Lecture Notes

EE230 Probability and Random Variables

Department of Electrical and Electronics Engineering
Middle East Technical University (METU)
Preface

These lecture notes were prepared with the purpose of helping the students to follow the lectures more easily and efficiently. This course is a fast-paced course (like many courses in the department) with a significant amount of material, and to cover all of this material at a reasonable pace in the lectures, we intend to benefit from these partially-complete lecture notes. In particular, we included important results, properties, comments and examples, but left out most of the mathematics, derivations and solutions of examples, which we do on the board and expect the students to write into the provided empty spaces in the notes. We hope that this approach will reduce the note-taking burden on the students and will enable more time to stress important concepts and discuss more examples.

These lecture notes were prepared mainly from our textbook titled "Introduction to Probability" by Dimitry P. Bertsekas and John N. Tsitsiklis, by revising the notes prepared earlier by Elif Uysal-Biyikoglu and A. Ozgur Yilmaz. Notes of A. Aydin Alatan and discussions with fellow lecturers Elif Vural, S. Figen Oktem and Emre Ozkan were also helpful in preparing these notes.

This is the first version of the notes. Therefore the notes may contain errors and we also believe there is room for improving the notes in many aspects. In this regard, we are open to feedback and comments, especially from the students taking the course.

Fatih Kamisli
May 19th, 2017.
Contents

1 Sample Space and Probability ........................................... 5
   1.1 Set Theory ..................................................................... 5
      1.1.1 Set Operations ....................................................... 6
      1.1.2 Properties of Sets and Operations ......................... 7
   1.2 Probabilistic Models .................................................... 8
      1.2.1 Random Experiment ............................................ 8
      1.2.2 Sample Space ..................................................... 8
      1.2.3 Probability Law .................................................. 10
      1.2.4 Properties of Probability Laws .............................. 11
      1.2.5 Discrete Probabilistic Models ................................. 12
      1.2.6 Continuous Probabilistic Models ......................... 13
   1.3 Conditional Probability ............................................... 14
      1.3.1 Properties of Conditional Probability .................. 14
      1.3.2 Chain (Multiplication) Rule ................................. 16
   1.4 Total Probability Theorem and Bayes’ Rule .................. 17
   1.5 Independence .......................................................... 20
      1.5.1 Properties of Independent Events ...................... 21
      1.5.2 Conditional Independence ................................. 21
      1.5.3 Independence of a Collection of Events ................ 22
      1.5.4 Independent Trials and Binomial Probabilities .......... 24
   1.6 Counting ..................................................................... 26

2 Discrete Random Variables .............................................. 28
   2.1 Basics ....................................................................... 29
   2.2 Probability Mass Functions (PMF) ............................... 30
      2.2.1 How to calculate the PMF of a random variable X .... 30
      2.2.2 Properties of PMF .............................................. 31
   2.3 Some Discrete Random Variables ................................. 31
      2.3.1 The Bernoulli R.V. ............................................ 31
      2.3.2 The Binomial R.V. ............................................ 31
      2.3.3 The Geometric R.V. .......................................... 32
      2.3.4 The Poisson R.V. ............................................. 33
      2.3.5 The Uniform R.V. ............................................ 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>Functions of Random Variables</td>
<td>33</td>
</tr>
<tr>
<td>2.4.1</td>
<td>How to obtain $p_Y(y)$ given $Y = g(X)$ and $p_X(x)$?</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>Expectation, Mean, and Variance</td>
<td>34</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Variance, Moments, and the Expected Value Rule</td>
<td>35</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Properties of Mean and Variance</td>
<td>36</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Mean and Variance of Some Common R.V.s</td>
<td>37</td>
</tr>
<tr>
<td>2.6</td>
<td>Multiple Random Variables and their Joint PMFs</td>
<td>38</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Functions of Multiple Random Variables</td>
<td>40</td>
</tr>
<tr>
<td>2.6.2</td>
<td>More Than Two Random Variables</td>
<td>41</td>
</tr>
<tr>
<td>2.7</td>
<td>Conditioning</td>
<td>42</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Conditioning a random variable on an event</td>
<td>42</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Conditioning one random variable on another</td>
<td>43</td>
</tr>
<tr>
<td>2.7.3</td>
<td>Conditional Expectation</td>
<td>47</td>
</tr>
<tr>
<td>2.8</td>
<td>Independence</td>
<td>49</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Independence of a Random Variable from an Event</td>
<td>49</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Independence of Random Variables</td>
<td>50</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Variance of the Sum of Independent Random Variables</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>General Random Variables</td>
<td>53</td>
</tr>
<tr>
<td>3.1</td>
<td>Continuous Random Variables and PDFs</td>
<td>54</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Some Continuous Random Variables and Their PDFs</td>
<td>56</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Expectation</td>
<td>57</td>
</tr>
<tr>
<td>3.2</td>
<td>Cumulative Distribution Functions</td>
<td>59</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Properties of CDF</td>
<td>60</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Hybrid Random Variables</td>
<td>62</td>
</tr>
<tr>
<td>3.3</td>
<td>Normal (Gaussian) Random Variables</td>
<td>63</td>
</tr>
<tr>
<td>3.3.1</td>
<td>The Standard Normal R.V.</td>
<td>63</td>
</tr>
<tr>
<td>3.4</td>
<td>Multiple Continuous Random Variables</td>
<td>66</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Joint CDFs, Mean, More than Two R.V.s</td>
<td>68</td>
</tr>
<tr>
<td>3.5</td>
<td>Conditioning</td>
<td>69</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Conditioning One R.V. on Another</td>
<td>71</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Conditional CDF</td>
<td>73</td>
</tr>
<tr>
<td>3.5.3</td>
<td>Conditional Expectation</td>
<td>74</td>
</tr>
<tr>
<td>3.6</td>
<td>Independent Random Variables</td>
<td>75</td>
</tr>
<tr>
<td>3.7</td>
<td>The Continuous Bayes’ Rule</td>
<td>76</td>
</tr>
<tr>
<td>3.7.1</td>
<td>Inference about a discrete random variable</td>
<td>77</td>
</tr>
<tr>
<td>3.7.2</td>
<td>Inference based on discrete observations</td>
<td>78</td>
</tr>
<tr>
<td>3.8</td>
<td>Some Problems</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>Further Topics on Random Variables</td>
<td>80</td>
</tr>
<tr>
<td>4.1</td>
<td>Derived Distributions</td>
<td>80</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Finding $f_Y(y)$ directly from $f_X(x)$</td>
<td>82</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Functions of Two Random Variables</td>
<td>84</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>4.2 Covariance and Correlation</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>4.3 Conditional Expectations Revisited (Iterated Expectations)</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>4.4 Transforms (Moment Generating Functions)</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>4.4.1 From Transforms to Moments</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>4.4.2 Inversion of Transforms</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>4.4.3 Sums of Independent R.V.s</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>4.4.4 Transforms Associated with Joint Distributions</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>4.5 Sum of a Random Number of Independent R.V.s</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>5 Limit Theorems</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>5.1 Markov and Chebychev Inequalities</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>5.2 The Weak Law of Large Numbers (WLLN)</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>5.3 Convergence in Probability</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>5.4 The Central Limit Theorem (CLT)</td>
<td>101</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 1

Sample Space and Probability

Contents

1.1 Set Theory ...................................................... 5
   1.1.1 Set Operations ............................................ 6
   1.1.2 Properties of Sets and Operations ....................... 7
1.2 Probabilistic Models ........................................... 8
   1.2.1 Random Experiment ....................................... 8
   1.2.2 Sample Space ............................................... 8
   1.2.3 Probability Law ........................................... 10
   1.2.4 Properties of Probability Laws ......................... 11
   1.2.5 Discrete Probabilistic Models ......................... 12
   1.2.6 Continuous Probabilistic Models ..................... 13
1.3 Conditional Probability ....................................... 14
   1.3.1 Properties of Conditional Probability ................ 14
   1.3.2 Chain (Multiplication) Rule ......................... 16
1.4 Total Probability Theorem and Bayes’ Rule ................. 17
1.5 Independence .................................................... 20
   1.5.1 Properties of Independent Events .................... 21
   1.5.2 Conditional Independence ............................ 21
   1.5.3 Independence of a Collection of Events ............ 22
   1.5.4 Independent Trials and Binomial Probabilities ...... 24
1.6 Counting .......................................................... 26

1.1 Set Theory

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

Ex: Some representations
Element of a set (and not element of a set)

Null set=empty set=∅={}. 

Quantity classification of sets

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

Some relations

- A is a subset of B (A ⊂ 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.

1.1.1 Set Operations

1. Union

2. Intersection

3. Complement of a set
4. Difference

5. Some definitions related to set operations
   - Disjoint sets
     - Two sets $A$ and $B$ are called disjoint (or mutually exclusive) if $A \cap B = \emptyset$.
     - More generally, multiple sets are called disjoint if no two of them have a common element.

   - Partition of a set
     - 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 $S$.

1.1.2 Properties of Sets and Operations
   - Commutative: $A \cup B = B \cup A$, $A \cap B = B \cap A$
   - Associative: $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$
   - Distributive: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
   - $A \cap \emptyset = A \cup \emptyset = \emptyset^c = \Omega^c = A^c = A \cap \Omega = A \cup \Omega = A \cap \Omega$
   - De Morgan’s laws
     - $(A \cup B)^c =$
     - $(A \cap B)^c =$
   - 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:
• Size (cardinality) of a set: The number of elements in the set, shown for a set $A$ as $|A|$. 
  Ex:

• The power set $\mathcal{P}(A)$ of a set $A$: the set of all subsets of $A$. 
  Ex:

1.2 Probabilistic Models

Definition 2 A probabilistic model is a mathematical description of an uncertain situation.

A probabilistic model consists of three components:

1. Random experiment
2. Sample space
3. Probability law.

1.2.1 Random Experiment

Definition 3 Every probabilistic model involves an underlying process called the random experiment, that will produce exactly one out of several possible outcomes.

Ex:

1.2.2 Sample Space

Definition 4 The set of all possible results (or outcomes) of a random experiment is called the sample space ($\Omega$) of the experiment.
Definition 5 A subset of the sample space $\Omega$ (i.e. a collection of possible outcomes) is called an event.

Ex: In the above example where we toss a coin 3 times, define event $A$ as the event that exactly two H’s occur.

Some definitions

- $\Omega$: certain event, $\emptyset$: impossible event
- An event with a single element is called a singleton (or elementary event).
- Trial is a single performance of an experiment
- An event $A$ is said to have occurred if the outcome of the trial is in $A$.

Choosing an appropriate sample space

- The sample space should have enough detail to distinguish between all outcomes of interest while avoiding irrelevant details for the given problem.
  
  Ex: Consider two games both involving 5 successive coin tosses.
  
  Game 1: We get $1 for each H.
  
  Game 2: We get $1 for each toss upto first H. Then we receive $2 for each toss upto second H and so on. In particular, the received amount per toss doubles each time a H comes up. We are interested in the amount of money we win in both games. In Game1, only total number of H’s is important, while in Game2 knowing the tosses in which H occur is also important.
Sequential models

Many experiments have sequential character. For example:

- Tossing a coin 3 times
- Receiving 5 successive bits at a communication receiver

Tree-based sequential descriptions are very useful for such experiments to describe the experiment and the associated sample space.

Ex: Two rolls of a 4 sided die

1.2.3 Probability Law

**Definition 6** The probability law assigns to every event $A$ a nonnegative number $P(A)$ called the probability of event $A$.

- Intuitively, the probability law specifies the "likelihood" of any outcome or any event.
- $P(A)$ is often intended as a model for the frequency with which the experiment produces a value in $A$ when repeated many times independently.

Ex: We have a biased coin with $P("H") = 0.6$. Throw the coin 100 times. How many "Heads" do you expect to observe?

**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) $P(\Omega) = 1$
To visualize the probability law, consider a mass of 1, which is spread over the sample space. Then, $P(A)$ is the total mass that was assigned to the elements of $A$.

**Ex:** Single toss a fair coin.

### 1.2.4 Properties of Probability Laws

Many properties of a probability law can be derived from the axioms.

(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) $A \subset B \implies P(A) \leq P(B)$

(e) $P(A \cup B) \leq P(A) + P(B)$  
(More generally, $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$)
(f) \( P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C) \)

### 1.2.5 Discrete Probabilistic Models

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

For discrete probabilistic models:

- The probability of any event \( A = \{s_1, s_2, \ldots, s_k\} \) is the sum of the probabilities of its elements.

\[
P(A) = P(\{s_1, s_2, \ldots, s_k\}) = P(\{s_1\}) + P(\{s_2\}) + \ldots + P(\{s_k\})
\]

- Discrete uniform probability law: The sample space consists of \( |\Omega| = n \) possible outcomes which are equally likely. Then the probability of any event \( A = \{s_1, s_2, \ldots, s_k\} \) can be obtained by counting the elements in \( A \) and \( \Omega \):

\[
P(A) = \frac{|A|}{|\Omega|}.
\]

**Ex:** Roll a pair of 4-sided die. Let us assume the die are fair and therefore each possible outcome is equally likely.

**Ex:** A box contains 100 light bulbs. The probability that there exists at least one defective bulb is 0.1. The probability that there exist at least two defective bulbs is 0.05. Let the number of defective bulbs be the outcome of the random experiment.

(a) Define the sample space.
(b) Find $P(\text{no defective bulbs})$.

(c) $P(\text{exactly one defective bulb})=?$

(d) $P(\text{at most one defective bulb})=?$

1.2.6 Continuous Probabilistic Models

The sample space $\Omega$ is an uncountable set in continuous models.

**Ex:** The angle of an arbitrarily drawn line.

Uniform probability law in continuous models is discussed in the example below.

**Ex:** Romeo and Juliet arrive equiprobably between 0 and 1 hour. The first one arrives and waits for $\frac{1}{4}$ hours and leaves. What is the probability that they meet?
1.3 Conditional Probability

Conditional probabilities reflect our beliefs about events based on partial information about the outcome of the random experiment.

Ex: There are 200 light bulbs in a box, where each bulb is either a 50, 100 or 150 Watt bulb, and each bulb is either defective or good. (A table indicating the number of each type of bulb is given.) Let’s choose a bulb at random from the box.

Definition 7 Let $A$ and $B$ two events with $P(A) \neq 0$. The conditional probability $P(B|A)$ of $B$ given $A$ is defined as

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

1.3.1 Properties of Conditional Probability

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

Proof 1
• If $A$ and $B$ are disjoint, $P(B|A) = .$

• If $A \subset B$, $P(B|A) = .$

• If $P(A) = 0$, leave $P(B|A)$ undefined.

• When all outcomes are equally likely, $P(B|A) =$

• Since conditional probabilities form a legitimate probability law, general properties of probability law remain valid. For example

  \[ P(B) + P(B^c) = 1 \implies P(B|A) + P(B^c|A) = 1. \]

  \[ P(B \cup C) \leq P(B) + P(C) \implies P(B \cap C|A) \leq P(B|A) + P(C|A) \]

\[ P(B|A)P(A) = P(A|B)P(B) \]

**Ex:** A fair coin is tossed three successive times. Two events are defined as $A = \{\text{more H than T come up}\}, B = \{1^{st} \text{ toss is a H}\}$. Determine the sample space and find $P(A|B)$.

**Ex:** An urn has balls numbered with 1 through 5. A ball is drawn and not put back. A second ball is drawn afterwards. What is the probability that the second ball drawn has an odd number given that

(a) the first ball had an odd number on it?

(b) the first ball had an even number on it?

**Ex:** Two cards are drawn form a deck in succession without replacement. What is the probability that first card is spades and second card is clubs?
Ex: Radar detection problem. If an aircraft is present in a certain area, a radar detects it and generates an 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 not present with probability 0.95.

(a) What is the probability of no aircraft presence and a false alarm?

(b) What is the probability of aircraft presence and no detection?

1.3.2 Chain (Multiplication) Rule

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

\[ P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|\bigcap_{i=1}^{n-1} A_i). \]

The multiplication rule follows from the definition of conditional probability.

Tree-based representation is also useful for the multiplication rule.
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 were all numbered with 1?

Ex: The Monty Hall problem. Look at Example 1.12 (pg.27) from textbook or look at this video on youtube.

1.4 Total Probability Theorem and Bayes’ Rule

Theorem 2 (Total Probability) Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space and assume that $P(A_i) > 0$ for all $i$. Then, for any event $B$, we have

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \ldots + P(B \cap A_n)$$
$$= P(B | A_1)P(A_1) + P(B | A_2)P(A_2) + \ldots + P(B | A_n)P(A_n).$$

Proof 2
• Special case, \( n = 2 \): \( P(B) = P(B|A)P(A) + \)

Ex: A box of 250 transistors all equal in size, shape etc. 100 of them are manufactured by \( A \), 100 by \( B \), and the rest by \( C \). The transistors from \( A \), \( B \), \( C \) are defective by 5, 10, 25\%, respectively.

(a) A transistor is drawn at random. What is the probability that it is defective?
(b) Find the probability that it is made by \( A \) given that it is defective.

Theorem 3 (Bayes’ Rule) Let \( A_1, \ldots, A_n \) be disjoint events that form a partition of the sample space and assume that \( P(A_i) > 0 \) for all \( i \). Then, for any event \( B \) with \( P(B) > 0 \), we have

\[
P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \ldots + P(B|A_n)P(A_n)}.
\]

Proof 3

By the definition of conditional probability, we have

Ex: A coin is tossed and a ball is drawn. If the coin is heads, the ball is drawn from Box H with 3 red and 2 black balls. If the coin is tails, Box T is used which has 2 red and 8 black balls.

(a) Find \( P(\text{red}) \).
(b) \( P(\text{black}) = ? \)
(c) If a red balls is drawn, what is the probability that a heads is thrown?
(d) Find \( P(T|\text{red}) \) using Bayes’ theorem.
**Ex:** Radar detection problem revisited. $A = \{\text{an aircraft is present}\}$, $B = \{\text{the radar generates an alarm}\}$. $P(A) = 0.05$, $P(B|A) = 0.99$, $P(B|A^c) = 0.1$. What is the probability that an aircraft is actually present if the radar generates an alarm?

**Ex:** 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?

**The conditional version of the Total Probability Theorem:**

Let $A_1, \ldots, A_n$ be disjoint events that form a partition of the sample space and the conditional probabilities used below be defined. Then

$$P(B|C) = \sum_{i=1}^{n} P(B|C \cap A_i)P(A_i|C).$$

Pf. (see Problem 28 in textbook.)
1.5 Independence

Conditional probability $P(A|B)$ captures partial information that event $B$ provides about event $A$.

- In some cases, the occurrence of event $B$ may provide no such information and not change the probability of $A$, i.e. $P(A|B) = P(A)$.
- When the above equality holds, we say $A$ is independent of $B$.
- Note that $P(A|B) = P(A)$ is equivalent to $P(A \cap B) = P(A)P(B)$

**Definition 8** Two events $A$ and $B$ are independent if and only if $P(A \cap B) = P(A)P(B)$.

**Ex:** Toss 2 fair coins. Event $A$: first toss is H. Event $B$: second toss is T. Are $A$, $B$ independent?

**Notes:**

- Independence is often easy to grasp intuitively. If the occurrence of two events $A, B$ are actuated by distinct and noninteracting physical processes, then the events are independent.
- Independence is not easily visualized in terms of sample space.

**Ex:** Toss a fair coin and roll a fair 6-sided die. Find the probability that the coin toss come up Heads and the die roll results in 5.

**Ex:** An urn has 4 balls where two of them are red and the others green. Each color has a ball numbered with 1, and the other is numbered with 2. A ball is randomly drawn. $A$: event that the ball drawn is red. $B$: event that the ball drawn has an odd number on it.

(a) Are $A$ and $B$ independent?
(b) Add another red ball with 3 on it. Are $A$ and $B$ independent now?

### 1.5.1 Properties of Independent Events

- 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 $A$.)

- If $A$ and $B$ are disjoint with $P(A) > 0$ and $P(B) > 0$, they are always dependent.

- For independent events $A$ and $B$, we have $P(A \cup B) =$

### 1.5.2 Conditional Independence

Recall that conditional probabilities form a legitimate probability law. We can then talk about various events with respect to this probability law.

**Definition 9** Given an event $C$, events $A$ and $B$ are **conditionally independent** if and only if

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

Conditioning may affect/change independence, as shown in the following example.

**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 \( A \) and \( B \) independent? Are they independent given \( C \)?

**Ex:** We have two unfair coins, A and B with \( P(H|\text{coin } A) = 0.9, P(H|\text{coin } B) = 0.1 \). We choose either coin with equal probability and perform independent tosses with the chosen coin.

(a) Once we know coin A is chosen, are future tosses independent?

(b) If we don’t know which coin is chosen, are future tosses independent? Compare \( P(2\text{nd toss is a T}) \) and \( P(2\text{nd toss is a T}|\text{first toss is T}) \).

### 1.5.3 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

\[
P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i) \text{ for every subset } S \text{ of } \{1, 2, \ldots, n\}.
\]

- For 3 events, we have the following conditions for independence:
• How many equations do you need to check to verify independence of 4 events?

Ex: Consider two independent tosses of a fair coin. \( A = \{\text{First toss is H}\} \), \( B = \{\text{Second toss is H}\} \), \( C = \{\text{Tosses have same outcomes}\} \). Are these events independent?

As seen in the example above,

• Pairwise independence does not imply independence! (Checking \( P(A_i \cap A_j) = P(A_i)P(A_j) \) for all \( i \) and \( j \) is not sufficient for confirming independence!)

• For three events, checking \( P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3) \) is not enough for confirming independence!

**Intuition behind independence of a collection of events** is similar to the case of two events. Independence means that the occurrence or non-occurrence of any number of events from the collection carries no information about the remaining events or their complements. For example, **if events** \( A_1, A_2, \ldots, A_6 \) **are independent, then**

• \( A_1 \) independent of \( A_2 \cup A_3 \)

• \( P((A_5 \cup A_2) \cap (A_1 \cup A_4)^c | A_3 \cup A_6^c) = P((A_5 \cup A_2) \cap (A_1 \cup A_4)^c) \)

Ex: (Network Connectivity) A computer networks connects nodes A and B via intermediate nodes C,D,E. For every pair of directly connected nodes (e.g. nodes i and j), there is a probability \( p_{i,j} \) that the link from i to j is up/working. Assuming link failures are independent of each other, what is the probability that there is a path from A to B in which all links are up?
1.5.4 Independent Trials and Binomial Probabilities

**Independent trials**: A random experiment that involves a sequence of independent but identical stages (e.g. 5 successive rolls of a die.)

**Bernoulli trials**: Independent trials where each identical stage has only two possible outcomes (e.g. 5 tosses of a coin.) The two possible outcomes can be anything (e.g. Yes or No, Male or female,...) but we will often think in terms of coin tosses and refer to the two results as H and T.

**Consider the Bernoulli trials** where we have \( n \) tosses of a biased coin with \( P(H) = p \). For this experiment, we have the following important result :

\[
P(k \text{ H's in an } n\text{-toss sequence}) =
\]

To obtain this results, let us consider the case where \( n = 3 \) and \( k = 2 \), and visualize the Bernoulli trials by means of a sequential description :
Let us now generalize to the case with \( n \) tosses:

- Probability of any such given sequence with \( k \) H’s in \( n \) tosses:
- Number of such \( n \)-toss sequences with \( k \) H’s and \( n - k \) T’s:
  (known as binomial coefficient)
- Overall: \( P(k \) H’s in an \( n \)-toss sequence) = \( \binom{n}{k} p^k (1 - p)^{n-k} \)
  (known as binomial probability.)

**Ex:** A fair die is rolled four times independently. Find the probability that ’3’ will show up only twice.

**Ex:** Binary symmetric channel with independent errors.

(a) What is the probability that \( k^{th} \) symbol is received correctly?

(b) What is the probability that a sequence 1111 is received correctly?

(c) Given that a 100 is sent, what is the probability that 110 is received?

(d) 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?

(e) Suppose that repetition encoding is used. What is the probability that 0 is sent given that 101 is received?
Ex: An Internet service provider (ISP) has $c$ modems to serve $n$ customers. At any time, each customer needs a connection with probability $p$, independent of others. What is the probability that there are more customers needing a connection than there are modems?

1.6 Counting

In random experiments with finite number of possible outcomes that are all equally likely, finding probabilities reduces to counting the number of outcomes in events and the sample space.

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

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

The counting principle: Based on divide-and-conquer approach where counting is performed by multiplying the number of possibilities in each stage.

Ex: Cities and roads.

Permutations: The number of different ways that one can pick $k$ out of $n$ objects when order of picking is important: $\frac{n!}{(n-k)!}$.

Ex: Count the number of 4-letter password with distinct letters where letters are from the English alphabet and are small case only.
**Combinations** : The number of different ways that one can pick \( k \) out of \( n \) objects when order of picking is **not important** :  
\[ \binom{n}{k} = \frac{n!}{k!(n - k)!}. \]

**Ex:** Count the number of 3-person teams out of 8 people.

**Partitions** : Number of different ways to partition a set with \( n \) elements into \( r \) disjoint subsets with the \( i^{th} \) subset having size \( n_i \) :  
\[ \left( \begin{array}{c} n \n_1, n_2, \ldots, n_r \end{array} \right) = \frac{n!}{n_1!n_2!\ldots n_r!}. \]

**Ex:** How many different words can be formed by the letters in the word RATATU?

**Ex:** Four balls colored red, green, blue and yellow are distributed randomly to three boxes, Box1, Box2 and Box3. Each ball can go to any box with equal probability.

1. Find the probability that Box1 contains the red ball and Box2 the remaining balls.

2. Find probabilities of events A, B and C:
   - A=\{ Box1 contains no balls \}
   - B=\{ Box1 contains at least two balls\}
   - C=\{ Box1 has exactly one ball\}

3. Given Box1 contains exactly one ball, what is the probability that exactly two boxes contain equal number of balls.
# Chapter 2

## Discrete Random Variables

### Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Basics</td>
<td>29</td>
</tr>
<tr>
<td>2.2</td>
<td>Probability Mass Functions (PMF)</td>
<td>30</td>
</tr>
<tr>
<td>2.2.1</td>
<td>How to calculate the PMF of a random variable $X$</td>
<td>30</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Properties of PMF</td>
<td>31</td>
</tr>
<tr>
<td>2.3</td>
<td>Some Discrete Random Variables</td>
<td>31</td>
</tr>
<tr>
<td>2.3.1</td>
<td>The Bernoulli R.V.</td>
<td>31</td>
</tr>
<tr>
<td>2.3.2</td>
<td>The Binomial R.V.</td>
<td>31</td>
</tr>
<tr>
<td>2.3.3</td>
<td>The Geometric R.V.</td>
<td>32</td>
</tr>
<tr>
<td>2.3.4</td>
<td>The Poisson R.V.</td>
<td>33</td>
</tr>
<tr>
<td>2.3.5</td>
<td>The Uniform R.V.</td>
<td>33</td>
</tr>
<tr>
<td>2.4</td>
<td>Functions of Random Variables</td>
<td>33</td>
</tr>
<tr>
<td>2.4.1</td>
<td>How to obtain $p_Y(y)$ given $Y = g(X)$ and $p_X(x)$?</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>Expectation, Mean, and Variance</td>
<td>34</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Variance, Moments, and the Expected Value Rule</td>
<td>35</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Properties of Mean and Variance</td>
<td>36</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Mean and Variance of Some Common R.V.s</td>
<td>37</td>
</tr>
<tr>
<td>2.6</td>
<td>Multiple Random Variables and their Joint PMFs</td>
<td>38</td>
</tr>
<tr>
<td>2.6.1</td>
<td>Functions of Multiple Random Variables</td>
<td>40</td>
</tr>
<tr>
<td>2.6.2</td>
<td>More Than Two Random Variables</td>
<td>41</td>
</tr>
<tr>
<td>2.7</td>
<td>Conditioning</td>
<td>42</td>
</tr>
<tr>
<td>2.7.1</td>
<td>Conditioning a random variable on an event</td>
<td>42</td>
</tr>
<tr>
<td>2.7.2</td>
<td>Conditioning one random variable on another</td>
<td>43</td>
</tr>
<tr>
<td>2.7.3</td>
<td>Conditional Expectation</td>
<td>47</td>
</tr>
<tr>
<td>2.8</td>
<td>Independence</td>
<td>49</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Independence of a Random Variable from an Event</td>
<td>49</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Independence of Random Variables</td>
<td>50</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Variance of the Sum of Independent Random Variables</td>
<td>51</td>
</tr>
</tbody>
</table>
2.1 Basics

Many random experiments have outcomes that are numerical. For example, number of people in a bus, the time a customer will spend in line in a bank etc.

In random experiments with outcomes that are not numerical, it is very useful to map outcomes to numerical values. For example, in a coin toss experiment: \( H \to 1 \) and \( T \to 0 \).

A random variable associates a number with each outcome of a random experiment.

**Definition 10** A random variable is a real-valued function of the outcome of a random experiment, i.e. it is a mapping (function) from the sample space to real numbers. (We denote random variables with capital letters, such as \( X \).)

**Ex:** Two rolls of a 4-faced die. Let \( X \) be the maximum of the two rolls.

**Ex:** Experiment involving a sequence of 5 tosses of a coin.

- The 5-long sequence of heads and tails is not a random variable since it does not have an explicit numerical value.
- The number of heads is a random variable since it is an explicit numerical value.

**Definition 11** The range of a random variable is the set of values it can take.

**Ex:** Two rolls of a 4-faced die. Let \( X \) be the maximum of the two rolls.
Definition 12 A random variable is **discrete** if its range is a countable (finite or infinite) set.

**Ex:** Consider the experiment of choosing a point in $[-1, 1]$ randomly.

Note that a function of a random variable is another mapping from the sample space to real numbers, hence another random variable.

### 2.2 Probability Mass Functions (PMF)

The most important way to characterize a discrete random variable is through its Probability Mass Functions (PMF).

**Definition 13** If $x$ is any possible value that a random variable $X$ can take, the probability mass of $x$, denoted by $p_X(x)$, is the probability of the event $\{X = x\}$ that consists of all outcomes that are mapped to $x$ by $X$:

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

#### 2.2.1 How to calculate the PMF of a random variable $X$

For each possible value $x$ of $X$:

1. Collect all possible outcomes of the random experiment that give rise to the event $\{X = x\}$.
2. Add their probabilities to obtain $p_X(x)$.

**Ex:** Two rolls of a fair 4-faced die. Let $X$ be the maximum of the two rolls. Find $p_X(k)$. 
Ex: A fair coin is tossed twice. The random variable $X$ is defined as the number of times heads shows up. Define events \( \{X = 0\} \), \( \{X = 1\} \), \( \{X = 2\} \) and find PMF $p_X(k)$.

2.2.2 Properties of PMF

1. $\sum_x p_X(x) =$

2. $P(X \in S) = \sum_{x \in S} p_X(x)$

2.3 Some Discrete Random Variables

2.3.1 The Bernoulli R.V.

$$p_X(k) = \begin{cases} p, & \text{if } k = 1 \\ 1 - p, & \text{if } k = 0 \end{cases}$$

- A simple r.v. that can take only one of two values: 1 and 0.
- Bernoulli r.v. can model random experiments with only two possible outcomes. For ex.,
  - a coin toss,
  - an experiment with possible outcome of either success or failure.
- Despite its simplicity, Bernoulli r.v. is very important because multiple Bernoulli r.v. can be combined to model more complex random variables.
- Short-hand notation:

2.3.2 The Binomial R.V.

$$p_X(k) = P(X = k) =$$

- The Binomial r.v. is the number of 1’s in a random experiment that consists of $n$ stages where each stage is an independent and identical Bernoulli random variable (Ber(p)). E.g.,
– number of H’s in \( n \) tosses of a coin,
– number of successes in \( n \) trials.

• Short-hand notation : \( X \sim \text{Binom}(n, p) \)

\[
\begin{align*}
\text{Let us check the legitimacy of this PMF:}
\end{align*}
\]
2.3.4 The Poisson R.V.

A Poisson r.v. with a positive parameter \( \lambda \) has the following PMF:

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

- A Poisson r.v. expresses the probability of how many times an event occurs in a fixed period of time if these events
  - occur with known average rate of \( \lambda \)
  - and independently of each other.
- The poisson r.v. is used to model traffic in many practical problems, such as number of cars arriving at an intersection, or number of Internet packets arriving at a router, or number of customers arriving at a supermarket, etc..

**Ex:** Traffic engineers often use a Poisson r.v. to model flow of cars in traffic. Suppose that during the hours from 9 to 10 each day, on average 4.7 cars pass through a red light without stopping. During such an hour, what is the probability that \( k \) cars pass through without stopping at that red light without stopping?

2.3.5 The Uniform R.V.

A uniform random variable has a finite range given by set \( S \) and has uniformly distributed probabilities for all values it can take from \( S \).

\[
p_X(k) = \frac{1}{|S|} \quad \text{for all} \quad k \in S.
\]

**Ex:** A four-sided die is rolled. Let \( X \) be the outcome.

2.4 Functions of Random Variables

- An engineering problem: system with input \( x \) and output \( y \) related through \( y = g(x) \).

- If the input is random, the output is also random.
- Output \( Y \) is a function of the input \( X \) : \( Y = g(X) \).
Ex: A uniform r.v. $X$ whose range is the integers in $[-2, 2]$. It is passed through a transformation $Y = |X|$. Find PMF of $Y$.

2.4.1 How to obtain $p_Y(y)$ given $Y = g(X)$ and $p_X(x)$?

To obtain $p_Y(y)$ for any $y$, 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.

2.5 Expectation, Mean, and Variance

We are sometimes interested in a summary of certain properties of a random variable. Ex: You don’t have to compare your grade to all the other grades in a class as a first approximation. Comparison to the average grade can tell a lot.

Definition 14 The expected value (also called expectation or mean) of a r.v. $X$ is defined as

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

- The mean of a r.v. $X$ is the weighted average (in proportion to probabilities) of the possible values of $X$.
- The mean corresponds to the center of gravity of the PMF.

Ex: $X$: outcome of fair die roll. $E[X] =$?
2.5.1 Variance, Moments, and the Expected Value Rule

Another very important property of a random variable is its variance.

**Definition 15** The variance of a r.v. $X$ provides a measure of spread of its PMF around its mean and is defined as

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

- The variance is always nonnegative.
- The standard deviation $\sigma_X = \sqrt{\text{var}(X)}$. It is usually simpler to interpret since its unit is the same as the unit of $X$.
- One way to calculate $\text{var}(X)$ is to calculate the mean of $Z = (X - E[X])^2$.

**Ex:** Consider the length (in cm) of gilt-head bream (čipura, Sparus aurata) sold at market as a random variable $X$. Find the variance with the following PMFs.

(a) $p_X(15) = p_X(20) = p_X(25) = \frac{1}{3}$.

(b) $p_X(20) = \frac{98}{100}, p_X(18) = p_X(22) = \frac{1}{100}$.

A typically easier way to calculate $\text{var}(X)$ is due to the following rule, called the **expected value rule**.

**Theorem 4** 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:**
The expected value rule can be utilized to evaluate the variance.

$$\text{var}(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 p_X(x).$$

**Ex:** The çipura example revisited.

**Definition 16** The $n^{th}$ moment of a r.v. $X$ is defined as

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

### 2.5.2 Properties of Mean and Variance

- Linearity of the expectation

- The expected value of $Y = aX + b$

- The variance of $Y = aX + b$

- Variance in terms of moments: $\text{var}(X) = E[X^2] - (E[X])^2$
• Unless $g(X)$ is linear, $E[g(X)] \neq g(E[X])$. (see example below)

**Ex:** Average speed vs average time. On the highway from Gebze to Kartal, one drives at a speed of 120kmh when there is no traffic jam. The speed drops to 30kmh in case of traffic jam. The probability of traffic jam is 0.2 and the distance from Gebze to Kartal through the highway is 60km. Find the expected value of the driving time.

### 2.5.3 Mean and Variance of Some Common R.V.s

The Bernoulli r.v.

The discrete uniform r.v. :
First consider $X$ defined over $\{1, 2, ..., n\}$. Then consider $Y$ defined over $\{a, a + 1, ..., b\}$.

The Poisson r.v.
2.6 Multiple Random Variables and their Joint PMFs

More than one random variable can be associated with the same random experiment, in which case one can consider a mapping from the sample space $\Omega$ to the real plane $\mathbb{R}^2$.

Ex: A coin is tossed three times. $X$: number of heads, $Y$: number of tails.

One can now talk about events such as $\{X = x, Y = y\}$.

**Definition 17** For two random variables $X$ and $Y$ that are associated with the same random experiment, the **joint PMF** of $X$ and $Y$ is defined as

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

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

- How to calculate joint PMF $p_{X,Y}(x, y)$? For each possible pair of values $(x, y)$ that $X$ and $Y$ can take:
  1. Collect all possible outcomes of the random experiment that give rise to the event $\{X = x, Y = y\}$.
  2. Add their probabilities to obtain $p_{X,Y}(x, y)$.

Ex: In the above example, assume the coin is fair and find joint PMF $p_{X,Y}(x, y)$.

- The joint PMF $p_{X,Y}(x, y)$ determines the probability of any event involving $X$ and $Y$:

$$P((X, Y) \in A) = \sum_{(x, y) \in A} p_{X,Y}(x, y).$$
Ex: In the above example, what is the probability that both $X$ and $Y$ are less than 3?

The term **marginal PMF** is used for $p_X(x)$ and $p_Y(y)$ to distinguish them from the joint PMF $p_{X,Y}(x,y)$.

- Can one find marginal PMFs from the joint one? How?
  
  - $p_X(x) = \sum_y p_{X,Y}(x,y)$
  - $p_Y(y) = \sum_x p_{X,Y}(x,y)$

  Pf. : 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,

  $$p_X(x) = P(X = x) = P(\{X = x\}) = P(\bigcup_y \{X = x, Y = y\}) = \sum_y P(\{X = x, Y = y\}) = \sum_y P(X = x, Y = y)$$

  $$= \sum_y p_{X,Y}(x,y).$$

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

- The above equations indicate that if the joint PMF are arranged in a table, then one can find marginal PMFs by **adding** the table entries along the columns or rows (this method is called tabular method).

  **Ex:** Two r.v.s $X$ and $Y$ have the joint PMF $p_{X,Y}(x,y)$ given in the 2-D table below. Find marginal PMFs. Find the probability that $X$ is smaller than $Y$?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1/10</td>
<td>2/10</td>
</tr>
<tr>
<td>2</td>
<td>1/10</td>
<td>1/15</td>
<td>1/30</td>
</tr>
<tr>
<td>3</td>
<td>2/10</td>
<td>2/10</td>
<td>1/10</td>
</tr>
</tbody>
</table>
• Note that one can find marginal PMFs from the joint PMF, but the reverse is not true in general.

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

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

Find $c$ and the marginal PMFs.

### 2.6.1 Functions of Multiple Random Variables

Let $Z = g(X, Y)$. The PMF of $Z$ can be found by

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

**Ex:** Two r.v.s $X$ and $Y$ have the joint PMF $p_{X,Y}(x, y)$ given in the 2-D table below. Find PMF of $Z = 2X + Y$. Find $E[Z]$.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1/10</td>
<td>2/10</td>
</tr>
<tr>
<td>2</td>
<td>1/10</td>
<td>1/15</td>
<td>1/30</td>
</tr>
<tr>
<td>3</td>
<td>2/10</td>
<td>2/10</td>
<td>1/10</td>
</tr>
</tbody>
</table>
2.6.2 More Than Two Random Variables

The joint PMF of three random variables $X$, $Y$ and $Z$ is defined in analogy as

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

Corresponding marginal PMFs can be found by analogous equations

$$p_{X,Y}(x, y) = \sum_z p_{X,Y,Z}(x, y, z)$$
$$p_X(x) = \sum_y \sum_z p_{X,Y,Z}(x, y, z)$$

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

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

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

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

Ex: Expectation of the Binomial r.v.

Ex: Suppose that $n$ people throw their hats in a box and then the hats are randomly distributed back to the people. What is the expected value of $X$, the number of people that get their own hats back?
2.7 Conditioning

2.7.1 Conditioning a random variable on an event

The occurrence of a particular event may affect the PMF of a r.v. $X$.

**Ex:** There are two coins with different probabilities of H’s : $1/2$ and $1/4$. One of the coins is selected in the beginning. The number of heads after 4 coin tosses is denoted by $X$.

**Definition 18** The conditional PMF of the random variable $X$, conditioned on the event $A$ with $P(A) > 0$ is defined by

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

- Let us show that $p_{X|A}(x)$ is a legitimate PMF. (Expand $P(A)$ using the total probability theorem)

- How can we calculate the conditional PMF $p_{X|A}(x)$ ?
  The above definition is applied, i.e. the probabilities of the outcomes that give rise to $X = x$ and belong to the conditioning event $A$ are added and then normalized by dividing with $P(A)$.

**Ex:** Let $X$ be the outcome of one roll of a tetrahedral die, and $A$ be the event that the outcome is not 1.
Ex: Ali has a total of $n$ chances to pass his motorcycle license test. Suppose each time he 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 $a + n$ ($a > 0$, $n > 0$). 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 $k$.

2.7.2 Conditioning one random variable on another

Consider a random experiment that is associated with two r.v.s $X$ and $Y$. 
• The knowledge that $Y$ equals a particular value $y$ may provide a partial knowledge on $X$.

• A conditional PMF of a random variable $X$ conditioned on another rv $Y$ is defined by using events $A$ of the form $\{Y = y\}$: (i.e. $p_{X|A}(x)$ where $A = \{Y = y\}$)

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

• Using definition of conditional probabilities, we have

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

Pf.

• Note that using the joint PMF, one can obtain the marginal and the conditional PMFs.

• Let us fix some value $y$ (with $p_Y(y) > 0$) and consider conditional PMF $p_{X|Y}(x|y)$ as a function of only $x$:

  – Then conditional PMF $p_{X|Y}(x|y)$ assigns non-negative values (i.e. probabilities) to each possible $x$ and these values add up to 1:

$$\sum_x p_{X|Y}(x|y) = 1$$

  – Furthermore, joint PMF $p_{X|Y}(x|y)$ has the same shape as joint PMF $p_{X,Y}(x,y)$ except that it is divided by $p_Y(y)$ which performs normalization.
**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

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

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

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

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

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

Conditional PMFs involving more than two random variables are defined similarly:

- $p_{X,Y|Z}(x, y|z) = P(X = x, Y = y|Z = z)$
- $p_{X|Y,Z}(x|y, z) = P(X = x|Y = y, Z = z)$

**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 $p_{Y}(y)$.  

45
Ex: (From textbook) Prof. Right answers each student question incorrectly with probability \( \frac{1}{4} \), independent of other questions. In each lecture Prof. Right is asked 0, 1, or 2 questions with equal probability of \( \frac{1}{3} \). Let \( X \) and \( Y \) be the number of questions Prof. Right is asked and the number of questions she answers incorrectly in a given lecture, respectively. Find the joint PMF \( p_{X,Y}(x, y) \).

Ex: A transmitter is sending messages over a computer network. Let \( X \) be the travel time of a given message and \( Y \) be the length of a given message. The length of the message \( Y \) can take two possible values: \( y = 10^2 \) bytes with probability \( \frac{5}{6} \) and \( y = 10^4 \) bytes with probability \( \frac{1}{6} \). Travel time \( X \) of the message depends on its length \( Y \) and the congestion in the network at the time of transmission. In particular, the travel time \( X \) is \( 10^{-4}Y \) seconds with probability \( \frac{1}{2} \), \( 10^{-3}Y \) seconds with probability \( \frac{1}{3} \), and \( 10^{-2}Y \) seconds with probability \( \frac{1}{6} \). Find the PMF of travel time \( X \).
2.7.3 Conditional Expectation

Conditional Expectations are defined similar to ordinary expectations except that conditional PMFs are used:

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

The expected value rule also extends similarly to conditional expectations:

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

Let us recall the Total Probability Theorem in the context of random variables.

- For \( A_i \)'s forming a partition of the sample space, we have

\[ p_X(x) = P(\{X = x\}) = \sum_i P(\{X = x\}|A_i)P(A_i) \]
\[ = \sum_i \]

- Similarly, the sets \( \{Y = y\} \) as \( y \) goes over the entire range of \( Y \) form a partition of the sample space, and we have

\[ p_X(x) = P(\{X = x\}) = \sum_y P(\{X = x\}|\{Y = y\})P(\{Y = y\}) \]
\[ = \sum_y \]

Let us evaluate the expectation of \( X \) based on the above formulations of \( p_X(x) \):

\[ E[X] = \sum_x xp_X(x) = \sum_x x \sum_i p_{X|A_i}(x)P(A_i) \]
\[ = \]
\[ = \]

\[ E[X] = \sum_x xp_X(x) = \sum_x x \sum_y p_{X|Y}(x|y)p_Y(y). \]
\[ = \]
\[ = \]
The equalities obtained above are collectively called the **Total Expectation Theorem**:

\[
E[X] = \sum_i E[X|A_i]P(A_i)
\]

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

**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:** \(X\) 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:** 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 \(\{X > 1\}\).
Consider two rolls of a fair die. Let $X$ be the total number of 6’s, and $Y$ be the total number of 1’s. Find $E[X|Y = y]$ and $E[X]$.

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

### 2.8 Independence

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

#### 2.8.1 Independence of a Random Variable from an Event

For the independence of a rv $X$ and an event $A$, the events $\{X = x\}$ and $A$ should be independent for all $x$ values.

**Definition 19** The random variable $X$ is independent of the event $A$ if and only if

$$P(\{X = x\} \cap A) = P(X = x)P(A) = p_X(x)P(A) \quad \text{for all } x.$$
Equivalent characterization in terms of conditional PMF:

**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.

### 2.8.2 Independence of Random Variables

For two random variables $X$ and $Y$ to be independent, knowledge on $X$ should convey no information on $Y$, and vice versa. In other words, the events $\{X = x\}$ and $\{Y = y\}$ should be independent for all $x, y$:

$$P(\{X = x\} \cap \{Y = y\}) = \text{for all } x, y.$$  

**Definition 20** Two random variables $X$ and $Y$ are independent if and only if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y) \quad \text{for all } x, y.$$  

- Equivalent characterization in terms of conditional PMFs:

- If $X$ and $Y$ are independent, then $E[XY] = E[X]E[Y]$. (Reverse may not be true.)
• If $X$ and $Y$ are independent, then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

• **Conditional independence** of two random variables conditioned on an event $A$ is obtained again through events $\{X = x\}$ and $\{Y = y\}$:

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

• The independence definition given above can be extended to multiple random variables in a straightforward way. For example, three random variables $X, Y, Z$ we have

**Ex:** Joint PMF of $X$ and $Y$ are given in the table. Are $X$ and $Y$ independent?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1/10</td>
<td>2/10</td>
</tr>
<tr>
<td>2</td>
<td>1/10</td>
<td>1/15</td>
<td>1/30</td>
</tr>
<tr>
<td>3</td>
<td>2/10</td>
<td>2/10</td>
<td>1/10</td>
</tr>
</tbody>
</table>

2.8.3 Variance of the Sum of Independent Random Variables

For two independent random variable $X$ and $Y$, consider their sum called $Z = X + Y$. Let’s find expectation and variance of $Z$.

• $E[Z] = E[X + Y] = E[X] + E[Y]$ (this is true even if $X$ and $Y$ are not independent!)

• $\text{var}(Z) = \text{var}(X + Y) =$

Pf.
If one repetitively uses the above results, the general formula for the sum of multiple independent random variables is obtained.

- \( E[X_1 + X_2 + \ldots + X_n] = \)
- \( \text{var}(X_1 + X_2 + \ldots + X_n) = \)

**Ex:** The variance of the Binomial random variable.

**Ex:** (Mean and variance of the sample mean) Let \( X_1, X_2, \ldots, X_n \) be independent Bernoulli random variables with common mean and variance (i.e. they are i.i.d.). Let us define the sample mean and find its mean and variance.

**Ex:** (Estimating probabilities by simulation) In many practical situations, the analytical calculation of the probability of some event of interest is very difficult. If so, we may resort to (computer) simulations where we observe the outcomes of a certain experiment performed many times independently. Say we are interested in finding \( P(A) \) of an event \( A \) defined from the experiment. Define the sample mean and find its mean and variance.
In some random experiments, one can define random variables that can take on a continuous range of possible values. For example, your weight is a continuous random variable. (If you round your weight to an integer, then it becomes a discrete random variable.) The use of continuous models may result in insights not possible with discrete modeling.

All of the concepts and tools introduced for discrete random variables, such as PMFs, expectation and conditioning, have continuous counterparts and will be discussed in this chapter.
3.1 Continuous Random Variables and PDFs

In Chapter 2, we found the probability of an event $A$ associated with a discrete random variable $X$ by summing up its probability mass function over the values in that set:

$$ P(X \in A) = \sum_{x \in A} p_X(x). $$

To find the probability of an event $A$ associated with a continuous random variable $X$, summation of probabilities over the values in that set is not feasible. Instead, a probability density function is integrated over the values in $A$:

$$ P(X \in A) = \int_{x \in A} f_X(x) \, dx. $$

Definition 21 A random variable $X$ is continuous if there is a nonnegative function $f_X$, called the probability density function (PDF) of $X$, such that

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

for every subset $B$ of the real line. In particular, the probability that $X$ is in an interval is

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

Hence, for a continuous rv $X$:

- For any single value $a$, we have $P(X = a) = \int_a^a f_X(x) \, dx = 0$.

- Thus, $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$.

For a valid PDF $f_X(x)$, the following must hold.

1. Non-negativity: $f_X(x) \geq 0$ (otherwise, ..)

2. Normalization:
To interpret the intuition behind PDF, consider a small interval \([x, x + \delta]\) where \(\delta \ll 1\):

- \(P(x < X \leq x + \delta) = \int_{x}^{x+\delta} f_X(a) da \approx \)

- \(f_X(x)\) can be viewed as the "probability mass per unit length near \(x\)" (or density of probability mass near \(x\))
- \(f_X(x)\) is not itself an event’s probability
- \(f_X(x)\) can be larger than 1 at some \(x\)

**Ex:** (PDF can be larger than 1) PDF of random variable \(X\) is given below. Find \(c\) and \(P(\mid X\mid^2 \leq 0.5)\).

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

**Ex:** (A PDF can take arbitrarily large values) Consider a r.v. with

\[
f_X(x) = \begin{cases} 
  cx^{-1/2}, & 0 \leq x \leq 2 \\
  0, & \text{o.w.}
\end{cases}
\]

Find \(c\) and inspect the graph of the PDF.
3.1.1 Some Continuous Random Variables and Their PDFs

Continuous Uniform R.V.

If the probability density is uniform (i.e. constant) over the set of values that the rv takes, we have a continuous uniform rv.

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

Gaussian (Normal) R.V.

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

Figure 3.1: 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 \( \lambda \) is a positive parameter.

![PDF for \( \lambda = 0.5 \) and \( \lambda = 2 \)](image)

Figure 3.2: Exponential PDF

### 3.1.2 Expectation

By changing summation with an integral, the following definition for the expected value (or mean or expectation) of a continuous r.v. is obtained.

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

\( E[X] \) of a continuous rv can be interpreted as (like in chapter 2)

- the center of gravity of the PDF,
- average obtained over a large number of independent trials of an experiment.

Many properties are directly carried over from the discrete counterpart.

- If \( X \) is a continuous rv, then \( Y = g(X) \) is also a random variable, however it can be continuous or discrete depending on \( g(.) \):
\[ Y = g_1(X) = 2X \]
\[ Y = g_2(X) = u(X) \]

- **Expected value rule**: \( E[g(X)] = \)

- **\( n^{th} \) moment**: \( E[X^n] = \)

- **\( var(X) \) =**

- **If** \( Y = aX + b \)

  - **\( E[Y] = \)**
  - **\( var(Y) = \)**

**Ex:** Mean and variance of a uniform r.v.

**Ex:** Mean and variance of an exponential r.v.
3.2 Cumulative Distribution Functions

A discrete r.v. is characterized by its PMF whereas a continuous one with its PDF. Cumulative distribution functions (CDF) are defined to characterize all sorts of r.v.s.

Definition 22 The cumulative distribution function (CDF) of a random variable $X$ is defined as

$$F_X(x) = P(X \leq x) = \begin{cases} \text{, if } X \text{ is discrete} \\ \text{, if } X \text{ is continuous} \end{cases}$$

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

Ex: CDF of a discrete r.v.

Hence, for discrete rvs, CDFs $F_X(x)$ are

- piecewise constant .
- continuous from right but not from left at the jump points.

Ex: CDF of a continuous r.v.
Hence, for **continuous rvs**, CDFs $F_X(x)$ are

- continuous, i.e. from right and left. (i.e. no jumps in CDFs)

### 3.2.1 Properties of CDF

(a)  
- $0 \leq F_X(x) \leq 1$
- $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) =$
- $F_X(\infty) = \lim_{x \to \infty} F_X(x) =$

(b) $P(X > x) = 1 - F_X(x)$

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

(d)  
- If $X$ is discrete, CDF $F_X(x)$ is a piecewise constant function of $x$.
- If $X$ is continuous, CDF $F_X(x)$ is a continuous function of $x$.

(e) $P(a < X \leq b) = F_X(b) - F_X(a)$.
    
    **Proof:**

(f) $0 = F_X(x^+) - F_X(x)$, where $F_X(x^+) = \lim_{\delta \to 0} F_X(x + \delta)$.
    
    **Proof:**

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

- If $X$ is discrete and takes only integer values,

$$p_X(k) = F_X(k) - F_X(k - 1).$$

- If $X$ is continuous,

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

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

Note that sometimes, to calculate the PMF of a discrete rv or PDF of a continuous rv, it is more convenient to calculate the CDF first, and then obtain the PMF or PDF.

**Ex:** (Maximum of several r.v.s) Let $X_1, X_2$ and $X_3$ be independent rvs and $X = \max\{X_1, X_2, X_3\}$. $X_i$ are discrete uniform taking values in $\{1, 2, 3, 4, 5\}$. Find $p_X(k)$.

**Ex:** The geometric and exponential CDFs
3.2.2 Hybrid Random Variables

Random variables that are neither continuous nor discrete are called hybrid or mixed random variables.

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

---

**Ex:** Assume when you go to a bus station, there is a bus with probability of \( \frac{1}{3} \) and you wait for a bus with probability of \( \frac{2}{3} \). If there is no bus at the stop, the next bus arrives anytime in the next 10 minutes equiprobably. Let \( X \) be the waiting time for the bus. Find \( F_X(x) \), \( P(X \leq 5) \), \( f_X(x) \) and \( E[X] \).
3.3 Normal (Gaussian) Random Variables

**Definition 23** A random variable $X$ is said to normal or Gaussian if its PDF is in the form

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

where $\mu$ and $\sigma$ are scalar parameters characterizing the PDF and $\sigma \geq 0$.

- One can show that $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$. (see Problem 14 of Section 3 in textbook)
- The **mean** is $\mu$ since the PDF is symmetric around $\mu$.
- The **variance** can be found as follows.

$$\text{var}(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma^2}e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx =$$

- A common notation for Normal rvs:
- A Gaussian r.v. has **several special properties**. One of these properties is as follows. (Other properties are discussed in more advanced probability or random process courses)

  - **Normality (Gaussian) is preserved by linear transformations** (Proof in Chp4.)

    If $X$ is a normal rv with mean $\mu$ and variance $\sigma^2$, then the r.v. $Y = aX + b$ is also a normal rv with $E[Y] = a\mu + b$ and $\text{var}(Y) = a^2\sigma^2$.

3.3.1 The Standard Normal R.V.

**Definition 24** A Gaussian (normal) rv with **zero mean** and **unit variance** is called a standard normal.s

The CDF for a standard normal r.v. $N$ is:

$$F_N(x) = P(N \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}} dt$$
• This CDF has a special notation: \( \Phi(x) = F_N(x) \).

• \( \Phi(-c) = 1 - \Phi(c) \) (from symmetry of \( \Phi(x) \))

• \( \Phi(.) \) cannot be directly evaluated, however, it can be calculated numerically (i.e. using numerical software.) We will use a table to find \( \Phi(x) \) for several \( x \) values:

<table>
<thead>
<tr>
<th>( x )</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
<th>.03</th>
<th>.04</th>
<th>.05</th>
<th>.06</th>
<th>.07</th>
<th>.08</th>
<th>.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.5000</td>
<td>.5040</td>
<td>.5080</td>
<td>.5120</td>
<td>.5160</td>
<td>.5199</td>
<td>.5239</td>
<td>.5279</td>
<td>.5319</td>
<td>.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>.5398</td>
<td>.5438</td>
<td>.5478</td>
<td>.5517</td>
<td>.5557</td>
<td>.5596</td>
<td>.5636</td>
<td>.5675</td>
<td>.5714</td>
<td>.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>.5793</td>
<td>.5832</td>
<td>.5871</td>
<td>.5910</td>
<td>.5948</td>
<td>.5987</td>
<td>.6026</td>
<td>.6064</td>
<td>.6103</td>
<td>.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>.6179</td>
<td>.6217</td>
<td>.6255</td>
<td>.6293</td>
<td>.6331</td>
<td>.6368</td>
<td>.6406</td>
<td>.6443</td>
<td>.6480</td>
<td>.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>.6554</td>
<td>.6591</td>
<td>.6628</td>
<td>.6664</td>
<td>.6700</td>
<td>.6736</td>
<td>.6772</td>
<td>.6808</td>
<td>.6844</td>
<td>.6879</td>
</tr>
<tr>
<td>0.5</td>
<td>.6915</td>
<td>.6950</td>
<td>.6985</td>
<td>.7019</td>
<td>.7054</td>
<td>.7088</td>
<td>.7123</td>
<td>.7157</td>
<td>.7190</td>
<td>.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>.7257</td>
<td>.7291</td>
<td>.7324</td>
<td>.7357</td>
<td>.7389</td>
<td>.7422</td>
<td>.7454</td>
<td>.7486</td>
<td>.7517</td>
<td>.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>.7580</td>
<td>.7611</td>
<td>.7642</td>
<td>.7673</td>
<td>.7704</td>
<td>.7734</td>
<td>.7764</td>
<td>.7794</td>
<td>.7823</td>
<td>.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>.7881</td>
<td>.7910</td>
<td>.7939</td>
<td>.7967</td>
<td>.7995</td>
<td>.8023</td>
<td>.8051</td>
<td>.8078</td>
<td>.8106</td>
<td>.8133</td>
</tr>
<tr>
<td>0.9</td>
<td>.8159</td>
<td>.8186</td>
<td>.8212</td>
<td>.8238</td>
<td>.8264</td>
<td>.8289</td>
<td>.8315</td>
<td>.8340</td>
<td>.8365</td>
<td>.8389</td>
</tr>
<tr>
<td>1.0</td>
<td>.8413</td>
<td>.8438</td>
<td>.8461</td>
<td>.8485</td>
<td>.8508</td>
<td>.8531</td>
<td>.8554</td>
<td>.8577</td>
<td>.8599</td>
<td>.8621</td>
</tr>
<tr>
<td>1.1</td>
<td>.8643</td>
<td>.8665</td>
<td>.8686</td>
<td>.8708</td>
<td>.8729</td>
<td>.8749</td>
<td>.8770</td>
<td>.8790</td>
<td>.8810</td>
<td>.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>.8849</td>
<td>.8869</td>
<td>.8888</td>
<td>.8907</td>
<td>.8925</td>
<td>.8944</td>
<td>.8962</td>
<td>.8980</td>
<td>.8997</td>
<td>.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>.9032</td>
<td>.9049</td>
<td>.9066</td>
<td>.9082</td>
<td>.9099</td>
<td>.9115</td>
<td>.9131</td>
<td>.9147</td>
<td>.9162</td>
<td>.9177</td>
</tr>
<tr>
<td>1.4</td>
<td>.9192</td>
<td>.9207</td>
<td>.9222</td>
<td>.9236</td>
<td>.9251</td>
<td>.9265</td>
<td>.9279</td>
<td>.9292</td>
<td>.9306</td>
<td>.9319</td>
</tr>
<tr>
<td>1.5</td>
<td>.9332</td>
<td>.9345</td>
<td>.9357</td>
<td>.9370</td>
<td>.9382</td>
<td>.9394</td>
<td>.9406</td>
<td>.9418</td>
<td>.9429</td>
<td>.9441</td>
</tr>
<tr>
<td>1.6</td>
<td>.9452</td>
<td>.9463</td>
<td>.9474</td>
<td>.9484</td>
<td>.9495</td>
<td>.9505</td>
<td>.9515</td>
<td>.9525</td>
<td>.9535</td>
<td>.9545</td>
</tr>
<tr>
<td>1.7</td>
<td>.9554</td>
<td>.9564</td>
<td>.9573</td>
<td>.9582</td>
<td>.9591</td>
<td>.9599</td>
<td>.9608</td>
<td>.9616</td>
<td>.9625</td>
<td>.9633</td>
</tr>
<tr>
<td>1.8</td>
<td>.9641</td>
<td>.9649</td>
<td>.9656</td>
<td>.9664</td>
<td>.9671</td>
<td>.9678</td>
<td>.9686</td>
<td>.9693</td>
<td>.9699</td>
<td>.9706</td>
</tr>
<tr>
<td>1.9</td>
<td>.9713</td>
<td>.9719</td>
<td>.9726</td>
<td>.9732</td>
<td>.9738</td>
<td>.9744</td>
<td>.9750</td>
<td>.9756</td>
<td>.9761</td>
<td>.9767</td>
</tr>
<tr>
<td>2.0</td>
<td>.9772</td>
<td>.9778</td>
<td>.9783</td>
<td>.9788</td>
<td>.9793</td>
<td>.9798</td>
<td>.9803</td>
<td>.9808</td>
<td>.9812</td>
<td>.9817</td>
</tr>
<tr>
<td>2.1</td>
<td>.9821</td>
<td>.9826</td>
<td>.9830</td>
<td>.9834</td>
<td>.9838</td>
<td>.9842</td>
<td>.9846</td>
<td>.9850</td>
<td>.9854</td>
<td>.9857</td>
</tr>
<tr>
<td>2.2</td>
<td>.9861</td>
<td>.9864</td>
<td>.9868</td>
<td>.9871</td>
<td>.9875</td>
<td>.9878</td>
<td>.9881</td>
<td>.9884</td>
<td>.9887</td>
<td>.9890</td>
</tr>
<tr>
<td>2.3</td>
<td>.9893</td>
<td>.9896</td>
<td>.9898</td>
<td>.9901</td>
<td>.9904</td>
<td>.9906</td>
<td>.9909</td>
<td>.9911</td>
<td>.9913</td>
<td>.9916</td>
</tr>
<tr>
<td>2.4</td>
<td>.9918</td>
<td>.9920</td>
<td>.9922</td>
<td>.9925</td>
<td>.9927</td>
<td>.9929</td>
<td>.9932</td>
<td>.9933</td>
<td>.9934</td>
<td>.9936</td>
</tr>
<tr>
<td>2.5</td>
<td>.9938</td>
<td>.9940</td>
<td>.9941</td>
<td>.9943</td>
<td>.9945</td>
<td>.9946</td>
<td>.9948</td>
<td>.9949</td>
<td>.9951</td>
<td>.9952</td>
</tr>
<tr>
<td>2.6</td>
<td>.9953</td>
<td>.9955</td>
<td>.9956</td>
<td>.9957</td>
<td>.9959</td>
<td>.9960</td>
<td>.9961</td>
<td>.9962</td>
<td>.9963</td>
<td>.9964</td>
</tr>
<tr>
<td>2.7</td>
<td>.9965</td>
<td>.9966</td>
<td>.9967</td>
<td>.9968</td>
<td>.9969</td>
<td>.9970</td>
<td>.9971</td>
<td>.9972</td>
<td>.9973</td>
<td>.9974</td>
</tr>
<tr>
<td>2.8</td>
<td>.9974</td>
<td>.9975</td>
<td>.9976</td>
<td>.9977</td>
<td>.9977</td>
<td>.9978</td>
<td>.9979</td>
<td>.9979</td>
<td>.9980</td>
<td>.9981</td>
</tr>
<tr>
<td>2.9</td>
<td>.9981</td>
<td>.9982</td>
<td>.9982</td>
<td>.9983</td>
<td>.9984</td>
<td>.9984</td>
<td>.9985</td>
<td>.9985</td>
<td>.9986</td>
<td>.9986</td>
</tr>
<tr>
<td>3.0</td>
<td>.9987</td>
<td>.9987</td>
<td>.9987</td>
<td>.9988</td>
<td>.9988</td>
<td>.9989</td>
<td>.9989</td>
<td>.9989</td>
<td>.9990</td>
<td>.9990</td>
</tr>
<tr>
<td>3.1</td>
<td>.9990</td>
<td>.9991</td>
<td>.9991</td>
<td>.9991</td>
<td>.9992</td>
<td>.9992</td>
<td>.9992</td>
<td>.9992</td>
<td>.9993</td>
<td>.9993</td>
</tr>
<tr>
<td>3.2</td>
<td>.9993</td>
<td>.9993</td>
<td>.9994</td>
<td>.9994</td>
<td>.9994</td>
<td>.9994</td>
<td>.9994</td>
<td>.9995</td>
<td>.9995</td>
<td>.9995</td>
</tr>
<tr>
<td>3.3</td>
<td>.9995</td>
<td>.9995</td>
<td>.9995</td>
<td>.9996</td>
<td>.9996</td>
<td>.9996</td>
<td>.9996</td>
<td>.9996</td>
<td>.9996</td>
<td>.9997</td>
</tr>
<tr>
<td>3.4</td>
<td>.9997</td>
<td>.9997</td>
<td>.9997</td>
<td>.9997</td>
<td>.9997</td>
<td>.9997</td>
<td>.9997</td>
<td>.9997</td>
<td>.9997</td>
<td>.9998</td>
</tr>
</tbody>
</table>

The standard normal table. The entries in this table provide the numerical values of \( \Phi(y) = P(Y \leq y) \), where \( Y \) is a standard normal random variable, for \( y \) between 0 and 3.49. For example, to find \( \Phi(1.71) \), we look at the row corresponding to 1.7 and the column corresponding to 0.01, so that \( \Phi(1.71) = .9554 \). When \( y \) is negative, the value of \( \Phi(y) \) can be found using the formula \( \Phi(y) = 1 - \Phi(-y) \).
• The importance of $\Phi(x)$ comes from the fact that it is used to find probabilities (or CDF) of any normal rv $Y$ with arbitrary mean $\mu$ and variance $\sigma^2$:

1. "Standardize" $Y$ by defining a new random variable $X$ as...
2. Since new rv $X$ is a linear function of $Y$, $X$ is...
3. 
4. Hence, any probability defined in terms of $Y$ can be redefined in terms of $X$:

Ex: The average height of men in Turkey is 175cm where it is believed that the height has a normal distribution. Find the probability that the next baby boy to be born has a height more than 200cm if the variance is 10cm. (All numbers are made up. Assume that new generations are not growing taller.)

Ex: Signal Detection (Ex 3.7 from the textbook.) A binary message is transmitted as a signal $S$, which is either +1 or −1. The communication channel corrupts the transmission with additive Gaussian noise with mean $\mu = 0$ and variance $\sigma^2$. The receiver concludes that the signal +1 (or −1) was transmitted if the received value is not negative (or negative, respectively). What is the probability of error?
3.4 Multiple Continuous Random Variables

The notion of PDF is now extended to multiple continuous random variables. Notions of joint, marginal and conditional PDF’s are discussed. Their intuitive interpretation and properties are parallel to the discrete case of Chapter 2.

**Definition 25** Two random variables associated with the same sample space are said to be jointly continuous if there is a joint probability density function \( f_{X,Y}(x,y) \) such that for any subset \( B \) of the two-dimensional real plane,

\[
P((X,Y) \in B) = \int_{(x,y) \in B} f_{X,Y}(x,y) \, dx \, dy.
\]

- When \( B \) is a rectangle:

- Normalization (\( B \) is the entire real plane):

- The joint PDF at a point can be approximately interpreted as the "probability per unit area" (or density of probability mass) 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 (i.e. marginal PDFs), and their dependencies (i.e. conditional PDFs).

  - As a special case, the probability of an event associated with only one of the rvs:

    \[
    P(X \in A) = P(X \in A, Y \in (-\infty, \infty)) = \int_{A} \int_{-\infty}^{\infty} f_{X,Y}(x,y) \, dy \, dx.
    \]

  - Since \( P(X \in A) = \int_{A} f_{X}(x) \, dx \), the marginals are evaluated as follows:

    \[
    f_{X}(x) = \\
    f_{Y}(y) =
    \]
**Ex:** Suppose that a steel manufacturer is concerned about the total weight of orders s/he received during the months of January and February. Let $X$ and $Y$ be the weight of items ordered in January and February, respectively. The joint probability density is given as

$$f_{X,Y}(x,y) = \begin{cases} c, & 5000 < x \leq 10000, 4000 < y \leq 9000 \\ 0, & \text{o.w.} \end{cases}$$

Determine the constant $c$ and find $P(B)$ where $B = \{X > Y\}$.

**Ex:** Random variables $X$ and $Y$ are jointly uniform in the shaded area $S$. Find the constant $c$ and the marginal PDFs.

**Ex:** Random variables $X$ and $Y$ are jointly uniform in the shaded area $S$. Find the constant $c$ and the marginal PDFs.
3.4.1 Joint CDFs, Mean, More than Two R.V.s

Joint CDF defined for random variables associated with the same random experiment:

\[ F_{X,Y}(x, y) = P(X \leq x, Y \leq y). \]

- Advantage of joint CDF is again that it applies equally well to discrete, continuous and hybrid random variables.

- If \( X \) and \( Y \) are continuous rvs:
  \[
  F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = f_{X,Y}(x, y) =
  \]

- If \( X \) and \( Y \) are discrete rvs:
  \[
  F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = p_{X,Y}(x, y) =
  \]

**Ex:** \( X \) and \( Y \) are jointly uniform in the shaded area. Find joint CDF \( F_{X,Y}(x, y) \).
For a function $Z = g(X, Y)$ of two continuous rvs $X$ and $Y$:

- Finding $f_Z(z)$, the marginal PDF of $Z$, will be discussed in Chapter 4.
- As before, $E[g(X, Y)] = \int \int f_{X,Y}(x, y) dy dx$.
- In case $Z = g(X, Y)$ is linear, recall the linearity of expectation:
  $$E[aX + bY + c] = \int \int f_{X,Y}(x, y) dy dx.$$

More than two continuous rvs can be elaborated in a similar fashion as two rvs:

$$P((X, Y, Z) \in B) = \int \int \int_{(x,y,z) \in B} f_{X,Y,Z}(x, y, z) dx dy dz$$

$$f_{X,Y}(x, y) = \int$$

$$f_X(x) = \int$$

$$E[g(X, Y, Z)] = \int$$

$$E[a_1X_1 + a_2X_2 + \ldots + a_kX_k] =$$

### 3.5 Conditioning

**Definition 26** The **conditional PDF** of a continuous random variable $X$, given and event $A$ with $P(A) > 0$, is defined as a nonnegative function $f_{X|A}$ that satisfies

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

for every subset $B$ of the real line.

- If $B$ is entire real line (Normalization property):

- In the special case where we condition on an event of the form $\{X \in A\}$ with $P(X \in A) > 0$:
  $$P(X \in B|X \in A) = \frac{P(X \in B, X \in A)}{P(X \in A)} =$$

Hence,

$$f_{X|\{X \in A\}} = \begin{cases} 
  \frac{f_X(x)}{P(\{X \in A\})}, & \text{if } x \in A \\
  0, & \text{otherwise.}
\end{cases}$$
As in discrete case, conditional PDF is zero outside the conditioning set.

Within conditioning set, conditional PDF has same shape as $f_X(x)$

**Ex:** (The exponential random variable is memoryless.) Let $T$ be the lifetime of a lightbulb, exponential with parameter $\lambda$. Given that you check the bulb at time $t$ and it was fine, find the conditional CDF of the remaining lifetime $X$ of the bulb.

**Conditional PDF for multiple rvs** given event $C = \{(X, Y) \in A\}$:

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

• $f_{X|C}(x) =$
Total probability theorem applied on PDFs. Let $A_1, A_2, \ldots, A_n$ partition the sample space with $P(A_i) > 0$ for each $i$.

\[
P(X \leq x) = \sum_{i=1}^{n} P(A_i)P(X \leq x|A_i)
\]

\[
\int_{t=\infty}^{x} f_X(t)dt = f_X(x) = \int_{t=\infty}^{x} f_X(t)dt
\]

Ex: (textbook) 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 start 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?

3.5.1 Conditioning One R.V. on Another

Definition 27 Let $X$ and $Y$ be two rvs 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 PDF $f_{X|Y}(x|y)$ as a function of the single variable $x$ only (as we did in Chapter 2)
  - Conditional PDF $f_{X|Y}(x|y)$ has the same shape as joint PDF $f_{X,Y}(x,y)$ except that it is divided by $f_Y(y)$ which does not depend on $x$. 

71
* Hence, finding $f_{X|Y}(x|y)$ from $f_{X,Y}(x,y)$ is similar to Chapter 2’s procedure: take a vertical or horizontal slice and normalize (see example below.)

- Normalization property can be shown easily for conditional PDF $f_{X|Y}(x|y)$:

$$\int_{-\infty}^{\infty} = 1$$

Pf.

Ex: (Same example we considered before to find marginals from $f_{X,Y}(x,y)$)

To interpret conditional PDF, let us find the probability of $A = \{x \leq X \leq x + \delta_1\}$ conditioned on $B = \{y \leq Y \leq y + \delta_2\}$ where $\delta_1$ and $\delta_2$ are “small”:

$$P(x \leq X \leq x + \delta_1|y \leq Y \leq y + \delta_2) = \frac{P(x \leq X \leq x + \delta_1, y \leq Y \leq y + \delta_2)}{P(y \leq Y \leq y + \delta_2)}$$

≈

= 

- $f_{X|Y}(x|y) \cdot \delta_1$ provides the probability that $X \in [x, x + \delta_1]$ given that $Y \in [y, y + \delta_2]$.

- As $\delta_2 \to 0$, we have

$$P(x \leq X \leq x + \delta_1|Y = y) \approx P(X \in A|Y = y) = \int_{A} f_{X|Y}(x|y)dx$$

Ex: A grandmom gives $X$ TL to her grandson, where $X$ is uniformly distributed between 0 and 10. The kid spends $Y$ TL of it in the grocery store. Not having reached the age when he would spend all his money, the amount of money $Y$ he spends is uniformly distributed between 0 and $X$. Find $f_Y(y)$. 

72
3.5.2 Conditional CDF

Conditional CDF of \( X \) given event \( A \):

\[
F_{X|A}(x) =
\]

- \( F_{X|A}(x) \) has similar properties as a regular CDF.
- \( f_{X|A}(x) = \frac{dF_{X|A}(x)}{dx} \)

As a special case, consider \( F_{X|A}(x) \) where \( A = \{a < X \leq b\} \):

\[
F_{X|A}(x) = P(X \leq x | a < X \leq b) =
\]

**Ex:** \( X \) is uniform in \([0, 1]\). Let \( A = \{X > \frac{1}{2}\} \). Find \( f_{X|A}(x) \) and \( F_{X|A}(x) \).
Also note that one can define **conditional CDF of one rv conditioned on another**:

\[ F_{Y|X}(y|x) = \int_{-\infty}^{y} \]

\[ f_{Y|X}(y|x) = \]

### 3.5.3 Conditional Expectation

The **conditional expectation** of a continuous rv \( X \) is defined similar to its expectation except that conditional PDFs are used.

\[ E[X|A] = \int \]

\[ E[X|Y = y] = \int \]

For any function \( g(X) \),

\[ E[g(X)|A] = \int \]

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

**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

\[ E[X] = \sum_{i=1}^{n} E[X|A_i]P(A_i) \]

\[ E[X] = \int_{y=-\infty}^{\infty} E[X|Y = y]f_Y(y)dy \]

Pfs.
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.

3.6 Independent Random Variables

Definition 28 Two random variables $X$ and $Y$ are independent if and only if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \text{for all } x,y.$$ 

- Equivalent independence characterization in terms of conditional PDFs:

- Independence of multiple random variables. For example, for three random variables $X,Y,Z$ we have

- If $X$ and $Y$ are independent, we have (note that the reverses are not necessarily true)
  $$P(X \in A,Y \in B) =$$
  $$F_{X,Y}(x,y) =$$
  $$E[XY] =$$
\[ - E[g(X)h(Y)] = \]

\[ - \text{var}(X + Y) = \]

**Ex:** Are \( X, Y \) independent for the joint PDF given below?

\[ f_{X,Y}(x,y) = \begin{cases} 
8xy, & 0 < x \leq 1, 0 < y \leq x \\
0, & \text{o.w.}
\end{cases} \]

**Ex:** Let \( X \) and \( Y \) be independent Normal random variables with means \( \mu_X \) and \( \mu_Y \) and variances \( \sigma_X^2 \) and \( \sigma_Y^2 \). Write the expression for their joint PDF and plot the contours on which it is constant.

### 3.7 The Continuous Bayes’ Rule

Schematic of the inference problem:
For **continuous rvs** $X$ and $Y$:

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

**Ex:** A light bulb has an exponentially distributed lifetime $Y$. The parameter $\lambda$ of $Y$ is a uniform random variable in $[1, \frac{3}{2}]$. You test a light bulb and record its lifetime $y$. What can you say about $\lambda$?

### 3.7.1 Inference about a discrete random variable

If the **unobserved** phenomenon is described by an **event** $A$

$$P(A|Y = y) \approx$$

If **unobserved** event $A$ is defined through a **discrete rv** as in $A = \{N = n\}$:

$$P(N = n|Y = y) =$$

**Ex:** A binary signal $S$ is transmitted with $P(S = 1) = p$ and $P(S = -1) = 1 - p$. The received signal is $Y = S + N$ where $N$ is a standard normal, independent of $S$. What is the probability that $S = 1$, as a function of the observed value $y$ of $Y$?
3.7.2 Inference based on discrete observations

The reverse of the previous case. The unobserved phenomenon is described by continuous rv $Y$ and the observed phenomenon by event $A$:

$$f_{Y|A}(y) = \frac{f_Y(y)P(A|Y = y)}{P(A)} =$$

(This expression can be obtained by rearranging the previous expression for $P(A|Y = y)$.)

If $A$ is defined through a discrete rv as in $A = \{N = n\}$:

$$f_{Y|N}(y|n) = \frac{f_Y(y)p_{N|Y}(n|y)}{p_N(n)} =$$

**Ex:** In a coin tossing experiment, the probability of Heads is not deterministically known, but modeled as a random variable $P$ uniform in $[0.4, 0.6]$. The coin is tossed twice and

(a) only one H is observed. Call this observation event $A$ and find $f_{P|A}(p)$.

(b) two H’s are observed. Call this observation event $B$ and find $f_{P|B}(p)$.

3.8 Some Problems

**Ex:** (The sum of a random number of r.v.s) You visit a random number $N$ of stores and in the $i^{th}$ store, spend a random amount of money $X_i$. The total amount of money you spend is

$$T = X_1 + X_2 + \ldots + X_N$$
where $N$ is a positive integer r.v. with mean $E[N]$ and variance $var(N)$. The r.v.s $X_i$ are independent and identically distributed (i.i.d.), and have mean $E[X]$ and variance $var(X)$. Evaluate the mean and variance of $T$.

**Ex:** The treasury of an underdeveloped country produces coins whose probability of heads is a r.v. $P$ with PDF

$$f_P(p) = \begin{cases} \; \; pe^p & \text{if } p \in [0, 1] \\ 0, & \text{o.w.} \end{cases}.$$  

(a) Find the probability that a coin toss results in heads.

(b) Given that a coin toss resulted in heads, find the conditional PDF of $P$.

(c) Given that the first coin toss resulted in heads, find the conditional probability of heads on the next toss.
Chapter 4

Further Topics on Random Variables

Contents

4.1 Derived Distributions ........................................... 80
   4.1.1 Finding \( f_Y(y) \) directly from \( f_X(x) \) ....................... 82
   4.1.2 Functions of Two Random Variables ............................. 84
4.2 Covariance and Correlation .................................... 87
4.3 Conditional Expectations Revisited (Iterated Expectations) .... 90
4.4 Transforms (Moment Generating Functions) ...................... 91
   4.4.1 From Transforms to Moments ..................................... 92
   4.4.2 Inversion of Transforms ........................................ 93
   4.4.3 Sums of Independent R.V.s ..................................... 93
   4.4.4 Transforms Associated with Joint Distributions ................. 94
4.5 Sum of a Random Number of Independent R.V.s .................. 94

This section covers a number of advanced topics on random variables.

4.1 Derived Distributions

Remember that in Chapter 2 we discussed how to find the PMF of a discrete rv \( Y = g(X) \) from the PMF of discrete rv \( X \) and the function \( g(.) \):

This topic covers the same problem for continuous rvs. In particular, we discuss how to find the distribution of \( Y = g(X) \) from the PDF of continuous rv \( X \) and the function \( g(.) \):
There is a general **two-step** procedure for deriving the distribution (i.e. PDF) of $Y$ (hence the name derived distribution):

1. Calculate the CDF $F_Y(y)$:

   $$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{\{x|g(x)\leq y\}} f_X(x)dx$$

2. Differentiate CDF $F_Y(y)$ to obtain the PDF $f_Y(y)$:

   $$f_Y(y) = \frac{d}{dy} F_Y(y)$$

**Ex:** Let $X$ be uniform on $[0, 1]$ and $Y = \sqrt{X}$. Derive distribution of $Y$, i.e. find $f_Y(y)$.

**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 $X$, $f_X(x)$.
4.1.1 Finding $f_Y(y)$ directly from $f_X(x)$

When function $g(.)$ has special properties, some formulas can be obtained for $f_Y(y)$.

**Linear case:** \[ Y = aX + b \quad \implies \quad f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right) \]

**Ex:** Linear function of a normal rv $X$ with mean $\mu_X$ and variance $\sigma_X^2$.

**Monotonic case:** \[ Y = g(X), \quad \text{where } y = g(x) \text{ is a strictly monotonic function with} \]
\[ \text{inverse } x = h(y) \]
\[ \implies \quad f_Y(y) = \frac{f_X(h(y))}{\left| \frac{d}{dx} g(x) \right|_{x=h(y)}} = f_X(h(y)) \left| \frac{d}{dy} h(y) \right| \]
Note: The formula above can be directly used when the function \( g(x) \) is monotonic in the support set of \( f_X(x) \) even though \( g(x) \) is not monotonic.

**Monotonic in pieces:** \( Y = g(X) \) where \( g(\cdot) \) is a general function such that for a value \( y_0 \), \( y_0 = g(x) \) has roots \( x_1, x_2, \ldots, x_n \)

\[
\Rightarrow \quad f_Y(y_0) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|\frac{d}{dx}(x)|_{x=x_i}}
\]

**Ex:** Let \( Y = g(X) = X^3, X \sim U[0, 1] \). Find \( f_Y(y) \).

**Ex:** Let \( Y = g(X) = X^2, X \sim U[-1, 1] \). Find \( f_Y(y) \).
**Ex:** (exercise) Consider random variables $X$ and $Y$ which are related as follows:

$$Y = g(X) = \begin{cases} 
1, & -2 \leq X < -1, \\
|X|, & -1 \leq X < +1, \\
(X - 2)^2, & +1 \leq X \leq +2 \\
0, & \text{otherwise}.
\end{cases}$$

a) If $X$ is uniformly distributed in $[0, 1]$, determine the PDF $f_Y(y)$ explicitly and plot it.

b) If $X$ is uniformly distributed in $[-1, 2]$, determine the PDF $f_Y(y)$ explicitly and plot it.

c) If $X$ is uniformly distributed in $[-2, 2]$,

i. determine the CDF $F_Y(y)$ explicitly and plot it.

ii. state whether $Y$ is a continuous or discrete random variable and briefly justify your answer.

### 4.1.2 Functions of Two Random Variables

Let $Z = g(X, Y)$ be a function of two random variables. $f_Z(z)$ can be determined using same/similar two-step procedure:

1. Calculate the CDF $F_Z(z)$:

$$F_Z(z) = P(Z \leq z) = P(g(X, Y) \leq z) = \int \int_{\{(x,y)|g(x,y)\leq z\}} f_{X,Y}(x, y) \, dx \, dy$$

2. Differentiate CDF $F_Z(Z)$ to obtain the PDF $f_Z(z)$:

$$f_Z(z) = \frac{d}{dz} F_Z(z)$$

**Ex:** Let $X$ and $Y$ be two independent continuous random variables and $Z = X + Y$. Show that, the PDF of $Z$ is given by the "convolution" of the PDFs $f_X(x)$ and $f_Y(y)$. 
Ex: Let $X$ and $Y$ be two \textbf{independent} discrete random variables and $Z = X + Y$. Show that, the PMF of $Z$ is given by the ”convolution” of the PMFs $p_X(x)$ and $p_Y(y)$.

Ex: Let $X$ and $Y$ be both exponentially distributed with parameter $\lambda$ and independent. Let $Z = \min\{X, Y\}$. Find the PDF of $Z$.

Ex: Let $X$ and $Y$ be independent uniform r.v.s in $[-1, 1]$. Compute the PDF of $Z = (X + Y)^2$. 


Ex: Let $X_1, X_2, X_3, \ldots$, 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 and in the limit the shape will be exactly the Gaussian PDF. It also turns out that the Gaussian PDF is a fixed point for convolution: convolving two Gaussian PDFs results in another Gaussian PDF.)

Figure 4.1: PDF of the sum of $n$ uniform random variables
4.2 Covariance and Correlation

The **covariance** of two random variables, $X$ and $Y$, is defined as:

\[ \text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \]

- It is a qualitative 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 $\text{cov}(X, Y) = 0$, we say $X$ and $Y$ are "uncorrelated".

- A **positive** (negative) covariance indicates that $(X - E[X])$ and $(Y - E[Y])$ tend to have the same (opposite) sign.
  - **positive covariance** $\Rightarrow$ $X$ and $Y$ tend to increase or decrease together.
  - **negative covariance** $\Rightarrow$ when $X$ increases, $Y$ tends to decrease, and vice versa.

- $\text{cov}(X, Y) = E[XY] - E[X]E[Y]$
  Pf

  \[ \text{cov}(X, X) = \]
  \[ \text{cov}(X, aY + b) = \]
  \[ \text{cov}(X, Y + Z) = \]

- Independence implies uncorrelatedness : (reverse not necessarily true)
  $X$ and $Y$ are indep. \[ E[XY] = E[X]E[Y] \quad \text{cov}(X, Y) = 0, \text{i.e. } X \text{ and } Y \text{ are uncorrel.} \]

- Correlatedness implies dependence : (derivable from above statement, reverse not necessarily true)
  $X$ and $Y$ are correlated \[ E[XY] \neq E[X]E[Y] \quad X \text{ and } Y \text{ are dependent} \]

- Dependence, uncorrelatedness do not necessarily imply something.
  - Summary of implications :
Ex: Discrete rvs $X$ and $Y$ uniformly distributed on the 4 points shown below.

- $\text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2) + 2\text{cov}(X_1, X_2)$
  Pf.

- $\text{var}(X_1 + X_2 + \ldots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \ldots + \text{var}(X_n) + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} \text{cov}(X_i, X_j)$
  Pf.

The **correlation coefficient** $\rho(X, Y)$ is a normalized form of covariance:

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

- $-1 \leq \rho(X, Y) \leq 1$ (pf based on the Schwartz inequality $(E^2[AB] \leq E[A^2]E[B^2])$)

- $\rho(X, Y) = +1 \iff Y = aX + b, \ a > 0$
\[ \rho(X, Y) = -1 \iff Y = aX + b, \ a < 0 \]

**Ex:** Consider \( n \) independent fair coin tosses with probability heads equal to \( p \). \( X \): number of heads, \( Y \)=number of tails. Find the correlation coefficient of \( X \) and \( Y \).

**Ex:** Let \( X_1, X_2, X_3, \ldots \), be a sequence of IID Bernoulli\((p)\) random variables ("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. Define another r.v. based on this process as \( Y_n = \sum_{i=1}^{n} X_i \).

(a) Find \( var(Y_n) \).

(b) Deviating from the Bernoulli process, let us assume that there is correlation between \( X_i \)'s:

\[ cov(X_i, X_j) = c|\!|i-j|\!|\sigma_X^2 \] with \( 1 > c > 0 \). Does the variance change?
4.3 Conditional Expectations Revisited (Iterated Expectations)

Remember the conditional expectation definitions:

- Hence, \( E[X|Y = y] \) is a function of \( y \):
- Then, introduce notation \( E[X|Y] \), which is a function of rv \( Y \):
- Hence, \( E[X|Y] \) is a **random variable** that takes value \( E[X|Y = y] \) when \( Y = y \).
- Calculate the **mean** of \( E[X|Y] \):

\[
E[X] = E[E[X|Y]] \quad \text{(Iterated Expectations)}
\]

**Ex:** Break a stick uniformly at random. Keep the piece that contains the left end. Now, break this piece and let \( X \) be the length of the piece that contains the left end. Find \( E[X] \).

Note that the same idea can be extended to expected value of a function of \( X \):

\[
E[g(X)] = E[E[g(X)|Y]]
\]
Ex: Same stick breaking example. Find \( \text{var}(X) \).

4.4 Transforms (Moment Generating Functions)

Transforms often provide us convenient ways to do certain mathematical manipulations.

The transform or moment generating function (MGF) of a rv \( X \) is defined by

\[
M_X(s) = E[e^{sX}] = \begin{cases} 
\text{, } X \text{ continous} \\
\text{, } X \text{ discrete}
\end{cases}
\]

Ex: Find MGF of \( X \) which has PMF given by

\[
p_X(x) = \begin{cases} 
1/2, & x = 2 \\
1/6, & x = 3 \\
1/3, & x = 5.
\end{cases}
\]

Ex: MGF of an exponential rv

Ex: MGF of a linear function of a rv
4.4.1 From Transforms to Moments

The name moment generating function follows from the following property.

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

Proof:

Ex: For discrete rv \( X \) in previous question, find first and second moments.

Ex: Mean and variance of an exponential rv.

Ex: The third moment \( (E[X^3]) \) of a standard normal (Gaussian) rv.
4.4.2 Inversion of Transforms

**Inversion Property:** The transform $M_X(s)$ associated with a r.v. $X$ uniquely determines the CDF of $X$, assuming that $M_X(s)$ is finite for all $s$ in some interval $[-a, a]$, where $a$ is a positive number.

- Transform $M_X(s)$ of $X$ uniquely determines the CDF (and therefore PDF or PMF) of $X$.
- Thus given $M_X(s)$ of $X$, one can invert it to find the PDF/PMF of $X$.
  - There are formulas for inversion of transforms which are usually difficult to use.
  - Transforms are rather inverted by *pattern matching* based on tables.

**Ex:** The transform associated with a r.v. $X$ is given as $M_X(s) = \frac{1}{3} e^{-2s} + \frac{1}{2} + \frac{1}{6} e^{3s}$. What is the corresponding distribution?

**Ex:** Find the distribution of $X$ for $M_X(s) = \frac{pe^s}{1-(1-p)e^s}$.

**Ex:** Find distribution for $M_X(s) = \frac{16 - 4s + 8 - 4/3s}{(6 - s)(4 - s)}$.

4.4.3 Sums of Independent R.V.s

Let $Z = X + Y$ where $X$ and $Y$ are independent. Then MGF of $Z$ is
Similarly, if $Z = X_1 + X_2 + ... + X_n$ where $X_i$ are independent, then MGF of $Z$ is

**Linear function of independent Gaussian rvs:** Linear function of independent Gaussian rvs is also Gaussian.

**Ex:** Show above property of independent Gaussian rvs.

**Ex:** Sum of independent Poisson rvs

### 4.4.4 Transforms Associated with Joint Distributions

Consider rvs $X_1, X_2, \ldots, X_n$ with a joint PDF. Their MGF (or multivariate transform) is then defined by

$$M_{X_1, X_2, \ldots, X_n}(s_1, s_2, \ldots, s_n) = E[e^{s_1 X_1 + s_2 X_2 + \ldots + s_n X_n}].$$

### 4.5 Sum of a Random Number of Independent R.V.s

Consider the sum $Y$ of random number $N$ of independent random variables $X_i$:

$$Y = X_1 + X_2 + \ldots + X_N$$

**N:**
$X_1, X_2, \ldots, X_i$ are iid
$N, X_1, X_2, \ldots, X_i$ are independent
Show the following equalities, which relate the mean, variance and the transform of $Y$ to similar quantities of $X_i$ and $N$:

1. $E[Y] = E[X]E[N]$
2. $var(Y) = E[N]var(X) + var(N)(E[X])^2$
3. $M_Y(s) = M_N(log(M_X(s)))$

- $E[Y] = E[X]E[N]$
- $var(Y) = E[N]var(X) + var(N)(E[X])^2$
Ex: Behzat visits a number of bookstores, looking for *The Loneliness of the Long Distance Runner*. Any given bookstore carries the book with probability $p$, independent of the others. In a typical bookstore visited, Behzat spends a random amount of time, exponentially distributed with parameter $\lambda$, until he finds the book or determines that the bookstore doesn’t carry it. We assume that Behzat will keep visiting bookstores until he buys the book and that the time spent in each is independent of everything else. We wish to find the mean, variance, and PDF of the total time spent in this quest.
Chapter 5

Limit Theorems

Contents

5.1 Markov and Chebychev Inequalities .............................................. 97
5.2 The Weak Law of Large Numbers (WLLN) .................................... 99
5.3 Convergence in Probability ......................................................... 100
5.4 The Central Limit Theorem (CLT) .............................................. 101

5.1 Markov and Chebychev Inequalities

Markov Inequality: If a random variable $X$ takes only nonnegative values, then

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Proof:

**Ex:** Use Markov inequality to bound the probability that a standard normal random variable exceeds twice its standard deviation, in absolute value.
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 the following.

**Chebychev Inequality:** For random variable $X$ with mean $E[X]$ and variance $\sigma^2$,

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

Proof:

Note that Chebychev’s inequality uses more information about $X$ than the Markov inequality and thus can provide a tighter bound about the probabilities related to $X$.

- In addition to the mean (a first-order statistic), Chebychev’s inequality 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, an exponential r.v. with mean 1 and variance 1, and a discrete random variable that takes on the values 0.9, and 1.1 equally probably. Markov inequality does not distinguish between these distributions, whereas Chebychev inequality does.)

Ex: Apply Chebychev inequality instead of Markov in the two examples above.
5.2 The Weak Law of Large Numbers (WLLN)

**WLLN:** Consider a sequence $X_1, X_2, \ldots$ of IID (independent and identically distributed) random variables, with $E[X_i] = \mu$, and $var(X_i) = \sigma^2$. Then the sample mean $M_n$ defined by

$$M_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

satisfies the following expression for any $\epsilon > 0$:

$$P(|M_n - \mu| \geq \epsilon) = P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| \geq \epsilon\right) \to 0 \text{ as } n \to \infty.$$

Proof: (apply the Chebychev’s inequality to sample mean $M_n$)

• The WLLN asserts that the sample mean $M_n$ of a large number of iid rvs is very close to the true mean $E[X] = \mu$, with high probability.

  - That is, consider interval $[\mu - \epsilon, \mu + \epsilon]$ around $\mu$, then the probability that $M_n$ is outside this interval goes to zero as $n$ increases (i.e. ”almost all” of the probability of PDF/PMF of $M_n$ is concentrated within $\epsilon$ of $\mu$ for large values of $n$.)

• If we refer to $\epsilon$ as **accuracy level** and $\delta = \frac{\sigma^2}{n\epsilon^2}$ as the **confidence level** : for any given level of accuracy and confidence, sample mean $M_n$ will be equal to $\mu$, within these levels of accuracy and confidence, by just choosing $n$ sufficiently large.

**Ex:** (Probabilities and frequencies) An event $A$ with probability $p = P(A)$. The empirical frequency of $A$ in $n$ repetitions is the fraction of time event $A$ occurs.
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?

5.3 Convergence in Probability

Convergence of a deterministic sequence: Let $a_1, a_2, ...$ be a sequence of real numbers and let $a$ be another real number. We say that the sequence $a_n$ converges to $a$, or $\lim_{n \to \infty} a_n = a$, if for every $\epsilon > 0$ there exists some $n_0$ such that

$$|a_n - a| \leq \epsilon \quad \text{for all } n \geq n_0.$$  

- Intuitively, if $\lim_{n \to \infty} a_n = a$, then the magnitude difference between $a$ and $a_n$ can be made smaller than any desired accuracy level $\epsilon$ by just choosing $n$ sufficiently large.

Convergence in probability: Let $Y_1, Y_2, ...$ be a sequence of random variables (not necessarily independent), and let $a$ be a real number. We say that the sequence $Y_n$ converges to $a$ in probability if for every $\epsilon > 0$ and $\delta > 0$ we have

$$P(|Y_n - a| \geq \epsilon) \leq \delta \quad \text{for all } n \geq n_0.$$  

$$\left( \lim_{n \to \infty} P(|Y_n - a| \geq \epsilon) = 0 \right)$$
• If $Y_n$ converges in probability to $a$, this means that ”almost all” of the probability of PDF/PMF of $Y_n$ is concentrated within $\epsilon$ of $a$ for large values of $n$.

• If we refer to $\epsilon$ as **accuracy level** and $\delta$ as the **confidence level** : If $Y_n$ converges in probability to $a$, then for any given level of accuracy and confidence, $Y_n$ will be equal to $a$, within these levels of accuracy and confidence, by just choosing $n$ sufficiently large.

• The WLLN states that the sample mean sequence $M_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$ converges in probability to $\mu$.

# 5.4 The Central Limit Theorem (CLT)

**CLT:** Let $X_1, X_2, \ldots$ be a sequence of IID random variables with $E[X_i] = \mu$, $var(X_i) = \sigma^2$, and define

$$S_n = X_1 + X_2 + \ldots + X_n$$

and also a normalized version of it

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$  

(Note that

- $E[S_n] = \mu$, $var(S_n) = \sigma^2$
- $E[Z_n] = 0$, $var(Z_n) = 1$.)

Then, the CDF of $Z_n$ converges to the standard normal CDF $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ in the sense that

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

for every $z$.

• **CLT** suggests that the **sum of a large number of IID rvs is approximately normal.**

  - **CLT approximation :** $Z_n \sim \Phi(z)$ or equivalently $S_n = (X_1 + X_2 + \ldots X_n) \sim \mathcal{N}(n\mu, n\sigma^2)$

• This is a very common situation where a random effect originates from many small random factors. Practically, CLT simplifies probabilistic elaboration of many problems.

• 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.)

**Simple Proof of the Central Limit Theorem:** We will use transforms. For simplicity, assume $E[X] = 0$, $\sigma = 1$. Consider the transform of $Z_n$, $M_{Z_n}(s)$ when $n \to \infty$. Recall that $e^z = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$. Note that $\frac{X}{\sqrt{n}}$ is very close to 0 for large $n$.  

101
\[
M_{Z_n}(s) = E\left[e^{\frac{S_1 + S_2 + \ldots + S_n}{\sqrt{n}}}ight] = E\left[e^{\frac{\sum S_i}{\sqrt{n}}}ight] = \prod_{i=1}^{n} E\left[e^{\frac{X_i}{\sqrt{n}}}ight]
\]

\[
= \prod_{i=1}^{n} E\left[1 + s\frac{X_i}{\sqrt{n}} + \frac{s^2 X_i^2}{2n} + \frac{s^3 X_i^3}{3!n^{3/2}} + \ldots\right]
\]

\[
\lim_{n \to \infty} M_{Z_n}(s) = \prod_{i=1}^{n} \left(1 + \frac{s^2/2}{n}\right) = \left(1 + \frac{s^2/2}{n}\right)^n = e^{s^2/2}
\]

(Last equality follows from: \(\lim_{n \to \infty} (1 + \frac{x}{n})^n = e^x\))

Since the transform \(M_{Z_n}(s)\) converges to that of a standard Gaussian, and there is a one-to-one mapping between transforms and distributions, we conclude that CDF of \(Z_n\) converges to that of a standard Gaussian.

**Ex:** You download 300 mp3 files from a website each month where your limit in MBytes is 1700. The mp3 file sizes are uniformly distributed in [3, 8] and independent from the other file sizes. What is the probability that you have to pay extra in a month?

**Ex:** Repeat polling question with CLT approximation.

**Ex:** Promoting a new product, a grocery store has an eye-catching stand which makes people buy the product with probability 3/10. Assuming that the purchasing behavior of the 100 people who visited the store are independent, what is the probability that less than 40 items are sold?