$E[X^n] = \frac{d^n}{ds^n} M_X(s) \,|_{s=0}$$

Proof:

Ex: Mean and variance of an exponential r.v.

Note: The transform $M_X(s)$ of a r.v. X uniquely determines the PDF of X. That is, one can always find $f_X(x)$ from $M_X(s)$.

4.3.2 Mixture of two distributions

Ex: The length in KBytes of IP packets received at a switch are, with 80 % probability, exponentially distributed with mean 10, and with 20 % probability, exponentially distributed with mean 100. Determine the MGF of the length of a randomly (uniformly) selected packet.

4.3.3 Sums of Independent R.V.s

When $X_1, X_2, \ldots, X_k$ are independent r.v.s, the MGF of their sum $Y = \sum_{i=1}^{k} X_i$ has a simple form. In deriving it, one may interpret e^{sX_i} as a function of X_i.

Ex: Sum of independent Poisson r.v.s

Ex: The sum of two independent Gaussian random variables is Gaussian. Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

Chapter 5

Limit Theorems

5.1 Markov and Chebychev Inequalities

Markov Inequality is a -typically loose- bound on the value of a non-negative random variable X with a known mean $E(X)$.

Markov Inequality: If a random variable X can take only non-negative values, then

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

Proof:

Ex: Let X be uniform in $[5, 10]$. Compute probabilities that X exceeds certain values and compare them with the bound given by Markov Inequality.

To be able to bound probabilities for general random variables (not necessarily positive), and to get a tighter bound, we can apply Markov Inequality to $(X - E(X))^2$ and obtain:

Chebychev Inequality: For random variable X with mean $E(X)$ and variance σ^2, and any real number $a > 0$,

$$\mathrm{P}(|X - E(X)| \geq a) \leq \frac{\sigma^2}{a^2}$$

Proof: Bound $\mathrm{P}((X - E(X))^2 \geq a^2$ using Markov Inequality.

Note that Chebychev's Inequality uses more information about X in order to provide a tighter bound about the probabilities related to X. In addition to the mean (a first-order statistic), it also uses the variance, which is a second-order statistic. You can easily imagine two very different random variables with the same mean: for example, a zero-mean Gaussian with variance 2, and a discrete random variable that takes on the values +0.1, and -0.1 equally probably. Markov Inequality does not distinguish between these distributions, where as Chebychev Inequality does.

It is possible to get better bounds that use more information- such as the Chernoff bound (but those are beyond the scope of this course.)

Ex: Use Chebychev's Inequality to lower-bound the probability that a Gaussian within two standard deviations of its mean.

5.2 Probabilistic Convergence

Let us remember the definition of convergence for a deterministic sequence of numbers from basic Calculus. Let $\{a_n\}$ be a sequence of numbers, indexed by n. $\lim_{n\to\infty} a_n = a$ means, for every $\epsilon > 0$, there exists an n_o such that for all $n > n_o$, $|a_n - a| < \epsilon$.

5.2.1 Convergence "In Probability"

Now, instead of a sequence of numbers, consider a sequence of random variables $\{Y_n\}$, indexed by n. Y_n converges in probability to a number a if for every $\epsilon > 0$, $\lim_{n\to\infty} P(|Y_n - a| > \epsilon) = 0$. The notation is: $Y_n \stackrel{i.p}{\rightarrow} a$.

Ex: Suppose for each n, Y_n takes the value n with probability $1/n$, and the value zero with probability $1 - 1/n$. Does $\{Y_n\}$ converge, and if so, to what?

In proving "convergence in probability" we often use Chebychev's Inequality, as in the following example.

Ex: Flip a fair coin n times, independently. Let Y_n be equal to the number of heads minus $n/2$. Does $\frac{Y_n}{n}$ converge?

The following result is a generalization of the previous example.

5.3 The Weak Law of Large Numbers

The Weak Law of Large Numbers (WLLN) is an important special case of convergence in probability. Consider $X_1, X_2, \ldots$ IID, with $E(X_i) = \mu$, and $\text{var}X_i = \sigma^2 < \infty$ for all i. The sample mean sequence $M_n = \frac{X_1+X_2+\ldots+X_n}{n}$ converges to μ in probability.

Proof: (Use the Chebychev Inequality on M_n.)

Ex: Polling: We want to estimate the fraction of the population that will vote for XYZ. Let X_i be equal to 1 if the i^{th} person votes in favor of XYZ, and 0 otherwise. How many people should we poll, to make sure our error will be less than 0.01 with 95% probability? (Answer: with Chebychev Inequality, we get n=50,000. However, this is too conservative. Using the Central Limit Theorem, we will get that a poll over a much smaller number of people will suffice.)

5.4 The Central Limit Theorem

Let $X_1, X_2, \ldots$ be a sequence of IID random variables with $E(X_i) = \mu$, and $\mathrm{var} X_i = \sigma^2 < \infty$, and define

$$Z_n = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then, the **CDF** of Z_n converges to the standard normal CDF

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

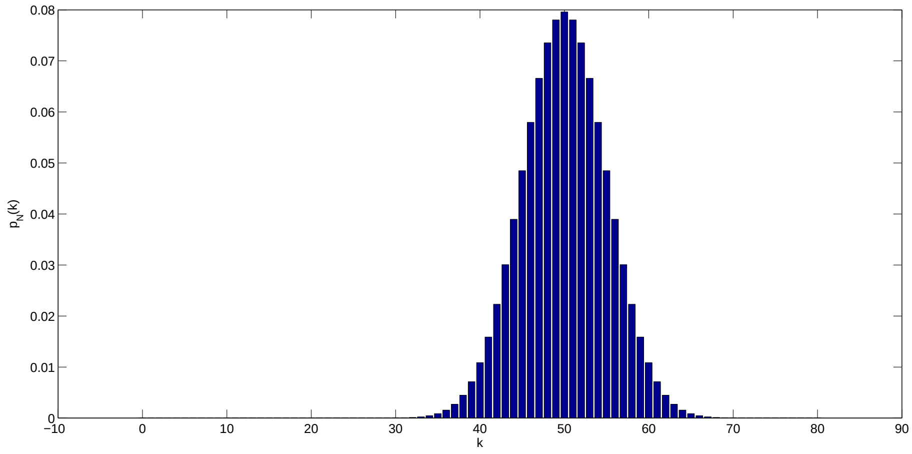
in the sense that

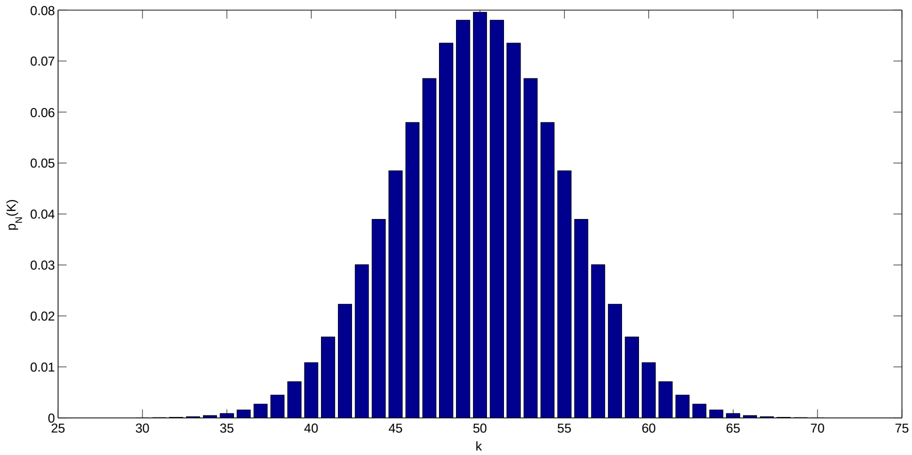
$$\lim_{n\to\infty} \mathrm{P}(Z_n \leq z) = \Phi(z)$$

for every z.

Let's see a very simple application of this that illustrates many of the issues.

Ex: Toss a fair coin 100 times. Find an approximation for the number of H's being between 45 and 55 (inclusive).


0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
$p_N(k)$
−10
0
10
20
30
40
50
60
70
80
90
k


0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
$p_N(K)$
25
30
35
40
45
50
55
60
65
70
75
k

Simple Proof of the Central Limit Theorem: We will use Transforms. For simplicity, assume $E(X) = 0$, $\sigma = 1$. Consider the transform of Z_n, $\Phi_{Z_n}(s)$.

$$\begin{aligned}
\Phi_{Z_n}(s) &= E[e^{s\frac{X_1+X_2+\ldots+X_n}{\sqrt{n}}}] \\
&= E[e^{s\frac{X_1}{\sqrt{n}}}e^{s\frac{X_2}{\sqrt{n}}}\ldots e^{s\frac{X_n}{\sqrt{n}}}] \\
&= \prod_{i=1}^{n} E[e^{s\frac{X_i}{\sqrt{n}}}] \\
&= (E[e^{s\frac{X_i}{\sqrt{n}}}])^n \\
&= (E[1 + \frac{sX_i}{\sqrt{n}} + \frac{(sX_i)^2}{2n}\ldots])^n (\text{ using Taylor Series expansion}) \\
&= (E[1 + \frac{(sX_i)^2}{2n}\ldots])^n (\text{ zero mean}) \\
&\rightarrow e^{s^2/2}
\end{aligned}$$

Since the transform of Φ_{Z_n} converges to that of a standard Gaussian, and there is a one-to-one mapping between transforms and distributions, we conclude that Φ_{Z_n} converges to a standard Gaussian.

There are other versions of the central limit theorem (CLT), but this basic version assumes that the random variables are independent, identically distributed, with finite mean and variance. Aside from these, there is no requirement on the distribution of X, it can be continuous, discrete, or mixed. One thing to be careful about while using the CLT is that this is a "central" property, i.e. it works well around the mean of the distribution, but not necessarily at the tails (consider, for example, X_i being strictly positive and discrete- then Z_n has absolutely no probability mass below zero, whereas the Gaussian distribution does.)

We shall end this topic with an application of the CLT to the polling problem.

Ex: Revisit the polling problem. Find approximately how large n should be such that the probability that the poll result differs from the true voter fraction in absolute value by more than 0.01 is less than 5 percent. (Answer: $n \geq 10,000$ will suffice.)

Chapter 6

The Bernoulli and Poisson Processes

6.1 The Bernoulli Process

By this time in the course, we have actually covered all the background for the Bernoulli Process under the disguise of "Bernoulli trials". All that remains is to introduce it as a *random process* (in other words, a stochastic process).

A discrete stochastic process is a sequence of random variables, $\{X_i\}$, indexed by i. Typically, $i = 1, 2, \ldots$.

The Bernoulli process with rate p is a sequence of IID Bernoulli random variables with parameter p:

$$X_i = \begin{cases} 1 & \text{, with probability p} \\ 0 & \text{, with probability } 1 - \text{p} \end{cases}$$

You can think of this as the result of a sequence of independent tosses of a coin with bias p. Thus it is easy to see that it has the following properties.

6.1.1 Properties

1. Binomial sums: Let S be the number of 1's among $X_1, X_2, \ldots, X_n$ is Binomial(n, p) (number of successes in n trials). In other words, $S = \sum_{i=1}^{n} X_i$. Then, $E(S) = np$, $\mathrm{var}(S) = np(1-p)$ and $p_S(k) =$, $k = 0, 1, \ldots, n$.

2. Geometric first arrival time: Let T be the time of the first success, $T \geq 1$. Then, $E(T) = \frac{1}{p}$, $\mathrm{var}(T) = \frac{1-p}{p^2}$ and $p_T(t) = (1-p)^{t-1}p$, $t = 1, 2, \ldots$.

3. Fresh-start: For any given time i, the sequence of random variables $X_{i+1}, X_{i+2}, \ldots$ (the future of the process) is also a Bernoulli process, and is independent of $X_1, X_2, \ldots, X_n$ (the past.)

4. Memorylessness and Geometric Inter-arrivals: Let $i + T$ be the time of the first success *after* time i, $T \geq 1$. Then, $E(T) = \frac{1}{p}$, $\mathrm{var}(T) = \frac{1-p}{p^2}$ and $p_T(t) = (1-p)^{t-1}p$, $t = 1, 2, \ldots$.

Ex: Ayse and Burak are playing a computer game. In each round of the game, Ayse wins with probability p, and otherwise, Burak wins. Find the PMF of the number of games won by Ayse between any two games won by Burak.

Ex: Let Y_k be the total number of games played up to and including the k^{th} game won by Ayse. Find the mean, variance and the PMF of Y_k. (This is called the Pascal PMF of order k with parameter p.)

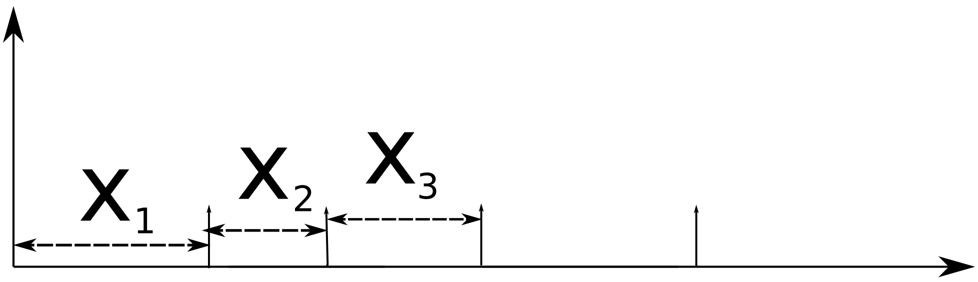
6.1.2 Splitting and Merging the Bernoulli Process

Ex: Show that when we split each arrival of a Bernoulli(p) process independently with probability r, the resulting subprocesses are Bernoulli with rates rp and $(1-r)p$. Are the resulting processes independent?

Ex: Show that when we merge two INDEPENDENT Bernoulli processes of rates p_1 and p_2, the resulting process is Bernoulli with rate $p_1 + p_2 - p_1 p_2$.

6.2 The Poisson Process

Consider arrivals (of busses, customers, photons, e-mails, etc) occurring at random points in time. We say that the arrival process is a Poisson Process if the times between arrivals are IID, Exponential random variables.



More precisely, let X_1, X_2, X_3, be the sequence of inter-arrival times as shown in the figure. The process is a Poisson process of rate λ if the $\{X_i\}$, $i \geq 1$ are independent and Exponential with rate λ. Note that the mean time between two arrivals is $\frac{1}{\lambda}$.

Ex: I am waiting for the bus, and bus arrivals are known to be a Poisson process at rate 1 bus per 10 minutes. Starting at time $t = 0$, what is the expected arrival time of the third bus?

Distribution of residual time: At an arbitrary time $t > 0$, let R be the duration until the next arrival. This is called the "residual time" because it is only part of the inter-arrival time that t falls into. It is easy to show that R has the same distribution as a regular inter-arrival time. This is a consequence of the Exponential being "memoryless". It also implies that the Poisson process has the "fresh-start" property.

Ex: Given that I arrive at the bus-stop at $t = 19$ and learn that I have missed the second bus by two minutes, how much do I expect to wait?

Equivalent Definition of the Poisson Process:

An arrival process that satisfies the following is a Poisson process.

1. The probability $\mathrm{P}(k, \tau)$ that there are k arrivals in any time interval of size τ is given by:

$$P(k, \tau) = \frac{(\lambda \tau)^k e^{-\lambda \tau}}{k!}.$$

   Note that this is the Poisson PMF, with mean $\lambda \tau$.

2. The numbers of arrivals in disjoint intervals are independent.

Ex: I get email according to a Poisson process at rate $\lambda = 0.1$ arrivals per minute. If I check my email every hour, what is the expected number of new messages I find in my inbox when I check my email? What is the probability that I find no messages? One message? Repeat for an e-mail checking period of two hours.

The time-reversed process is also Poisson: We can show that the reverse residual time distribution is the same as the inter-arrival time distribution.

Ex: In the bus problem, what is the expected number of people on the bus that I get on? (Hint: Consider the people that arrived in the two minutes before I arrived, as well as the people that arrive while I am waiting.)

The random incidence "paradox": When I arrive at random, the interval of time I arrive in has twice the expectation of a regular inter-arrival time. Recall the difference of interviewing bus drivers versus passengers, to understand how crowded a bus is on average.

Relationship to the Bernoulli process: Take a Poisson process at rate λ and discretize time finely, in chunks of size δ. Show that as $\delta \to 0$, the Poisson process can be approximated by a Bernoulli process.

The "Baby Bernoulli" definition of the Poisson process: We can equivalently define a Poisson process as a process where the probability of arrival in any time interval of size δ is $\lambda\delta + o(\delta)$, the probability of more than one arrival is, $o(\delta)$, and arrivals in disjoint intervals are independent.

6.2.1 Splitting and Merging Poisson Processes

Ex: Show that, when we send each arrival of a Poisson process at rate λ to a process A with probability p, and process B with probability $1-p$, the resulting processes A and B are Poisson with rates $p\lambda$ and $(1-p)\lambda$. Note also that processes A and B are independent of each other (this is unlike the Bernoulli case, where we can easily show that the split processes are not independent.) (Hint: Express the transform of the interarrival time as a geometric sum of exponentials.)

Ex: Show that when we merge two INDEPENDENT Poisson processes at rates λ_a and λ_b, we get a Poisson process at rate $\lambda_a + \lambda_b$. (Hint: consider two lightbulbs with exponential lifetimes running side by side. Find the distribution of the time that the first one burns out.)