# STAT 230 - Probability

Cameron Roopnarine

Last updated: June 5, 2020

# Contents

# Chapter 1

# Introduction to Probability

## 1.1   Definitions of Probability

We can define probability in three different ways:

### 1.1.1   Definition (Probability)

1. *Classical* definition:
$$P(\text{event}) = \frac{\text{\# ways the event can occur}}{\text{\# of all possible outcomes}}$$
this requires the outcomes to all be equally likely.
2. *Relative frequency* definition:

$$P(\text{event}) = \text{proportion of the time the event occurs in repeated experiments}$$

this requires the same conditions for each observation.
3. *Subjective probability* definition:

$$P(\text{event}) = \text{how certain we are that the event will occur}$$

However, all three of these definitions have serious limitations.

# Chapter 2

# Chapter 2: Mathematical Probability Models

## 2.1 Sample Spaces and Probability

We need a mathematical model to define probability.

### 2.1.1 Definition (Experiment, Trial, Outcome)

We define an *experiment* as a process that can be repeated with multiple possible results. We define a *trial* as a single repetition of an experiment. We define an *outcome* as the results on one trial of an experiment.

### 2.1.2 Definition (Set)

A set is a collection of well defined and distinct objects.

**REMARK 2.1.1.** A set is an unordered list with no repetition.

### 2.1.3 Definition (Sample Space)

A *sample space* $S$ is a set of distinct outcomes for an experiment or process, with the property that in a single trial, one and only one of these outcomes occur.

Sample spaces can be *discrete* or *non-discrete*.

### 2.1.4 Definition (Discrete)

Let $S$ be a sample space. We say $S$ is *discrete* if it consists of a finite or countably infinite set of simple events.

**Example**

Roll a fair 6-sided die repeatedly. Determine some possible sample spaces for this experiment.

5

**Solution.**

Sample space:

$$S_1 = \{1, 2, 3, 4, 5, 6\} \star \text{ easiest to work with}$$
$$S_2 = \{\text{odd, even}\}$$
$$S_3 = \{\text{prime, non-prime}\}$$
$$S_4 = \{6, \text{not } 6\} \star \text{ outcomes don't have to be equally likely in a sample space}$$
$$S_5 = \{\text{a number}\}$$

$\star$ need not have equally likely outcomes

### 2.1.5   Definition (Event, Simple Event, Compound Event)

Let $S$ be a discrete sample space. An *event* in a discrete sample space is a subset $A \subseteq S$. If the event is indivisible so it contains only one point, e.g. $A_1 = \{a_1\}$ we call it a *simple event*. An event $A$ made up of two or more simple events such as $A = \{a_1, a_2\}$ is called a *compound event*.

**REMARK 2.1.2.** The notation $A \subseteq S$ means $a \in A \implies a \in B$.

**Example**

Let $A = $ a 5 is rolled. Let $B = $ an odd # is rolled. Determine which of the events are simple events and compound events.

**Solution.**

$A = \{5\} \subseteq S$, $B = \{1, 3, 5\} \subseteq S$. Thus, $A$ is a simple event and $B$ is a compound event.

When the trial is conducted, the outcome determines which events occur.

If outcome is in the set, it occurs

5 rolled $\rightarrow A$ and $B$ both occur

3 rolled $\rightarrow A$ does not occur, $B$ occurs

2 rolled $\rightarrow$ neither events occur

### 2.1.6   Definition (Probability, Probability Distribution)

Let $S = \{a_1, \ldots\}$ be a discrete sample space $S$. Assign numbers (*probabilities*) $P(a_i)$ for $i = 1, \ldots$ to the $a_i$'s such that the following two conditions hold:
  (1) $0 \le P(a_i) \le 1$
  (2) $\sum\limits_{\text{all } i} P(a_i) = 1$
The set of probabilities $\{P(a_i), i = 1, \ldots\}$ is called a *probability distribution* on $S$.

### 2.1.7   Definition (Probability of an Event)

The *probability $P(A)$ of an event* $A$ is the sum of the probabilities for all the simple events that make up $A$ or
$$P(A) = \sum_{a \in A} P(a)$$

If a sample space $S$ has equally likely outcomes then,

$$P(\text{simple event}) = \frac{1}{|S|}$$

$$P(A) = \sum_{a \in A} P(a_i) = \frac{|A|}{|S|}$$

$$P(A) = \sum_{a \in A} P(a_i) = \frac{|A|}{|S|}$$

# Chapter 3

# Probability and Counting Techniques

## 3.1 Addition and Multiplication Rules

We need a systematic way to count outcomes without listing them.

### 3.1.1 Counting Rules

There are two basic counting rules:
1. The *Addition* Rule: Suppose we can do job $1$ in $p$ ways and job $2$ in $q$ ways. Then we can do either job 1 OR job 2 (but not both), in $p + q$ ways.
2. The *Multiplication* Rule: Suppose we can do job $1$ in $p$ ways and job $2$ in $q$ ways. Then we can do both job 1 AND job 2 in $p \times q$ ways.

## 3.2 Counting Arrangements or Permutations

**Sampling with replacement**: it is possible to obtain the same result more than once, e.g. die rolls, coin flip, slot machine, password.

**Sampling without replacement**: once a result occurs, it cannot happen again. e.g. drawing cards, balls from an urn, eating candy of different colour.

### 3.2.1 Definition (Permutation)

A *permutation* is an ordered selection of $k$ objects chosen from $n$ objects.
If we select the objects above without replacement, we write

$$^{n}P_{k} = n(n-1) \cdots (n-k+1) = n^{(k)}$$

If we select the objects above with replacement, we write

$$n^{k}$$

$n^{(0)} = 1$

$n^{(n)} = n!$

$k > n \to 0$ not possible

**Example**

IP addresses: an ordered sequence of four numbers between $0$ and $255$. e.g. 192.168.1.1, 129.97.95.107, etc. Determine the total possible outcomes with and without replacement.

**Solution.**

Since order matters, we are immediately looking at a permutation.

With replacement: $256^4$

Without replacement: $256^{(4)}$

---

<div align="center">

LECTURE 4

</div>

---

Always ask:

1. Can you get the same object twice?

- Yes $\to n^k$

- No $\to$ Step 2.

2. Does the order matter?

- Yes $\to n^{(k)}$

- No $\to$ today's lesson $\binom{n}{k}$

**Examples**

4 numbers between $0$ and $255$

- total possible: $256^4$
- all odd numbers $128^4$ so $P(\text{all odd}) = {}^1/_{16}$
- at least one odd number; work with the opposite: all even: $128^4$, so $P(\text{at least one odd}) = 1 - {}^{128^4}/_{256^4}$

**Example**

5 people $A$, $B$, $C$, $D$ 4 co-op jobs $1$, $2$, $3$, $4$

Find the probability that $A$ gets a job.

**Solution.**

1. order matters $\to$ (permutation of some sort)
2. 1 + without replacement $\to n^{(k)}$

so total ways is $5^{(4)} = 120$

$A, \_, \_, \_$ or $\_, A, \_, \_$ or $\_, \_, A, \_$ or $\_, \_, \_, A$
$4^{(3)} = 96$

So probability they do is $\frac{96}{120} = 0.8$

Alternatively, # of ways for $A$ to not get a job is $4^{(4)}$ or 4! (they are the same quantity). So probability they do is $1 - {}^{4!}/_{120} = 0.8$.

Intuitively this makes sense because each of the $5$ is equally likely (${}^1/_5$) to not get a job.

Find probability that $B$ and $C$ get adjacent jobs.

$BC, \_, \_$ or $\_, BC, \_$ or $\_, \_, BC$

$CB, \underbrace{\_,\_}_{3^{(2)}}$ or $\_, CB, \_$ or $\_, \_, CB$

So total ways is $6 \times 3^{(2)} = 36$, probability$= {}^{36}/_{120}$. Alternatively, treat $BC$ as one unit with $2$ ways it can look (BC or CB).

## 3.3   Counting Subsets or Combinations

### 3.3.1   Definition (Combination)

A *combination* is an unordered selection of $k$ objects chosen from $n$ objects.
If we select the objects above without replacement, we write

$$ {}^{n}C_k = \frac{n!}{k!(n-k)!} = \binom{n}{k} $$

How many ways? If we did care, $n^{(k)}$. Then deliberately forgot the order.

e.g. select $3$ digits $0-9$

$$ \{8, 2, 1\} $$

if we care about the order, each set is counted $3! = 6$ times.

So, there are ${}^{10^{(3)}}/_{3!} = 120$ possible sets of $3$ digits.

This quantity is called

$$ \binom{n}{k} = \frac{n^{(k)}}{k!} = \frac{n!}{k!(n-k)!} = {}^{n}C_k $$

"$n$ choose $k$", "binomial coefficient", "$\binom{n}{k}$ is the $k^{th}$ element of the $n^{th}$ row of Pascal's $\Delta$"

**Example**

Lotto $6/49$, choose $6$ winning # from $49$. The order of the numbers does not matter.

$$ \binom{49}{6} \text{ ways } \approx 13.9 \text{ million} $$

---

LECTURE 5

---

### 3.3.2   Theorem (Properties of Combinations)

Let $n$, $k \in \mathbb{Z}$ be non-negative.
1. $n^{(k)} = \frac{n!}{(n-k)!} = n(n-1)^{(k-1)}$ for $k \geq 1$
2. $\binom{n}{k} = \frac{n^{(k)}}{k!} = \frac{n!}{k!(n-k)!}$
3. $\binom{n}{k} = \binom{n}{n-k}$ for all $k = 1, \ldots, n$
4. If we define $0! = 1$, then $\binom{n}{0} = \binom{n}{n} = 1$
5. Pascal's Identity: $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$
6. Binomial Theorem: $(1+x)^n = \binom{n}{0} + \cdots + \binom{n}{n}x^n$

**Example**

Lotto $6/49$ $\binom{49}{6}$ possible sets of winning #'s. If your ticket contains all $6$ winning #'s you win.

$$P(\text{win}) = \frac{\binom{6}{6}\binom{43}{0}}{\binom{49}{6}}$$

$$P(\text{match } 5) = \frac{\binom{6}{5}\binom{43}{1}}{\binom{49}{6}}$$

**Example**

Suppose you select $5$ cards from $52$ ($13$ cards of each $4$ suits). Find the probability of $3$ of one rank, $2$ of another rank.

Total # of hands: $\binom{52}{5}$. # with $3$ of one, $2$ of another:

$$\underbrace{\text{rank of triple}}_{13} \times \binom{4}{3} \times \underbrace{\text{rank of pair}}_{12} \times \binom{4}{2}$$

## 3.4 Number of Arrangements When Symbols Are Repeated

Suppose we have $5$ objects, $2$ of which are alike:

$$\text{D I A N A}$$

If we arrange the objects in order, how many results can we get?

If we could tell them apart, then we have $5!$ ways, but every possible arrangement has a matching one with A's flipped.

So, $5!/2! = 60$ ways (removing double counting)

In general, if we have $n$ objects:

$$\left. \begin{array}{ccc} n_1 & \text{of} & \text{type } 1 \\ & \vdots & \\ n_k & \text{of} & \text{type } k \end{array} \right\} n_1 + \cdots + n_k = n$$

How many ways can the objects be arranged, where objects of the same type are identical?

$STATISTICS$: $n = 10$

$$\begin{array}{cc} S : 3 & n_1 \\ T : 3 & n_2 \\ A : 1 & n_3 \\ I : 2 & n_4 \\ C : 1 & n_5 \end{array}$$

ways to place:

$$\begin{array}{cc} S & \binom{10}{3} \\ T & \binom{7}{3} \\ A & \binom{4}{1} \\ I & \binom{3}{2} \\ C & \binom{1}{1} \end{array}$$

So in total there are

$$\binom{10}{3}\binom{7}{3}\binom{4}{1}\binom{3}{2}\binom{1}{1} = \frac{10!}{3!7!}\frac{7!}{3!4!}\frac{4!}{1!3!}\frac{3!}{2!1!}\frac{1!}{1!0!}$$
$$= \frac{10!}{3!3!2!}$$

In general,

$$\frac{n!}{n_1! \cdots n_k!}$$

---

## LECTURE 6

**Example**

7 Pokémon Go players, ($2$ M, $2$ I, $3$ V) are ranked $1 - 7$. Find the probability that $1$ and $7$ are on different teams.

Total # rankings: $\frac{7!}{2!2!3!} = 210$.

$M, \underbrace{\_,\_,\_,\_,\_}_{2I,3V}, M$: $\frac{3!}{2!3!} = 10$

$I, \underbrace{\_,\_,\_,\_,\_}_{2M,3V}, I$: $\frac{3!}{2!3!} = 10$

$V, \underbrace{\_,\_,\_,\_,\_}_{2M,2I,1V}, V$: $\frac{3!}{2!2!1!} = 50$

$210 - 50 = 160/210 < -\text{total}$

# Chapter 4

# Probability Rules and Conditional Probability

## 4.1 General Methods

$A \cup B$
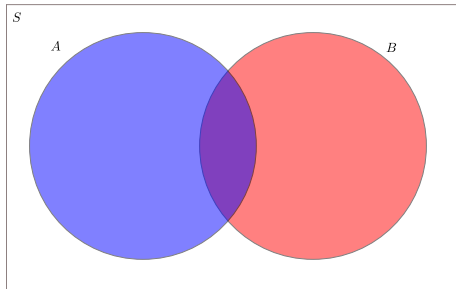


$A \cap B$
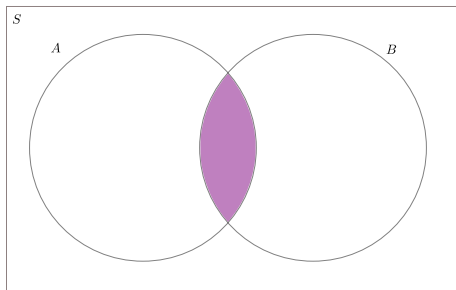


$\overline{A \cup B} = \overline{A} \cap \overline{B}$
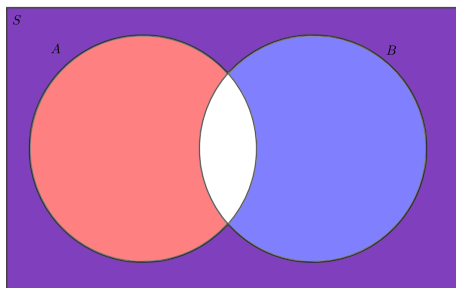
### 4.1.1 Theorem (De Morgan's Laws)

(1) $\overline{A \cup B} = \bar{A} \cap \bar{B}$
(2) $\overline{A \cap B} = \bar{A} \cup \bar{B}$

## 4.2 Rules for Unions of Events

### 4.2.1 Rule 4a (Addition Law of Probability or the Sum Rule)

Let $A$ and $B$ be any events. Then

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

### 4.2.2 Rule 4b (Probability of the Union of Three Events)

Let $A$, $B$ and $C$ be any events. Then

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

**Example**

$P(J) = {}^{19}/_{22}$

$P(C) = {}^{7}/_{22}$

$P(\text{neither}) = {}^{2}/_{22}$

Find $P(JC)$.

**Solution.**

$$\frac{(19 - x) + x + (7 - x) + 2}{22} = 1 \implies x = 6$$

This relied on the fact that the regions were non-overlapping so we could add them up.

### 4.2.3 Definition (Mutually Exclusive)

Events $A$ and $B$ are *mutually exclusive* if

$$A \cap B = \emptyset \text{ (the empty set)}$$

### 4.2.4 Rule 5a (Probability of the Union of Two Mutually Exclusive Events)

Let $A$ and $B$ be mutually exclusive events. Then

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

### 4.2.5  Rule 5c (Probability of the Union of n Mutually Exclusive Events)

Let $A_1, \ldots, A_n$ be mutually exclusive events. Then

$$P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i)$$

### 4.2.6  Rule 6 (Probability of the Complement of an Event)

For any event $A$,
$$P(A) = 1 - P(\bar{A})$$

*Proof.* $A$ and $\bar{A}$ are mutually exclusive. By rule 5a we have

$$A \cup \bar{A} = P(A) + P(\bar{A})$$

But since $P(A \cup \bar{A}) = P(S) = 1$,

$$1 = P(A) + P(\bar{A}) \implies P(A) = 1 - P(\bar{A})$$

$\square$

---

<center>LECTURE 7</center>

---

Roll two fair 12-sided die. What is the probability at least one of them is greater than 7.

$$
\begin{aligned}
1 - P(\text{neither}) &= 1 - P(\overline{A \cap B}) \\
&= 1 - P(\bar{A} \cup \bar{B}) \\
&= 1 - \frac{7}{12}\frac{7}{12}
\end{aligned}
$$

In this example, we relied on the multiplication rule to find a probability on both events, but this requires the events to not influence each other.

## 4.3  Intersections of Events and Independence

### 4.3.1  Definition (Independent, Dependent)

$A$ and $B$ are *independent* if and only if

$$P(AB) = P(A)P(B)$$

If the events are not independent, we call the events *dependent*.

We can use this in two ways.

1. If we know both events are independent, we can calculate $P(AB)$.

2. 2. Calculate/estimate all three probabilities and check whether independent. e.g. treatment vs recover, smoking vs cancer, income vs politics

Note for $A$, $B$, $C$ to be independent, we need

$$P(AB) = P(A)P(B)$$
$$P(BC) = P(B)P(C)$$
$$P(AC) = P(A)P(C)$$
$$P(ABC) = P(A)P(B)P(C)$$

**Example**

Roll $2$ fair $6$-sided dice. Let $A = $ the first die is $3$. Let $B = $ the total is $7$. Are $A$ and $B$ independent?

$$P(A) = \frac{1}{6}$$
$$P(B) = \frac{6}{36}$$
$$P(AB) = \frac{1}{36}$$
$$P(AB) = \frac{1}{36} = \frac{1}{6}\frac{1}{36} = P(A)P(B)$$

Now, let $C = $ the total is $8$. Are $A$ and $C$ independent?

$$P(C) = \frac{5}{36}$$
$$P(AC) = \frac{1}{36}$$
$$P(AC) = \frac{1}{36} \neq \frac{1}{6}\frac{5}{36} = P(A)P(C)$$

Why? With $7$ there is always a possible second roll, but with $8$ it's not always possible (e.g. if the first die was a $1$).

**Independence vs. Mutual Exclusive**

|       | ME | ID | Both |
|-------|----|----|------|
| math  | $AB = \emptyset$, $P(AB) = 0$ | $P(AB) = P(A)P(B)$ | $P(A)P(B) = 0$ |
| logic | both can't happen | one doesn't affect the other | at least one is impossible |

If events are dependent, we might want to quantify the effect of one on the other.

## 4.4 Conditional Probability

### 4.4.1 Definition (Conditional Probability)

The *conditional probability* of $A$, given $B$ is

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

provided $P(B) > 0$.

Why? The classical definition of probability:

$$\frac{\#\text{ ways }A\text{ can occur}}{\#\text{ ways }B\text{ can occur}}$$

Since we need to restrict $S$ to just be $B$,

$$P(A|C) = \frac{P(AC)}{P(C)} = \frac{\frac{1}{36}}{\frac{5}{36}} = \frac{1}{5} > P(A)$$

or $C : \{(2,6),(3,5),(4,4),(5,3),(6,2)\}$

$$P(C|A) = \frac{P(CA)}{P(A)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6} > P(C)$$

or $A : \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\}$

Dependence is a two way relationship. Both influence the other in the same direction.

---

<div align="center">LECTURE 8*</div>

---

## 4.5   Product Rules, Law of Total Probability and Bayes' Theorem

### 4.5.1   Rule 7 (Product Rules)

Let $A$, $B$, $C$, $D,\ldots$ be events in a sample space. Assume that $P(A) > 0$, $P(AB) > 0$, and $P(ABC) > 0$. Then

$$P(AB) = P(B|A)P(A)$$
$$P(ABC) = P(C|AB)P(AB)$$
$$P(ABCD) = P(D|ABC)P(ABC)$$

and so on.

### 4.5.2   Rule 8 (Law of Total Probability)

Let $A_1,\ldots,A_k$ be a partition of the sample space $S$ into disjoint (mutually exclusive events), that is

$$A_1 \cup \cdots A_k = S$$

and

$$A_i \cap A_j = \emptyset \qquad i \neq j$$

Let $B$ be an arbitrary event in $S$. Then

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

### 4.5.3   Theorem (Bayes' Theorem)

Suppose $A$ and $B$ are events defined on a sample space $S$. Suppose also that $P(B) > 0$. Then

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

## LECTURE 9

We might be interested in reversing the direction of a conditional probability.

- given a positive test, what is the probability that you have a disease?
- given an error in code, who wrote it?

**Example**

1. If you test for a disease, what is the probability that you have it?

$P(D) = 0.02$

$P(T|\bar{D}) = 0.05$ false positive

$P(\bar{D}|T) = 0.01$ false negative

We found $P(T) = 0.0688$, we want

$$P(D|T) = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})}$$
$$= \frac{0.99 \times 0.02}{0.99 \times 0.02 + 0.05 \times 0.98}$$
$$= 0.288$$

2. Given a line of code that has an error, what is the probability that $A$ wrote it?

$P(A) = 0.5$

$P(B) = P(C) = 0.25$

$P(E|A) = 0.01$

$P(E|B) = 0.02$

$P(E|C) = 0.05$

We want

$$P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|B)P(B) + P(E|C)P(C)}$$
$$= \frac{0.5 \times 0.01}{0.0225}$$
$$= 0.222 < P(A)$$

Similarly,

$$P(B|E) = 0.22 < P(B)$$
$$P(C|E) = 0.556 > P(C)$$

Note the three conditional probabilities sum to $1$ which they must since exactly one of $A$, $B$, or $C$ wrote the line.

3. Probability of a LoL player also playing Warcraft?

$$P(W|L) = \frac{P(L|W)P(W)}{0.0387} = \text{exercise}$$

## 4.6 Useful Series and Sums

### 4.6.1 Theorem (Geometric Series)

The geometric series $\sum\limits_{n=0}^{\infty} ar^n$ converges if $|r| < 1$ and diverges otherwise. If $|r| < 1$, then

$$\sum_{n=0}^{\infty} ar^n = a + ar + ar^2 + \cdots = \frac{a}{1-r}$$

### 4.6.2 Theorem (Binomial Theorem)

Let $n$ be a positive integer, $x \in \mathbb{R}$.

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$$

### 4.6.3 Theorem (Exponential Series)

Let $x \in \mathbb{R}$.

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

$$e^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$

### 4.6.4 Theorem (Hypergeometric Identity)

$$\binom{a+b}{n} = \sum_{x=0}^{n} \binom{a}{x}\binom{b}{n-x}$$

---

<div align="center">

LECTURE 10

</div>

---

**MLIW 3: Naïve Bayes' Classifier**

In ML classification, we use evidence to decide what category something belongs to. That is,

$$P(\text{category} \mid \text{evidence}) = \frac{P(\text{cat})P(\text{evidence} \mid \text{cat})}{\sum\limits_{\text{cat } i} P(\text{cat } i)P(\text{evidence} \mid \text{cat } i)}$$

The Naïve Bayes' Classifier assumes that if there are multiple pieces of evidence, they are conditionally independent (conditional on the category).

A very simple ML example of this is spam detection. Consider the machine learning problem of classifying incoming messages as spam. We define:

- $A_1 =$ message fails rDNS (reverse DNS lookup) check (i.e. the "from" domain doesn't match)
- $A_2 =$ message is sent to over $100$ people
- $A_3 =$ message contains a link with the URL not matching the alt

We will assume that the $A_i$'s are independent events, given that a message is a spam, and that they are also independent events, given that a message is $\overline{\text{spam}}$.

We estimate

- $P(\text{Spam}) = 0.25$
- $P(A_1 \mid \text{Spam}) = 0.3$
- $P(A_2 \mid \text{Spam}) = 0.2$
- $P(A_3 \mid \text{Spam}) = 0.1$
- $P(A_1 \mid \overline{\text{Spam}}) = 0.005$
- $P(A_2 \mid \overline{\text{Spam}}) = 0.04$
- $P(A_3 \mid \overline{\text{Spam}}) = 0.05$

We use Bayes' Theorem:

$P(\text{Spam} \mid A_1 A_2 A_3) =$

$$
= \frac{P(A_1 A_2 A_3 \mid \text{Spam})P(\text{Spam})}{P(A_1 A_2 A_3 \mid \text{Spam})P(\text{Spam}) + P(A_1 A_2 A_3 \mid \overline{\text{Spam}})P(\overline{\text{Spam}})}
$$

$$
= \frac{P(A_1 \mid \text{Spam})P(A_2 \mid \text{Spam})P(A_3 \mid \text{Spam})P(\text{Spam})}{P(A_1 \mid \text{Spam})P(A_2 \mid \text{Spam})P(A_3 \mid \text{Spam})P(\text{Spam}) + P(A_1 \mid \overline{\text{Spam}})P(A_2 \mid \overline{\text{Spam}})P(A_3 \mid \overline{\text{Spam}})P(\overline{\text{Spam}})}
$$

$$
= \frac{(0.3)(0.2)(0.1)(0.25)}{(0.3)(0.2)(0.1)(0.25) + (0.005)(0.04)(0.05)(0.75)}
$$

$$
\approx 0.9950
$$

Remember that $A_1$, $A_2$, and $A_3$ are NOT independent! They are only conditionally independent, given the type of email.

# Chapter 5

# Random Variables

## 5.1   Random Variables and Probability Functions

<div style="background-color:yellow">

### 5.1.1   Definition (Random Variable)

A *random variable* is a function that assigns a real number to each point in a sample space $S$.

</div>

We typically use $X$, $Y$, $Z$ as random variables and $x$, $y$, $z$ as the possible values the random variable can take on. There are two types of random variables based on the range.

<div style="background-color:yellow">

### 5.1.2   Definition (Discrete Random Variables)

*Discrete random variables* take integer values or, more generally, values in a countable set.

</div>

<div style="background-color:yellow">

### 5.1.3   Definition (Continuous Random Variables)

*Continuous random variables* take values in some interval of real numbers like $(0, 1)$ or $(1, \infty)$ or $(-\infty, \infty)$.

</div>

We can define multiple random variables on the same sample space $S$. For example, roll a fair 6-sided die 3 times.
$$S = \{(x, y, z) | 1 \leq x, y, z \leq 6\}$$

- Let $X$ = sum on the three die. $range(X) = \{3, \dots, 18\}$.
- Let $Y$ = product on the three die. $range(Y) = \{1, \dots, 216\}$.
- Let $Z$ = number on the first die. $range(Z) = \{1, \dots, 6\}$
- Let $\bar{X}$ = average. $range(\bar{X}) = \{1, {}^4\!/\!{}_3, \dots, 6\}$
- Let $W$ = # of dice that are $1$. $range(W) = \{0, 1, 2, 3\}$

Examples of continuous random variables include:

- $T$ = time until an event
- $P$ = positive in space of a particle
- $H$ = height of a random person.

For Chapter 5, we will focus on discrete random variables.

### 5.1.4   Definition (Probability Function, Probability Distribution)

Let $X$ be a discrete random variable with $range(X) = A$. The *probability function* of $X$ is the function

$$f(x) = P(X = x)$$

for all $x \in A$.
The set of pairs $\{(x, f(x)) : x \in A\}$ is called the *probability distribution* of $X$.

### 5.1.5   Theorem (Properties of Probability Functions)

All probability functions must have the two properties:
1. $0 \le f(x) \le 1$ for all $x \in A$
2. $\sum\limits_{\text{all } x \in A} f(x) = 1$

**Example**

Let $X = \#$ of dice that are $1$.

| $x$ | $0$ | $1$ | $2$ | $3$ |
|---|---|---|---|---|
| $f(x)$ | $5^3/6^3$ | $3 \times 5^2/6^2$ | $3 \times 5/6^3$ | $1^3/6^3$ |

---

### LECTURE 11

---

### 5.1.6   Definition (Cumulative Distribution Function)

The cumulative distribution function of $X$ is the function usually denoted by $F(x)$

$$F(x) = P(X \le x)$$

for all $x \in \mathbb{R}$.

**Example**

| $x$ | $0$ | $1$ | $2$ | $3$ |
|---|---|---|---|---|
| $f(x)$ | $125/216$ | $75/216$ | $15/216$ | $1/216$ |
| $F(x)$ | $125/216$ | $200/216$ | $215/216$ | $1$ |

### 5.1.7   Theorem (Properties of Cumulative Distribution Functions)

All cumulative distribution functions must have the following properties:
1. $F(x)$ is a non-decreasing function of $x$ for all $x \in \mathbb{R}$
2. $0 \le F(x) \le 1$ for all $x \in \mathbb{R}$
3. $\lim\limits_{x \to -\infty} F(x) = 0$
4. $\lim\limits_{x \to \infty} F(x) = 1$

### 5.1.8   Theorem

If $X$ takes on integer values for $x$ such that $x \in A$ and $(x - 1) \in A$,

$$f(x) = F(x) - F(x - 1)$$

*Proof.* $F(x) - F(x-1) = P(X \leq x) - P(X \leq x-1) = P(X = x) = f(x)$ □

## 5.2 Discrete Uniform Distribution

### 5.2.1 Definition (Discrete Uniform Distribution)

Suppose $X$ takes values $a, a+1, \ldots, b$ with all values being equally likely. Then $X$ has a Discrete Uniform distribution on the set $\{a, a+1, \ldots, b\}$ and we write

$$X \sim \mathcal{U}[a, b]$$

**Find the probability function, $f(x)$**

we note there are $(b - a + 1)$ values $X$ can take so the probability at each of these values must be $\frac{1}{b-a+1}$ so that $\sum\limits_{x=a}^{b} f(x) = 1$. Hence

$$f(x) = \begin{cases} \frac{1}{b-a+1}, & x = a, \ldots, b \\ 0, & \text{otherwise} \end{cases}$$

## 5.3 Hypergeometric Distribution

### 5.3.1 Definition (Hypergeometric Distribution)

Suppose we have a collection of $N$ objects which can be classified into two different types, a success (S) and a failure (F). Suppose there are $r$ success and $N - r$ failures. Pick $n$ objects at random without replacement. Let $X$ be the number of successes obtained. Then $X$ has a *Hypergeometric distribution* and we write

$$X \sim \text{Hyp}(N, r, n)$$

**Find the probability function, $f(x)$**

There are $\binom{N}{n}$ points in the sample space $S$ if we don't consider the order of selection. There are $\binom{r}{x}$ ways to choose the successes from the $r$ available AND $\binom{N-r}{n-x}$ ways to choose the remaining $(n - x)$ objects from the $(N - r)$ failures. Hence

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

for $x = 0, \ldots, \min(r, n)$.

**Example**

Suppose we have 10 cards, with 7 that are money cards and 3 non-money cards. Let $X = \#$ of money cards in your hand. Then

$$X \sim Hyp(10, 7, 5)$$

$$f(x) = \frac{\binom{7}{x}\binom{3}{5-x}}{\binom{10}{5}}$$

for $x = 2, 3, 4, 5$ (any less and you run out of non-money cards).

## 5.4   Binomial Distribution

### 5.4.1   Definition (Bernoulli Trials)

(1) Suppose an experiment has two distinct outcomes, call them a success (S) and a failure (F).
(2) Let their probabilities $p$ for S and $(1-p)$ for F.
(3) Repeat the experiment $n$ independent times.
Then, the $n$ individual experiments in the process are called *Bernoulli trials*.

### 5.4.2   Definition (Binomial Distribution)

Suppose an experiment follows the Bernoulli trials. Then $X$ has a *Binomial distribution* and we write

$$X \sim \text{Binomial}(n, p)$$

**Find the probability function, $f(x)$**

There are $\binom{n}{x}$ different arrangements of $x$ S's and $(n-x)$ F's over $n$ trials. Since the trials are independent, the probability of a success is $p$ multiplied $x$ times, and a failure is $(1-p)$ multiplied $(n-x)$ times. Thus we have

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, \ldots, n$.

---

LECTURE 12

---

**Example - Error Correcting Code**

Suppose you send a $4$-bit message over a noisy connection. Each bit is independently flipped with a probability $p = 0.1$. Find $P(\text{message is received correctly})$.

**Solution.**

Let $X = $ # of bits that get flipped.

$$X \sim \text{Binomial}(4, 0.1)$$

$$P(X = 0) = \binom{4}{0} 0.1^0 (1 - 0.1)^{4-0} = 0.6561$$

**Example cont.**

Now suppose we add $3$ parity bits that allow the receiver to detect and correct up to $1$ error in the message. Find $P(\text{message is received correctly})$.

**Solution.**

Let $Y = $ # of bits flipped.

$$Y \sim Bin(7, 0.1)$$

$$P(Y = 0) + P(Y = 1) = \binom{7}{0} 0.1^0 0.9^7 + \binom{7}{1} 0.1^1 0.9^6 \approx 0.8503$$

**Example (ION)**

Suppose there are $100$ people, $10$ people with no bus pass, and $20$ people are selected at random without replacement. Find the probability that there are two people with no bus pass.

**Solution.**

Let $X = \#$ people found with no bus pass.

$$X \sim \text{Hyp}(100, 10, 20)$$

$$P(X = 2) = \frac{\binom{10}{2}\binom{90}{18}}{\binom{100}{20}} \approx 0.3182$$

**Hypergeometric Approximation with Binomial Distribution**

If we did the selection with repetition, the Hypergeometric distribution would become a Binomial distribution.

If $N$ is extremely large compared to $n$, then it doesn't make much difference if the sampling is with our without.

In that case, a $\text{Hyp}(N, r, n)$ case can be approximated by $\text{Binomial}(n, r/N)$

**Example**

Suppose we have $100$ cards, with $70$ money cards and $30$ non-money cards. Select $4$ cards without replacement. Find a suitable approximation for this Hypergeometric distribution assuming $Z = \#$ of money cards selected.

**Solution.**

Let $Z = \#$ of money cards selected.

$$Z \sim \text{Hyp}(100, 70, 5)$$

$$P(Z = 4) = \frac{\binom{70}{4}\binom{30}{1}}{\binom{100}{5}} \approx 0.3654$$

With a Binomial approximation:

$$Z \sim \text{Binomial}(5, 70/100 = 0.7)$$

$$P(Z = 4) = \binom{5}{4}0.7^4(1 - 0.7)^{5-4} = 0.36015$$

which is clearly a very good approximation.

---

<div align="center">LECTURE 13</div>

---

## 5.5 Negative Binomial Distribution

### 5.5.1 Definition (Negative Binomial Distribution)

Suppose an experiment follows the Bernoulli trials, and continue doing this experiment until $k$ success are obtained. Let $X$ be the number of failures obtained before the $k^{\text{th}}$ success. Then $X$ has a *Negative Binomial distribution* and we write

$$X \sim \text{NB}(k, p)$$

**Find the probability function, $f(x)$**

There will be a total of $(x + k)$ trials ($x$ F's $k$ S's) and the last trial will be a success. In the first $(x + k - 1)$ trials we must have $x$ failures and $(k - 1)$ successes, where order does not matter (the last trial we know it

must be a success, that's why we are only looking at before the $k^{\text{th}}$ success!) Hence, we have a combination which is $\binom{x+k-1}{k-1}$. Each order will have a probability $p^k(1-p)^x$. Hence

$$f(x) = \binom{x+k-1}{k-1}p^k(1-p)^x$$

for $x \in [0, \infty)$.

In a picture:



### Example

How many tails until we get the $10^{\text{th}}$ head on a fair coin?

### Solution.

Let $X = $ # of tails before the $10^{\text{th}}$ head.

$$X \sim \text{NB}\left(10, {}^1\!/_2\right)$$

### Example

If courses were independent with probability $p$ of passing and you need $40$ courses, then the number of failed courses would be $\text{NB}(40, p)$.

### Example

Suppose a start-up is looking for $5$ investors. They ask investors repeatedly where each independently has a $20\%$ chance of saying yes. Let $X = $ total # of investors that they ask and note that $X$ does not follow a negative binomial distribution. Find $f(x)$ and $f(10)$.

### Solution.

Let $Y = $ # who say no before $5$ say yes. We know $X = Y + 5$.

$$Y \sim \text{NB}(5, 0.2)$$

$$\begin{aligned}
f(x) &= P(X = x) \\
&= P(Y + 5 = x) \\
&= P(Y = x - 5) \\
&= \binom{(x-5)+5-1}{5-1}(0.2)^5(0.8)^{x-5} \\
&= \binom{x-1}{4}(0.2)^5(0.8)^{x-5}
\end{aligned}$$

for $x \in [5, \infty)$.

$$f(10) = \binom{9}{4}(0.2)^5(0.8)^5 \approx 0.0132$$

Note that we have $\binom{9}{4}$ and not $\binom{10}{5}$ because the $10^{\text{th}}$ investor must have said yes.

---

## LECTURE 14

---

### Example

Suppose you send a bit string over a noisy connection with each bit independently having a probability $0.01$ of being flipped. What is the probability that

(a) it takes 50 bits to get 5 errors?

(b) a 50 bit message has 5 errors?

**Solution.**

(b) Let $Y = \#$ of errors in 50 bits.

$$Y \sim \text{Binomial}(50, 0.01)$$

$$P(Y = 5) = \binom{50}{5}(0.01)^5(0.99)^{45}$$

(a) Let $X = \#$ of correct bits until 5 errors.

$$X \sim \text{NB}(5, 0.01)$$

$$P(X = 45) = \binom{49}{4}(0.01)^5(0.99)^{45}$$

## 5.6   Geometric Distribution

### 5.6.1   Definition (Geometric Distribution)

Consider the exact same process as in the Negative Binomial distribution case, but with $k = 1$. That is, we repeat the Bernoulli trials until the first success. Let $X$ be the number of failures obtained before the first success. Then $X$ has a *Geometric distribution* and we write

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

**Find the probability function, $f(x)$**

Substitute $k = 1$ into the Negative Binomial distribution to obtain

$$f(x) = p(1 - p)^x$$

for $x \in [0, \infty)$.

**Prove the following:**

$$\sum_{x=0}^{\infty} f(x) = 1$$

*Proof.*

$$\sum_{x=0}^{\infty}(1-p)^x p = \underbrace{p + p(1-p) + \cdots}_{\text{(geo. series: } a = p,\, r = 1 - p)}$$

$$= \frac{p}{1 - (1 - p)}$$

$$= 1$$

$\square$

**Find the cumulative distribution function,** $F(x)$

$$
\begin{aligned}
F(x) &= P(X \le x) \\
&= 1 - P(X > x) \\
&= 1 - [f(x+1) + f(x+2) + \cdots] \\
&= 1 - \underbrace{\left[ p(1-p)^{x+1} + p(1-p)^{x+2} + \cdots \right]}_{\text{(geo. series: } a = p(1-p)^{x+1},\, r = 1-p)} \\
&= 1 - \frac{p(1-p)^{x+1}}{1-(1-p)} \\
&= 1 - (1-p)^{x+1}
\end{aligned}
$$

for $x \in [0, \infty)$.

If $x \in \mathbb{R}$, then

$$
F(x) = \begin{cases} 1 - (1-p)^{\lfloor x \rfloor + 1} & x \ge 0 \\ 0 & x < 0 \end{cases}
$$

---

<center>LECTURE 15</center>

---

**Example**

Naomi invites $12$ people to her party. If each independently comes with probability $p$. Let $X = \#$ of guests.

$$
X \sim \text{Binomial}(12, p)
$$

**Example**

$20$ toys in a machine. Each time you grab one with a claw. Let $X = \#$ of tries to get one toy you want.

*None.*

**Example**

Trying to catch a Pokémon, each time has a probability $p$ of succeeding. Let $X = \#$ of failed attempts.

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

**Example**

You have $5$ classes randomly scheduled in a row.

Let $X = \#$ of classes before your favourite.

The range of $X$ is: $0, 1, 2, 3, 4$, and the probability is $1/5$ for each of the range. We have the following:

$$
X \sim \mathcal{U}[1, 4]
$$

## 5.7 Poisson Distribution from Binomial

Suppose we have a $X \sim \text{Binomial}(n, p)$ where $n$ is very large and $p$ is very small. Then, as $n \to \infty$ and $p \to 0$ such that $np$ remains constant, the probability function of X approaches a limit.

Let $np = \mu$, so $p = \frac{\mu}{n}$. Then

$$
\begin{aligned}
\lim_{n \to \infty} f(x) &= \lim_{n \to \infty} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \lim_{n \to \infty} \frac{n(n-1)\cdots(n-x+1)}{x!} \frac{\mu^x}{n^x} \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-x} \\
&= \frac{\mu^x}{x!} \lim_{n \to \infty} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-x+1}{n} \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-x} \\
&= \frac{\mu^x}{x!} \lim_{n \to \infty} \left(1 - \frac{\mu}{n}\right)^n \\
&= \frac{e^{-\mu} \mu^x}{x!}
\end{aligned}
$$

We write: $X \sim \text{Poisson}(\mu)$, range: $0, 1, \ldots$

We can use the Poisson random variable as an approximation to the Binomial when $n$ is large, and $p$ is small. The only thing we need to do is $\mu = np$.

**Example**

Tim Hortons roll up the rim says $1$ in $6$ cups win a prize. Suppose you have $80$ cups. Find the probability that you get $10$ or fewer winners.

Let $X = \#$ of winning cups. $X \sim \text{Binomial}(80, 1/6)$ We want

$$
\begin{aligned}
F(10) &= P(X \leq 10) \\
&= \sum_{x=0}^{10} f(x) \\
&= \binom{80}{0}\left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{80} + \cdots + \binom{80}{10}\left(\frac{1}{6}\right)^{10}\left(\frac{5}{6}\right)^{70} \\
&= 0.2002 \text{ (tedious)}
\end{aligned}
$$

Try a Poisson approximation. $Y \sim \text{Poisson}(\mu = np = \frac{80}{6} \approx 13.33)$. Then,

$$
P(Y \leq 10) = e^{-13.33}\left[1 + \frac{13.33}{1!} + \cdots + \frac{13.33^{10}}{10!}\right] = 0.224
$$

Not a good approximation since $p$ was too large.

## 5.8   Poisson Distribution from Poisson Process

### 5.8.1   Definition (Poisson Process)

*Poisson Process*: Suppose events occur randomly in time or space according to three conditions:
   (1)  Independence: the number of events in one period cannot affect another non-overlapping period
   (2)  Individuality: events occur one at a time (cannot have two at the exact same time)
   (3)  Homogeneity or Uniformity: events occur at a constant rate

LECTURE 16

**Example**

Request coming in from a web server at a rate of $100$ requests per minute. $\lambda = 100, t = \frac{1}{60}$ The # of requests per second would be

$$\text{Poisson}\left(\mu = \frac{100}{60} = \frac{5}{3}\right)$$

## 5.9 Combining Other Models with the Poisson Process

Problems may involve many different random variables!

**Example (Continued)**

We say that a second is quiet if it has no requests.

(a) Find probability that a second is quiet

(b) In a minute ($60$ non-overlapping seconds), find the probability of $10$ quiet seconds

(c) Find the probability of having to wait $30$ non-overlapping seconds to get $2$ quiet seconds

(d) Given (c), find the probability of $1$ quiet second in the first $15$ seconds

(a) Let $X = $ # requests in a second. $X \sim \text{Poisson}(5/3)$.

We want $P(X = 0) = \frac{e^{-\frac{5}{3}}\left(\frac{5}{3}\right)^0}{0!} = 0.189$

(b) Let $Y = $ # quiet seconds out of 60. $Y \sim \text{Binomial}(60, 0.189)$.

We want $P(Y = 10) = \binom{60}{10}(0.189)^{10}(0.811)^{50} = 0.124$

(c) Let $Z = $ # non-quiet seconds before getting $2$ quiet seconds. $Z \sim \text{NB}(2, 0.189)$.

We want $P(Z = 28) = \binom{29}{1}(0.189)^2(0.811)^{28} = 0.003$

(d) $D_x = $ # of quiet seconds out of 15. $D_x \sim \text{Binomial}(15, 0.189)$.

$$P(D_x = 1) = \binom{15}{1}(0.189)^1(0.811)^{14}$$

We get,

$P(1 \text{ quiet second in the first 15 seconds} \mid \text{wait 30 to get 2 quiet}) =$

$$= \frac{P(1 \text{ quiet second in the first 15 seconds AND wait 30 to get 2 quiet})}{P(\text{wait 30 to get 2 quiet})} \tag{5.1}$$

$$= \frac{P(1 \text{ quiet second in the first 15 seconds AND wait an additional 15 to get 1 additional quiet})}{P(C)} \tag{5.2}$$

$$= \frac{P(1 \text{ quiet second in the first 15 seconds})P(\text{wait an additional 15 to get 1 additional quiet})}{P(C)} \tag{5.3}$$

$$= \frac{\binom{15}{1}(0.189)^1(0.811)^{14} \times (0.811)^{14}(0.189)}{\binom{29}{28}(0.189)^2(0.811)^{28}} \tag{5.4}$$

$$= \frac{\binom{15}{1}}{\binom{29}{28}} \tag{5.5}$$

$$= \frac{15}{29} \tag{5.6}$$

In (3) we used the independence of non-overlapping time intervals and constant probability of events.

|  | Discrete Uniform | Hypergeometric | Binomial | Negative Binomial | Geometric | Poisson |
|---|---|---|---|---|---|---|
| function | $\mathcal{U}[a,b]$ | $\mathrm{Hyp}(N,r,n)$ | $\mathrm{Binomial}(n,p)$ | $\mathrm{NB}(k,p)$ | $\mathrm{Geometric}(p)$ | $\mathrm{Poisson}(\mu)$ |
| range | $a, a+1, \ldots, b$ | bad | $0, 1, \ldots, n$ | $0, 1, \ldots$ | $0, 1, \ldots$ | $0, 1, \ldots$ |
| parameters |  |  |  |  |  | $\mu = np,\ \mu = \lambda t$ |
| $f(x)$ | $\frac{1}{b-a+1}$ | $\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ | $\binom{n}{x}p^x(1-p)^{n-x}$ | $\binom{x+k-1}{k-1}p^k(1-p)^x$ | $p(1-p)^x$ | $\frac{e^{-\mu}\mu^x}{x!}$ |
| $F(x)$ | $\frac{x-a+1}{b-a+1}$ |  |  |  | $1-(1-p)^{x+1}$ | $e^\mu\left[1+\frac{\mu^1}{1!}+\cdots\frac{\mu^x}{x!}\right]$ |
| how to tell | "equally likely" know min. and max. | know total # objects know # S's know # trials count # S's without replacement selecting a subset | Bernoulli trials know # trials count # S's | Bernoulli trials "until" "it takes… to get" "before" know how many S's we want | "until we get" "before the first" | Bin. with large amount of trials and small probability rate specified (#events/time) no pre-specified max. events occurring at any time (randomly) Poisson process, known time, count events doesn't make sense to ask how often an event did **not** occur |

# Chapter 7

# Expected Value and Variance

## 7.1 Summarizing Data on Random Variables

Let $X = $ # of kids in a family.

| Value | Frequency |
|-------|-----------|
| 1 | 3 |
| 2 | 10 |
| 3 | 1 |
| 4 | 1 |

### 7.1.1 Definition (Median)

The *median* of a sample is a value such that half the results are below it and half above it, when the results are arranged in numerical order.

### 7.1.2 Definition (Mode)

The *mode* of the sample is the value which occurs most often. There is no guarantee there will be only a single mode.

Mean: average $\rightarrow \frac{1\times3+2\times10+3\times1+4\times1}{15} = 2$

Median: $2$

Mode: $2$

---
### LECTURE 17*
---

## 7.2 Expectation of a Random Variable

Imagine we know the theoretical probability of each # of kids in a family.

| $x$ | 1 | 2 | 3 | 4 | 5 |
|------|------|-----|------|------|------|
| $f(x)$ | 0.43 | 0.4 | 0.12 | 0.04 | 0.01 |

Now we replace the observed proportion in the sample mean with $f(x)$.

$$\sum_{\text{all } x} xf(x) = (1)(0.43) + (2)(0.4) + (3)(0.12) + (4)(0.04) + (5)(0.01) = 1.8$$

which is the theoretical mean.

Why do we have sample mean > theoretical mean?

- urban vs rural population

- income level

- sampled max family size but theoretical includes growing families

- selection bias (if you randomly select people rather than families, people with lots of siblings will be over-represented)

### 7.2.1   Definition (Expected Value)

Let $X$ be a discrete random variable with range$(X) = A$ and probability function $f(x)$. The *expected value* (also called the mean or the expectation) of $X$ is given by

$$\mu = E[X] = \sum_{x \in A} x f(x)$$

**REMARK 7.2.1.** $\mu$ will be within the range but not necessarily equal to a possible value of $x$.

We might be interested in the expected value of some function of $X$, $g(X)$.

**Example**

Tax credit of $1000 plus $250 per kid. Find the average cost.

| $x$ | 1 | 2 | 3 | 4 | 5 |
|------|------|------|------|------|------|
| $g(x)$ | 1250 | 1500 | 1750 | 2000 | 2250 |

Average cost = weighted average of $g(x)$ values = $(1250)(0.43) + \cdots + (2250)(0.01) = 1450$

### 7.2.2   Theorem

Let $X$ be a discrete random variable with range$(X) = A$ and probability function $f(x)$. The expected value of a some function $g(X)$ of $X$ is given by

$$E[g(X)] = \sum_{x \in A} g(x) f(x)$$

Note that $g(x) = 1000 + 250x$ from last example.

$$E[g(X)] = 1000 + 250E[X] = 1450$$

What if tax credit $= \frac{2000}{x}$

$$E[g(X)] = (2000)(0.43) + (1000)(0.40) + \cdots + (400)(0.01) = 1364$$

But $\frac{2000}{E[X]} = \frac{2000}{1.8} = 1111.11$. Therefore

$$E[g(X)] \neq g(E[X])$$

unless $g$ is a linear function. That is, if $g(X) = aX + b$, then $E[g(X)] = aE[X] + b$

### 7.2.3   Theorem

Let $X$ be a discrete random variable with range$(X) = A$ and probability function $f(x)$. For constants $a$ and $b$ and some function $g(X)$,

$$E[ag(X) + b] = aE[g(X)] + b$$

*Proof.*

$$E[ag(X) + b] = \sum_{x \in A} (ag(x) + b) \, f(x)$$

$$= \sum_{x \in A} (ag(x)f(x) + bf(x))$$

$$= a \sum_{x \in A} g(x)f(x) + b \sum_{x \in A} f(x)$$

$$= aE[g(X)] + b \qquad \left[ \text{since } \sum_{x \in A} f(x) = 1 \right]$$

$\square$

### 7.2.4  Theorem

Let $X$ be a discrete random variable with range$(X) = A$ and probability function $f(x)$. For constants $a$ and $b$ and some functions $g_1(X)$, $g_2(X)$,

$$E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_2(X)]$$

*Proof.*

$$E[ag_1(X) + bg_2(X)] = \sum_{x \in A} (ag_1(x) + bg_2(x)) \, f(x)$$

$$= \sum_{x \in A} (ag_1(x)f(x) + bg_2(x)f(x))$$

$$= a \sum_{x \in A} g_1(x)f(x) + b \sum_{x \in A} g_2(x)f(x)$$

$$= aE[g_1(X)] + bE[g_2(X)]$$

$\square$

## 7.3  Some Applications of Expectation

**Example**

A web server has a cache. Takes 10ms to check, 20% of the requests are found (cache hit) and immediately shown. If it's not found (cache miss), it takes $\underbrace{50}_{\text{to server}} + \underbrace{70}_{\text{lookup}} + \underbrace{50}_{\text{to client}}$ additional milliseconds to get info and display. Is it worth it? Let $X = \#$ of milliseconds to display the information.

| $x$ | 10 | 10+50+70+50=180 |
|---|---|---|
| $f(x)$ | 0.2 | 0.8 |

$$E[X] = (10)(0.2) + (180)(0.8) = 146\text{ms}$$

Time no cache $= 50 + 70 + 50 = 170$ms.

Since $146$ms $< 170$ms, it's worth it!

**Example**

Roulette: each of $38$ numbers is equally likely

(1) If you bet $1$ dollar on number $7 \rightarrow$ pays $35:1$

OR

(2) If you bet $50$ cents on red $\rightarrow$ pays $1:1$ and $50$ cents on first $12$ numbers $\rightarrow$ pays $2:1$

(1)

| $x$ | $0$ | $36$ |
|---|---|---|
| $f(x)$ | $37/38$ | $1/38$ |

(2)

| $y$ | $0$ | $1$ | $1.50$ | $2.50$ |
|---|---|---|---|---|
| $f(y)$ | $14/38$ | $12/38$ | $6/38$ | $6/38$ |
| | neither | red | black | both red |

$$E[X] = 0\left(\frac{37}{38}\right) + 36\left(\frac{1}{38}\right) = 0.94737$$

$$E[Y] = 0\left(\frac{14}{38}\right) + 1\left(\frac{12}{38}\right) + 1.5\left(\frac{6}{38}\right) + 2.5\left(\frac{6}{38}\right) = 0.94737$$

---

LECTURE 18

---

## 7.4  Means and Variances of Distributions

The mean $E[X]$ tells us where the distribution is on average. We also need a way to describe how spread out a distribution is. Variance could be $E[X - \mu] = 0$.

What about $E[|X - \mu|]$

- need cases to evaluate
- non-differentiable at point $X - \mu$
- linear penalty for being away from the mean

Instead we use $E[(X - \mu)^2]$

### 7.4.1  Definition (Variance)

The *variance* of a random variable $X$, denoted by $Var(X)$ or by $\sigma^2$, is

$$\sigma^2 = Var(X) = E\left[(X - \mu)^2\right]$$

**Example**

$X = \#$ on fair 6-sided die

$E[X] = 3.5$

$E[(x - 3.5)^2]$

$E[X]^2 - 3.5^2$

| $x$ | $1$ | $2$ | $3$ | $4$ | $5$ | $6$ |
|---|---|---|---|---|---|---|
| $x^2$ | $1$ | $4$ | $9$ | $16$ | $25$ | $36$ |

Alternate form (calculation form)

$$
\begin{aligned}
Var(X) &= E[(X - E[X])^2] \\
&= E[X^2 - 2XE[X] + E[X]^2] \\
&= E[X^2] - 2E[X]E[X] + E[X]^2 \\
&= E[X^2] - 2(E[X])^2 + E[X]^2 \\
&= E[X^2] - E[X]^2 \\
&= \sum_{\text{all } x} x^2 f(x) - \left( \sum_{\text{all } x} x f(x) \right)^2
\end{aligned}
$$

## 7.4.2 Example (Roulette)

$X = 0$ or 36 (dollars)

| $x$ | 0 | 36 |
|---|---|---|
| $f(x)$ | $37/38$ | $1/38$ |

$$E[X] = 0.94737$$

$$Var(X) = E[X^2] - 0.94737^2 = 36^2(\frac{1}{36}) - 0.94737^2 = 33.207 \text{ dollars}^2$$

To interpret the variance better, we often take the square root to get the same units of the original variable.

### 7.4.3 Definition (Standard Deviation)

The *standard deviation* of a random variable $X$ is

$$\sigma = SD(X) = \sqrt{Var(X)}$$

$$SD(X) = \sqrt{33.207} = 5.76$$

What if we bet \$1 on red. Y=winnings

| $y$ | 0 | 2 |
|---|---|---|
| $f(y)$ | $20/38$ | $18/38$ |

$$E[Y] = 0.94737$$

$$Var(Y) = E[Y^2] - 0.94737^2 = 0.97723$$

$$SD(Y) = 0.9986$$

## 7.4.4 Linear Transformations

If $Y = aX + b$, and we know $E[X]$ and $Var(X)$, what can we say about $E[Y]$ and $Var(Y)$.

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

$$
\begin{aligned}
Var(Y) &= E[(Y - E[Y])^2] \\
&= E[(aX + b) - (aE[X] + b)^2] \\
&= E[a^2 X^2 - 2XE[X] + E[X]^2] \\
&= a^2 E[(X - E[X])^2]
\end{aligned}
$$

$$Var(Y) = a^2 Var(X)$$
$$SD(Y) = |a| SD(X)$$

LECTURE 19*

**Example**

Suppose $X$ has probability function:

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $y$ | 1 | 3 | 5 | 7 | 9 |
| $f(x)$ | 0.1 | 0.1 | 0.1 | 0.5 | 0.2 |

Let $Y = 2X + 1$.

$E[X] = 2.6$

$E[X^2] = 6.2$

$E[Y] = 6.2$

$E[Y^2] = 94.2$

$Var(X) = 8.2 - 2.6^2 = 1.44$

$SD(X) = 1.2$

$Var(Y) = 44.2 - 6.2 = 5.76$

$SD(Y) = 2.4$

Now we can verify,

$$\begin{aligned} E[Y] &= E[2X + 1] \\ &= 2E[X] + 1 \\ &= 2(2.6) + 1 \\ &= 6.2 \end{aligned}$$

$$Var(Y) = 2^2 Var(X) = 4(1.44) = 5.76$$
$$SD(Y) = |2| SD(X) = 2(1.2) = 2.4$$

Let $X \sim \text{Binomial}(n, p)$. Find $E[X]$.

$$E[X] = \sum_{\text{all } x} x f(x) \tag{7.1}$$

$$= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \tag{7.2}$$

$$= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \tag{7.3}$$

$$= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \tag{7.4}$$

$$= \sum_{x=1}^{n} x \frac{n!}{x(x-1)!(n-x)!} p^x (1-p)^{n-x} \tag{7.5}$$

$$= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \tag{7.6}$$

$$= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)![(n-1)-(x-1)]!} pp^{x-1} (1-p)^{(n-1)-(x-1)} \tag{7.7}$$

$$= np(1-p)^{n-1} \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{-(x-1)} \tag{7.8}$$

$$= np(1-p)^{n-1} \sum_{x=1}^{n} \binom{n-1}{x-1} \left(\frac{p}{1-p}\right)^{x-1} \tag{7.9}$$

From (2) to (3) we used to fact that when $x = 0$ the value of the expression is $0$. Provided that $x \neq 0$, we can expand $x!$ as $x(x-1)!$ as seen from (4) to (5). Let $y = x - 1$, we get

$$E[X] = np(1-p)^{n-1} \sum_{y=0}^{n} \binom{n-1}{y} \left(\frac{p}{1-p}\right)^y \tag{7.10}$$

$$= np(1-p)^{n-1} \left(1 + \frac{p}{1-p}\right)^{n-1} \tag{7.11}$$

$$= np(1-p)^{n-1} \frac{(1-p+p)^{n-1}}{(1-p)^{n-1}} \tag{7.12}$$

$$= np \tag{7.13}$$

From (10) to (11) we used the Binomial Theorem.

Let $X \sim \text{Poisson}(\mu)$. Find $E[X]$.

$$E[X] = \sum_{\text{all } x} x f(x) \tag{7.1}$$

$$= \sum_{x=0}^{\infty} x \frac{e^{-\mu} \mu^x}{x!} \tag{7.2}$$

$$= \sum_{x=1}^{\infty} x \frac{e^{-\mu} \mu^x}{x(x-1)!} \tag{7.3}$$

$$= \sum_{x=1}^{\infty} \mu \frac{e^{-\mu} \mu^{x-1}}{(x-1)!} \tag{7.4}$$

$$\tag{7.5}$$

Let $y = x - 1$, we get

$$E[X] = \mu e^{-\mu} \sum_{y=0}^{\infty} \frac{\mu^y}{y!} \tag{7.6}$$

$$= \mu e^{-\mu} e^{\mu} \tag{7.7}$$

$$= \mu \tag{7.8}$$

From (6) to (7) we used the fact that $e^x = \sum_{y=0}^{\infty} \frac{x^y}{y!}$.

Similarly,

$$X \sim \mathcal{U}[a, b], \ E[X] = \frac{a+b}{2}$$

$$X \sim \text{Hyp}(N, r, n), \ E[X] = \frac{nr}{N}$$

$$X \sim \text{NB}(k, p), \ E[X] = \frac{k(1-p)}{p}$$

$$X \sim \text{Geometric}(p), \ E[X] = \frac{1-p}{p}$$

Let $X \sim \text{Poisson}(\mu)$. Find $Var(X)$.

Since there's $x!$ in the denominator of $f(x)$, let's find $E[X(X-1)]$.

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)\frac{\mu^x e^{-\mu}}{x!}$$

$$= \sum_{x=2}^{\infty} x(x-1)\frac{\mu^x e^{-\mu}}{x(x-1)(x-2)!}$$

$$= \mu^2 e^{-\mu} \sum_{x=2}^{\infty} \frac{\mu^{x-2}}{(x-2)!}$$

Let $y = x - 2$, we get

$$E[X(X-1)] = \mu^2 e^{-\mu} \sum_{y=0}^{\infty} \frac{\mu^y}{y!}$$

$$= \mu^2$$

$$Var(X) = E[X(X-1)] + E[X] - E[X]^2$$
$$= \mu^2 + \mu - \mu^2$$
$$= \mu$$

---

## LECTURE 20*

---

**Example**

Suppose the amount of data you use on your phone (in units of 100MB) has a Poisson distribution with mean 7 per month. You pay 15 per month plus 3 per 100MB. Find the standard deviation of random month's phone bill.

Let $X$ = # of units of data used. $X \sim \text{Poisson}(7)$. Let $Y = 15 + 3X \rightarrow E[Y] = 15 + 3(7) = 36$.

$SD(Y) = 3SD(X) = 3\sqrt{7} = 7.94$.

# Chapter 8

# Continuous Random Variables

---

---

## 8.1 General Terminology and Notation

A *continuous random variable* $X$ maps points in a continuous sample space to real numbers such that the range is uncountably infinite.

EXAMPLES OF CONTINUOUS RANDOM VARIABLES

Let $X$ be the number the point spots at.

(1) temperature of a day

(2) length of time until a bus arrives

(3) height of a random person

(4) average height of 10 people

$F(x) = P(X \leq x)$

$F(a) = P(X \leq a)$

**Example**

For $x < 0$, no chance of the point stopping at a number $< 0$.

For $x > 4$, $F(x) = 1$ since the point is certain to stop at a number below $4$.

$P(0 < x \leq 1) = \frac{1}{4} = F(1)$

$$F(x) = \begin{cases} 0, \ x < 0 \\ \frac{x}{4}, \ 0 \leq x \leq 4 \\ 1, \ x > 4 \end{cases}$$

PROPERTIES OF $F(x)$

(1) For all $x$, $P(X = x) = 0$. So,

$$\begin{aligned} P(a < x \leq b) &= P(a \leq x \leq b) \\ &= P(a < x < b) \\ &= P(a \leq x < b) \end{aligned}$$

40

**REMARK 8.1.1.** End points don't matter.

(2)

$$\lim_{\varepsilon \to 0} F(x) - F(x - \varepsilon) = \lim_{\varepsilon \to 0} P(x - \varepsilon < X \leq x)$$
$$= P(X = x)$$
$$= 0$$

Thus $\lim_{\varepsilon \to 0} F(x - \varepsilon) = F(x)$, so $F(x)$ is continuous.

(3) $F(x)$ is non-decreasing.

(4) $\lim_{x \to +\infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$

(5) $0 \leq F(x) \leq 1$

### 8.1.1   Definition (Probability Density Function)

The *probability density function* (p.d.f) $f(x)$ for a continuous random variable $X$ is the derivative

$$f(x) = \frac{d}{dx} F(x)$$

where $F(x)$ is the cumulative distribution function for $X$.

**REMARK 8.1.2.** $f(x)$ is not a probability. It can be $> 1$ relative likelihood that $X$ takes a value near $X$.

PROPERTIES OF $f(x)$

(1)

$$P(a \leq X \leq b) = F(b) - F(a)$$
$$= \int_a^b f(x)dx$$

(2)

$$\int_{-\infty}^{+\infty} f(x)dx = F(+\infty) - F(-\infty)$$
$$= 1 - 0$$
$$= 1$$

(3) $f(x) \geq 0$ (since $F(x)$ is non-decreasing, it's derivative is non-negative)

(4)

$$F(x) = \int_{-\infty}^{x} f(u)du$$

**Example**

Suppose a continuous random variable $X$ is on the range $[0, 1]$ has the cumulative distribution function $F(x) = x^2$.

WHAT IS THE PROBABILITY DENSITY FUNCTION?

$f(x) = \frac{d}{dx}F(x) = 2x.$

WHAT IS $P(X = 0.25)$?

$P(X = 0.25) = 0$

WHAT IS $P(X \leq 0.25)$?

(1) $P(X \leq 0.25) = F(0.25) = (0.25)^2 = 0.0625$

(2)

$$P(X \leq 0.25) = \int_0^{0.25} f(x)dx = \int_0^{0.25} 2xdx = 0.625$$

Expectation:

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_{x \in \text{range}} xf(x)dx$$

Variance:

$$Var(X) = E[X^2] - E[X]^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx - \int_{-\infty}^{+\infty} xf(x)dx$$

### 8.1.2 Definition (Percentiles)

The $p^{\text{th}}$ *percentile* of a distribution $x_p$ such that $F(x_p) = p$.

---

LECTURE 22

---

**Example**

$F(x) = x^2$ for $0 < x < 1$.

FIND THE MEAN, MEDIAN, AND MODE

Mean:

$$E[X] = \int_0^1 x2xdx = \int_0^1 2x^2 dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}$$

Median: $x_{0.5}$ satisfies $F(x_{0.5}) = 0.5 \implies (x_{0.5})^2 = 0.5 \implies x_{0.5} = \sqrt{0.5} = 0.707$

Mode: 1 ($x$ value that maximizes $f(x)$)

## 8.2 Continuous Uniform Distribution

A *continuous* random variable takes real values between $a$ and $b$ with $a < b$ such that any interval of fixed size is equally likely.

NOTATION

$X \sim U(a, b)$

**REMARK 8.2.1.** Can include or exclude endpoints, doesn't matter.

FIND $f(x)$

$f(x) = c$, (since it can't depend on $x$). We need

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

$$\int_{a}^{b} c\, dx = 1$$

$[cx]_a^b = 1 \implies c(b-a) = 1 \implies c = \frac{1}{b-a}$

So,

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

FIND $F(x)$

$$F(x) = \int_{-\infty}^{x} f(u)du = \int_{a}^{x} \frac{1}{b-a} du = \left[\frac{u}{b-a}\right]_a^x = \frac{x-a}{b-a}$$

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

FIND THE MEAN, MEDIAN AND MODE

Mean:

$$E[X] = \int_{a}^{b} x \frac{1}{b-a} dx = \left[\left(\frac{x^2}{2}\right)\left(\frac{1}{b-a}\right)\right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

Median: is also $\frac{a+b}{2}$

Mode: no unique mode

Similarly,

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

SPECIAL CASE

$U \sim U(0,1)$ (i.e. $a = 0$, $b = 0$)

$$f(u) = \begin{cases} 1, & 0 < u < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$F(u) = \begin{cases} u, & 0 < u < 1 \\ 0, & u < 0 \\ 1, & u > 1 \end{cases}$$

$U(0,1)$ random variables are easy to generate.

## 8.2.1   Change of Variables

Suppose you know the distribution of $X$ and you want the distribution of $Y = g(X)$.

1. Write the cumulative distribution function of $Y$ in terms of the cumulative distribution function of $X$

2. Sub in what we know about $X$, then differentiate to get the pdf

3. Determine the range of $Y$

## 8.2.2   Example (Change of Variable)

Let $X \sim U(0,4)$, $F_X(x) = \frac{x}{4}$, $f_X(x) = \frac{1}{4}$ $x \in (0,4)$

Let $Y = \frac{1}{X}$

1.

$$
\begin{aligned}
F_Y(y) &= P\left(Y \leq y\right) \\
&= P\left(\frac{1}{X} \leq y\right) \\
&= P\left(X > \frac{1}{y}\right) \\
&= 1 - F_X\left(\frac{1}{y}\right)
\end{aligned}
$$

2.

$$
\begin{aligned}
F_Y(y) &= 1 - F_X\left(\frac{1}{y}\right) \\
&= 1 - \frac{\frac{1}{y}}{4} \\
&= 1 - \frac{1}{4y}
\end{aligned}
$$

$f_Y(y) = \frac{d}{dx}F_Y(y) = \frac{1}{4y^2}$

OR differentiate $F_Y(y)$ before substituting in the information about $X$. You need the chain rule!

$$
\frac{d}{dy}\left[1 - F_X\left(\frac{1}{y}\right)\right] = -f_X\left(\frac{1}{y}\right)\left(-\frac{1}{y^2}\right) = \frac{1}{4y^2}
$$

3. $y \in (\frac{1}{4}, \infty)$

---

## LECTURE 23

**Example**

Let $Y \sim U(0,1) \implies f_Y(y) = \frac{1}{1-0} = 1$, $F_Y(y) = \frac{y-0}{1-0} = y$

$X = 2\sqrt[3]{Y}$.

FIND $f_X(x)$

1.

$$F_X(x) = P(X \leq x)$$
$$= P\left(2\sqrt[3]{Y} \leq x\right)$$
$$= P\left(Y \leq \frac{x^3}{8}\right)$$
$$= F_Y\left(\frac{x^3}{8}\right)$$

2.

$$F_X(x) = F_Y\left(\frac{x^3}{8}\right)$$
$$= \frac{x^3}{8}$$

$f_X(x) = \frac{d}{dx}F_X(x) = \frac{3}{8}x^2$

3. $x \in (0, 2)$

## 8.3  Exponential Distribution

Suppose we have a Poisson Process with rate $\lambda$. Let $X =$ time until the next event occurs. $X$ has an exponential distribution.

FIND RANGE, $F(x)$, AND $f(x)$

$x \in (0, \infty)$

$$F(x) = P(X \leq x)$$
$$= P(\text{time to next event} \leq x)$$
$$= P(\text{number of events in } (0, x)) \geq 1$$
$$= P(Y \geq 1) \qquad Y \sim \text{Poisson}(\lambda x)$$
$$= 1 - P(Y = 0)$$
$$= 1 - \frac{(e^{-\lambda x})(\lambda x)^0}{0!}$$
$$= \begin{cases} 1 - e^{-\lambda x}, \ x > 0 \\ 0, \ x \leq 0 \end{cases}$$

Alternate forms: $\theta = \frac{1}{\lambda}$, so

$$F(x) = 1 - e^{-\frac{x}{\theta}}$$
$$f(x) = \frac{1}{\theta}e^{-\frac{x}{\theta}}$$

We say $X \sim Exp(\theta)$.

---

LECTURE 24

---

FIND THE MEAN AND VARIANCE

$$E[X] = \int_0^\infty x \frac{1}{\theta} e^{-\frac{x}{\theta}} \, dx \text{ IBP}$$

Trick: **Gamma Function**

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

where $\alpha > 0$

Properties of the Gamma Function

(1) if $\alpha > 1$, then $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$

(2) if $\alpha$ is an integer $\geq 1$,

$$\Gamma(1) = 1$$
$$\Gamma(2) = 1\Gamma(1) = 1$$
$$\Gamma(3) = 2\Gamma(2) = 2$$
$$\Gamma(4) = 3\Gamma(3) = 6$$

In general,

$$\Gamma(\alpha) = (\alpha - 1)!$$

So, back to our example:

$$E[X] = \int_0^\infty x \frac{1}{\theta} e^{-\frac{x}{\theta}} \, dx \qquad y = \frac{x}{\theta} \implies x = (\theta y) \wedge \theta dy = dx$$

$$= \int_0^\infty (\theta y) \frac{1}{\theta} e^{-y} \theta dy$$

$$= \theta \int_0^\infty y^{2-1} e^{-y} dy \qquad \Gamma(2) = (2-1)! = 1$$

$$= \theta$$

$$E[X] = \theta = \frac{1}{\lambda}$$

Why? If $\lambda$ is higher, events happen more often, which means shorter wait time.

To find $Var(X)$,

$$E[X]^2 = \int_0^\infty x^2 \frac{1}{\theta} e^{-\frac{x}{\theta}} \, dx \qquad y = \frac{x}{\theta} \implies x^2 = (\theta y)^2 \wedge \theta dy = dx$$

$$= \int_0^\infty (\theta y)^2 \frac{1}{\theta} e^{-y} \theta dy$$

$$= \theta^2 \int_0^\infty y^{3-1} e^{-y} dy \qquad \Gamma(3) = (3-1)! = 2$$

$$= 2\theta^2$$

So $Var(X) = 2\theta^2 - \theta^2 = \theta^2$, $SD(X) = \theta = E[X]$

### 8.3.1  Memoryless Property

**Example**

Suppose buses follow a Poisson process with average $5$ per hour.

(a) Find the probability that you wait $> 15$ minutes.

Let $X = $ time until next bus. $X \sim Exp(12)$

$$P(X > 15) = 1 - F(X \leq 15) = 1 - \left(1 - e^{-\frac{15}{12}}\right) = e^{-\frac{15}{12}} \approx 0.2865$$

(b) If you have been waiting 6 minutes already, what is the probability that you wait another $> 15$ more minutes.

$$
\begin{aligned}
P(X > 21 \mid X > 6) &= \frac{P(X > 21 \text{ AND } X > 6)}{P(X > 6)} \\
&= \frac{P(X > 21)}{P(X > 6)} \\
&= \frac{1 - F(21)}{1 - F(6)} \\
&= \frac{1 - (1 - e^{-\frac{21}{12}})}{1 - (1 - e^{-\frac{6}{12}})} \\
&= e^{-\frac{15}{12}} \approx 0.2865
\end{aligned}
$$

The memoryless property says the past is irrelevant in the future distribution. In general, if $s, t > 0$:

$$P(X > t + s \mid X > s) = P(X > t)$$

## 8.5   Normal Distribution

Many natural phenomena tend to follow a shape like this:
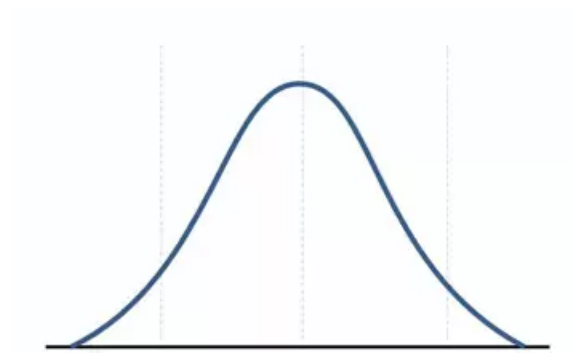


- amount of precipitation
- heights/weights of large populations
- measurement errors
- grades in courses

A normal random variable $X$ with parameters $\mu$ and $\sigma^2$ has pdf

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

for $x \in \mathbb{R}$

- symmetric around $\mu$
- both tails go to zero quickly
- $\frac{1}{\sqrt{2\pi}\sigma}$ makes it integrate to 1.
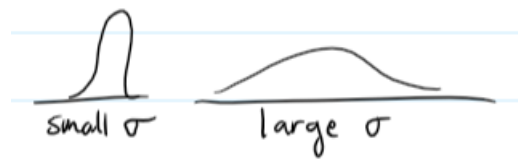


We can show that $E[X] = \mu$ and $Var(X) = \sigma^2$

### 8.5.1 Empirical rule
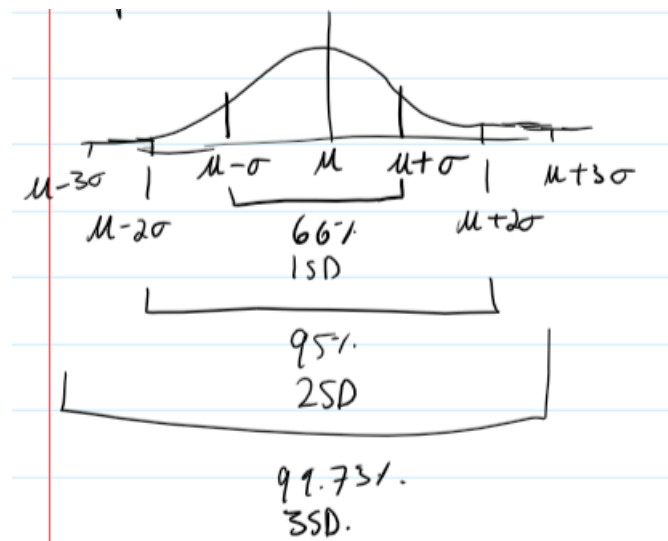


FIND $f(x)$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(u-\mu)^2/(2\sigma^2)} du$$

- not analytically integrable
- look it up or numerically evaluate

Standard Normal random variable (special case with $\mu = 0$, $\sigma^2 = 1$)

$Z \sim N(0,1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

$F(z)$ still has no closed form.

## LECTURE 25

FIND $P(Z \leq 2.31)$

$$P(Z \leq \underbrace{2}_{\text{row}} \cdot \underbrace{31}_{\text{col}}) = 0.98956$$

FIND $P(Z \leq -0.63)$

$$P(Z \leq -0.63) = P(Z > 0.63) = 1 - P(Z \leq 0.63) = 1 - 0.73565 = 0.26435$$

**MLIW 9: Setting the Threshold for a Classifier**

Imagine we send a voltage of $+2$ (for $1$) or $-2$ (for $0$) over a connection to convey a string of bits. The connection is noisy and adds a $N(0, 1)$ distributed amount of voltage to whatever signal is sent.

The person receiving the message on the other side must interpret the incoming signal as either a $1$ or $0$, based on a threshold $c$. If the voltage is above $c$, it will interpret it as a $1$, otherwise a $0$.

**Find $P(\textbf{error})$ if $c = 0.5$**

**Solution.**

$P(\text{error})$ if a $1$ was sent: $R = 2 + Z$. $Z \sim N(0, 1)$

$$\begin{aligned}
P(\text{error}) &= P(R < 0.5) \\
&= P(2 + Z < 0.5) \\
&= P(Z < -1.5) \\
&= P(Z > 1.5) \\
&= 1 - P(Z \leq 1.5) \\
&= 1 - 0.93319 \\
&= 0.06681
\end{aligned}$$

$P(\text{error})$ if a $0$ was sent: $R = -2 + Z$. $Z \sim N(0, 1)$

$$\begin{aligned}
P(\text{error}) &= P(R > 0.5) \\
&= P(-2 + Z > 0.5) \\
&= P(Z > 2.5) \\
&= 1 - P(Z \leq 2.5) \\
&= 1 - 0.99379 \\
&= 0.006621
\end{aligned}$$

Why? We had $c = 0.5$ closer to $2$ than $-2$, thus the probability of error is higher for $1$'s sent than for $0$'s.

If we wanted the probabilities of error to be equal no matter what input, we could set $c = 0$.

**Find $P(\textbf{error})$ if $c = 0$**

**Solution.**

$P(\text{error})$ if a 1 was sent: $R = 2 + Z$. $Z \sim N(0,1)$

$$
\begin{aligned}
P(\text{error}) &= P(R < 0) \\
&= P(2 + Z < 0) \\
&= P(Z < -2) \\
&= P(Z > 2) \\
&= 1 - P(Z \le 2) \\
&= 1 - 0.97725 \\
&= 0.02275
\end{aligned}
$$

$P(\text{error})$ if a 0 was sent: $R = -2 + Z$. $Z \sim N(0,1)$

$$
\begin{aligned}
P(\text{error}) &= P(R > 0) \\
&= P(-2 + Z > 0) \\
&= P(Z > 2) \\
&= 1 - P(Z \le 2) \\
&= 1 - 0.97725 \\
&= 0.02275
\end{aligned}
$$

FIND PERCENTILES OF $N(0,1)$

Suppose we want $c$ such that $P(Z < c) = 0.85$

- look in body of table for $\approx 0.85$ and read off row and column: $c$ is between $1.03$ and $1.04$
- use reverse table, look up row and column: $1.0364$

TRANSFORMING A NORMAL RANDOM VARIABLE

Suppose $X \sim (\mu, \sigma^2)$, $\mu, \sigma^2 < \infty$.

Claim: if

$$
Z = \frac{X - \mu}{\sigma}
$$

then $Z \sim N(0,1)$

Proof.

1.

$$
\begin{aligned}
F_Z(z) &= P(Z \le z) \\
&= P\left(\frac{X - \mu}{\sigma} \le z\right) \\
&= P\left(X \le z\sigma + \mu\right) \\
&= F_X(\sigma z + \mu)
\end{aligned}
$$

2. Differentiate

$$
\begin{aligned}
f_Z(z) &= \frac{d}{dz}F_Z(z) \\
&= \frac{d}{dz}F_X(\sigma z + \mu) \\
&= f_X(\sigma z + \mu)\sigma \qquad \text{Chain Rule} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}e^{-((\sigma z + \mu)-\mu)^2/(2\sigma^2)}\right)\sigma \\
&= \frac{1}{\sqrt{2\pi}}e^{-z^2/2}
\end{aligned}
$$

3. range of $Z$ is $\mathbb{R}$, so $Z \sim N(0,1)$

**Example**

MCAT scores are normal with mean $25.3$ and standard deviation $6.5$.

A SCORE OF $41$ IS HOW GOOD?

Find $P(X > 41)$ where $X \sim N(25.3, 6.5^2)$

$$
P\left(\frac{X - 25.3}{6.5} > \frac{41 - 25.3}{6.5}\right) = P(Z > 2.42) = 1 - 0.99202 = 0.00798
$$

**Example**

You want $98\%$ of the population to use a ride by height. $X = $ height $\sim N(69, 2.4^2)$. That is, find $h$ such that $P(X < h) = 0.98$, so

$$
P\left(\frac{X - 69}{2.4} < \frac{h - 69}{2.4}\right) = 0.98 \implies P\left(Z < \frac{h - 69}{2.4}\right) = 0.98
$$

Set $F(\frac{h-69}{2.4}) = 0.98$, and solve for $h$. You can also take $F^{-1}$ on each side.

$$
2.0537 = \frac{h - 69}{2.4} \implies h = (2.0537)(2.4) + 69 = 73.93 \text{ inches}
$$

In general,

$$
x_p = \sigma z_p + \mu
$$

---

## LECTURE 26

**Example**

If $Z \sim N(0,1)$, find $d$ such that $P(|Z| < d) = 0.9$.

$$
\begin{aligned}
P(|Z| < d) &= P(-d < Z < d) \\
&= P(Z < d) - P(Z > -d) \\
&= P(Z \leq d) - [1 - P(Z \leq d)] \\
&= 2P(Z \leq d) - 1
\end{aligned}
$$

$$2P(Z \leq d) - 1 = 0.90$$
$$\implies P(Z \leq d) = \frac{0.90 + 1}{2}$$
$$\implies F(d) = 0.95$$
$$\implies F^{-1}(F(d)) = F^{-1}(0.95)$$
$$\implies d = 1.6449$$

# Chapter 9

# Multivariate Distributions

## 9.1 Basic Terminology and Techniques

We have models for a single RV (both discrete or cont.) but we often care about two or more RV's at the same time (and their relationship) Examples:

- two stock returns
- heights and weights
- number of cards of a rank vs number of a suit
- treatment vs recovery time
- all machine learning classification and regression

In this course, we focus on all discrete random variables

### 9.1.1 Definition (Joint Probability Function)

Let $X_1, \ldots, X_n$ be $n$ discrete random variables. We define the *joint probability function* $f(x_1, \ldots, x_n)$ of $(X_1, \ldots, X_n)$ as

$$f(x_1, \ldots, x_n) = P(X_1 = x_1 \text{ and } \cdots \text{ and } X_n = x_n)$$
$$= P(X_1 = x_1, \ldots, X_n = x_n)$$

### 9.1.2 Theorem

- $\displaystyle\sum_{\text{all } (x_1, \ldots, x_n)} f(x_1, \ldots, x_n) = 1$
- $f(x_1, \ldots, x_n) \geq 0$ for all $(x_1, \ldots, x_n)$

**Example**

Suppose we flip a coin 3 times. Let $X = \#$ heads. Let

$$Y = \begin{cases} 1, & \text{if first flip is a H} \\ 0, & \text{otherwise} \end{cases}$$

Find $f(x, y)$.

| $y \backslash x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | $^1/_8$ | $^2/_8$ | $^1/_8$ | $^0/_8$ |
| 1 | $^0/_8$ | $^1/_8$ | $^2/_8$ | $^1/_8$ |

$f(x, y)$ can be represented in a table or as a function of $x$ and $y$ (not usually a histogram).

Now suppose we are only interested in one of the random variables, e.g. suppose we are only want to find out about $X$.

$$P(X = x) = f(0,0) + f(0,1) = \frac{1}{8} + 0 = \frac{1}{8}$$

### 9.1.3 Definition (Marginal Probability Function)

Let $X$ and $Y$ be two discrete random variables. We define the *marginal probability function* of $X$ as

$$f_X(x) = \sum_{\text{all } y} f(x, y)$$

and the *marginal probability function* of $Y$ as

$$f_Y(y) = \sum_{\text{all } x} f(x, y)$$

### 9.1.4 Definition (Independent Random Variables)

$X_1, \ldots, X_n$ are *independent random variables* if and only if

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

for all $(x_1, \ldots, x_n)$.

From example: Are $X$ and $Y$ independent? No. $f(0,0) = \frac{1}{8} \neq f_X(0)f_Y(0) = \frac{1}{8} \cdot \frac{1}{2}$ shortcut: any $0$ in your table $\rightarrow$ dependent.

---

LECTURE 27

---

## 9.1.5 Thought Question

For a full-time UW Math Faculty student, let $X$ = number of courses taking and $Y = 1$ if in co-op, or $0$ if in regular. The joint pf is given by (this is real data)

| $y \backslash x$ | 3 | 4 | 5 | 6 | $f_Y(y)$ |
|---|---|---|---|---|---|
| 0 | 0.09 | 0.17 | 0.22 | 0.01 | |
| 1 | 0.05 | 0.10 | 0.32 | 0.04 | 0.51 |
| $f_X(x)$ | | | 0.54 | | 1 |

Are $X$ and $Y$ independent?

(a) Yes, (b) No, (c) Not enough information

Correct answer is (b): No. $f(5, 1) = 0.32 \neq f_X(5)f_Y(1) = (0.54)(0.51) = 0.2754$

**Example**

Imagine you have a card game with a total of 12 cards. Classified in three different categories: 5 cards (money), 4 cards (action), 3 cards (useless). Draw a hand of them, in this case 3 without replacement, and let $X$ = # of useless, $Y$ = # action.

FIND THE JOINT PF AND RANGE

| $y \backslash x$ | 0 | 1 | 2 | 3 | $f_Y(y)$ |
|---|---|---|---|---|---|
| 0 | $10/220$ | $30/220$ | $15/220$ | $1/220$ | $56/220$ |
| 1 | $40/220$ | $60/220$ | $12/220$ | 0 | $112/220$ |
| 2 | $30/220$ | $18/220$ | 0 | 0 | $48/220$ |
| 3 | $4/220$ | 0 | 0 | 0 | $4/220$ |
| $f_X(x)$ | $84/220$ | $108/220$ | $27/220$ | $1/220$ | 1 |

Range: $x \in \{0, 1, 2, 3\}$, $y \in \{0, 1, 2, 3\}$ such that $x + y \le 3$

$f(0, 0)$ (no useless, no action) $= P$(all money)

$$\frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}$$

$f(1, 1)$ (1 useless, 1 action) $= P$(one of each type)

$$\frac{\binom{3}{1}\binom{4}{1}\binom{5}{1}}{\binom{12}{3}} = \frac{60}{220}$$

$$f(x, y) = \frac{\binom{3}{x}\binom{4}{y}\binom{5}{3-x-y}}{\binom{12}{3}}$$

Find marginal probability functions (sum), $X \sim \text{Hyp}(12, 3, 3)$. $Y \sim \text{Hyp}(12, 4, 3)$. Check that the marginal probability functions match.

Are they independent? No (don't have a cartesian product)

Recall: conditional probability:

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

### 9.1.6   Definition (Conditional Probability Function)

The *conditional probability function* of $X$ given $Y = y$ is

$$f(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x, y)}{f_Y(y)}$$

provided $f_Y(y) > 0$.
Similarly, the *conditional probability function* of $Y$ given $X = x$ is

$$f(y \mid x) = \frac{f(x, y)}{f_X(x)}$$

provided $f_X(x) > 0$.

**Example**

What is the probability that someone taking 4 courses is a co-op student?

In other words, $P(Y = 1 \mid X = 4) = \frac{0.1}{0.27} = 0.37$

For 6 courses, $\frac{0.04}{0.05} = 0.80$.

**Example**

If you have 1 action card, find the pf of the number of useless cards.

i.e. the pf of $X \mid Y = 1$

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $f(x \mid 1)$ | $40/112$ | $60/112$ | $12/112$ |

### 9.1.7 Functions of 2 or more random variables

Suppose $U$ is some function of $X$ and $Y$, e.g. $U = X - Y$. To find the pf of $U$.

| $y\backslash x$ | 3 | 4 | 5 | 6 |
|---|---|---|---|---|
| 0 | 3 | 4 | 5 | 6 |
| 1 | 2 | 3 | 4 | 5 |

1. determine the possible values of $U$ for each pair $(x, y)$, so the range is $u \in \{2, 3, 4, 5, 6\}$

2. $f(u)$ is the sum of $f(x, y)$ for all combos that map to $u$.

$$f(u) = \sum_{(x,y) \text{ s.t. } x-y=u} f(x, y)$$

| $u$ | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $f(u)$ | 0.05 | 0.19 | 0.49 | 0.24 | 0.01 |

$u = 2 \to (x = 3, y = 1) = 0.05$

$u = 3 \to (x = 4, y = 1) + (x = 3, y = 0) = 0.1 + 0.09 = 0.19$

Using the earlier table:

| $y\backslash x$ | 3 | 4 | 5 | 6 |
|---|---|---|---|---|
| 0 | 0.09 | 0.17 | 0.22 | 0.07 |
| 1 | 0.05 | 0.1 | 0.32 | 0.04 |

---

<div align="center">LECTURE 28</div>

### 9.1.8 Thought Question

Suppose $X = \#$ apple products and $Y = \#$ Microsoft products (given at least one of each) have a joint pf:

| $y\backslash x$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0.30 | 0.17 | 0.20 |
| 2 | 0.17 | 0.10 | 0.06 |

Find $P(X + Y = 4)$

(a) 0.10, (b) 0.20, (c) 0.30, (d) 0.40, (e) none

Correct answer is (c): $P(X + Y = 4) = (3, 1) + (2, 2) = 0.20 + 0.10 = 0.30$

### 9.1.9 Sums of random variables

Suppose $T = X + Y$, and $X, Y$ are non-negative.

The range of $T$ is $0, 1, \ldots, \max(X) + \max(Y)$ pf of $T$ is

$$f_T(t) = \sum_{x+y=t} \sum f(x, y)$$
$$= f(0, t) + f(1, t - 1) + f(2, t - 2) + \cdots + f(t, 0)$$
$$= \sum_{x=0}^{t} f(x, t - x)$$

If $X$ and $Y$ are independent, then

$$f_T(t) = \sum_{x=0}^{t} f_X(x) f_Y(y)(t-x)$$

This can be used to prove:

- sum of two independent Poisson is a Poisson random variable
- sum of $k$ independent $\mathrm{Geometric}(P)$ is $\mathrm{NB}(k,p)$

## 9.2 Multinomial Distribution

An extension of Binomial, where each independent trial can have $k$ possible outcomes.

The probability of type $i$ is $p_i$ which is constant.

$$p_1 + p_2 + \cdots + p_k = 1$$

We do $n$ trials and let $X_i = $ # of outcome $i$'s that occur.

$$X_1 + X_2 + \cdots + X_k = n$$

where $n$ is the total number of trials.

Then we say $X_1, \ldots, X_k \sim \mathrm{Multinomial}(n, p_1, p_2, \ldots p_k)$.

**REMARK 9.2.1.** $X_k$ can be written as $n - \sum_{i=1}^{k-1} x_i$ and $p_k$ can be written as $1 - \sum_{i=1}^{k-1} p_i$

**Example**

Roll a fair 6-sided die 10 times. $X_1 = $ # 1's $X_2 = $ # composites (4,6) $X_3 = $ # primes (2,3,5)

Find range: $X_i \in \{0, \ldots, n\}$ $n = 10$ in this case. So,

$$X_1 + X_2 + X_3 = 10$$

Find joint pf: $f(x_1, x_2, x_3) = P(X_1 1's, X_2 C's, X_3 P's)$. So,

$$\underbrace{\frac{10!}{x_1! x_2! x_3!}}_{\text{arrangements}} \underbrace{\left(\frac{1}{6}\right)^{x_1} \left(\frac{2}{6}\right)^{x_2} \left(\frac{3}{6}\right)^{x_3}}_{\text{outcomes}}$$

In general,

$$f(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

for $x_1 + \cdots + x_k = n$ OR

$$f(x_1, \ldots, x_{k-1}) = \frac{n!}{x_1! \cdots x_{k-1}!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}}$$

for $x_1 + \cdots + x_{k-1} \leq n$

Find marginal pf of $x_1$.

$$f_1(x_1) = \sum_{x_2}^{x_3} f(x_1, x_2, x_3)$$

$$= \sum_{x_2=0}^{10-x_1} f(x_1, x_2)$$

$$= \sum_{x_2=0}^{10-x_1} \frac{10!}{x_1! x_2! (10 - x_1 - x_2)!} \left(\frac{1}{6}\right)^{x_1} \left(\frac{1}{3}\right)^{x_2} \left(\frac{1}{2}\right)^{(10-x_1-x_2)}$$

$$= \frac{10!}{x_1!(10-x_1)!} \left(\frac{1}{6}\right)^{x_1} \sum_{x_2=0}^{10-x_1} \frac{(10-x_1)!}{x_2!(10-x_1-x_2)!} \left(\frac{1}{3}\right)^{x_2} \left(\frac{1}{2}\right)^{(10-x_1-x_2)}$$

$$= \binom{10}{x_1} \left(\frac{1}{6}\right)^{x_1} \sum_{x_2=0}^{10-x_1} \binom{10-x_1}{x_2} \left(\frac{1}{3}\right)^{x_2} \left(\frac{1}{2}\right)^{(10-x_1-x_2)}$$

$$= \binom{10}{x_1} \left(\frac{1}{6}\right)^{x_1} \left(\frac{1}{3} + \frac{1}{2}\right)^{10-x_1}$$

$$f(x_1) = \binom{10}{x_1} \left(\frac{1}{6}\right)^{x_1} \left(\frac{5}{6}\right)^{10-x_1} \sim \text{Binomial}(10, 1/6)$$

In general:

$$X_i \sim \text{Binomial}(n, p_i)$$

---

<center>LECTURE 30</center>

---

## 9.4  Expectation for Multivariate Distributions: Covariance and Correlation

- Steve $\neq$ Diana
- Midterm # M3-3116
- Covariance
- Correlation

If $X$ and $Y$ are independent:

$$E(XY) = \sum_x \sum_y xy f(x, y) \qquad \text{by defn}$$

$$= \sum_x \sum_y xy f_X(x) f_Y(y)$$

$$= \sum_x x f_X(x) \sum_y y f_Y(y) \qquad X, Y \text{ indep.}$$

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

$\implies Cov(X, Y) = E[XY] - E[X]E[Y]$. Thus, $X, Y$ independent $\implies Cov(X, Y) = 0$, but $Cov(X, Y) = 0$ **does not** mean $X, Y$ are independent.

Uncorrelated variables **could** still be dependent. If, for instance there is a non-linear relationship.

$\sim Cov \neq 0$

clearly a relationship.
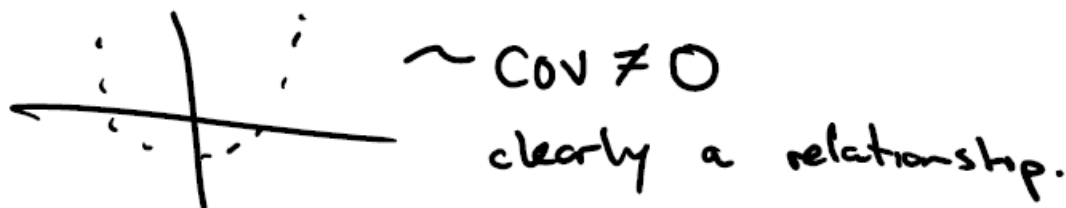
**Other points about covariance**:

- if $Cov > 0$, then $X \uparrow \iff Y \uparrow$ OR $X \downarrow \iff Y \downarrow$.

- if $Cov < 0$, then $X \downarrow \iff Y \uparrow$ OR $X \uparrow \iff Y \downarrow$.

- the magnitude of $Cov(X, Y)$ can't be interpreted. We need to rescale to a restricted range to interpret size.

**Correlation**

### 9.4.1   Definition (Correlation Coefficient)

The *correlation coefficient* $\rho_{xy}$ of $X$ and $Y$ is:
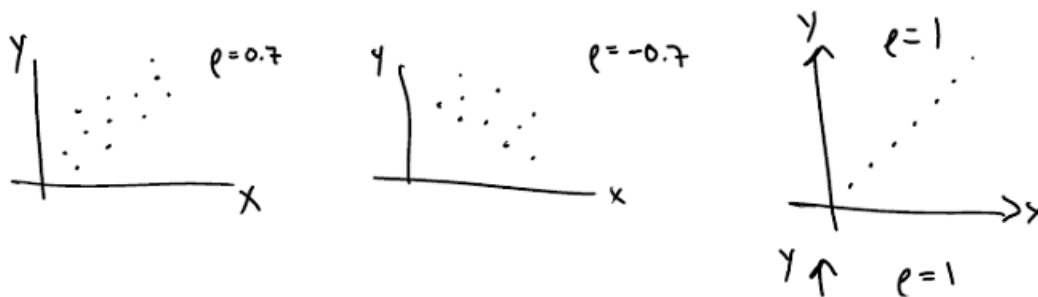
$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

**Notes**

- sign of $Corr =$ sign of $Cov$ for any given $X, Y$.

- $-1 \leq \rho_{xy} \leq 1$ (see course notes)

- only equal to $\pm 1$ if $Y = aX + b$

We interpret the magnitude of the correlation as the **strength** of the linear relationship.



**Important**: correlation **does not** imply causation!

if $\rho = 0.95$, then

$$\begin{cases} X \text{ causes } Y \text{ OR} \\ Y \text{ causes } X \text{ OR} \\ X, Y \text{ are caused by } Z \\ X, Y \text{ are correlated by chance} \end{cases}$$

**Example**

(from past lecture)

Suppose $X = $ # apple products and $Y = $ # Microsoft products (given at least one of each) have a joint pf:

| $y \backslash x$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0.30 | 0.17 | 0.20 |
| 2 | 0.17 | 0.10 | 0.06 |

$Cov(X, Y) = -0.0407$

$Var(X) = E(X^2) - 1.79^2 = 0.6859$

$Var(Y) = E(Y^2) - 1.33^2 = 0.2211$

$Corr(X, Y) = \frac{-0.0407}{\sqrt{0.6959}\sqrt{0.2211}} = -0.1045\ldots$ a weak negative correlation.

**Example**

Roll a fair 6-sided die 10 times. $X_1 = $ # 1's $X_2 = $ # even composites (4,6)

$Cov(X_1, X_2) = 5 - {^{10}/_6} \times {^{10}/_3} = -0.556$

$Var(X_1) = 10\left({^1/_6}\right)\left({^5/_6}\right) = 1.389 \; (npq)$

$Var(X_2) = 10\left({^1/_3}\right)\left({^2/_3}\right) = 2.222 \; (npq)$

$Corr(X, Y) = \frac{-0.556}{\sqrt{1.389}\sqrt{2.222}} = -0.316$

**Next class**: Linear Combinations of Random Variables

---

<div align="center">

LECTURE 31

</div>

---

# 9.5  Linear Combinations of Random Variables

Suppose two variables $X$ and $Y$ have **non-zero** covariance. What can we say?

(a) $X$ and $Y$ are independent. (b) $X$ and $Y$ are not independent. (c) we cannot tell if they are independent.

Same question, but for **zero** covariance.

**Today**

- Linear Combinations of random variables (9.5 & 9.6), which connects nicely to CLT
- a couple of examples

**Friday**

- Indicator Variables

**Rules of Linear Combinations**

$P = \alpha X + (1 - \alpha)Y \rightarrow$ two stocks

$S = 0.05A + 0.3M + 0.15Q + 0.5F$

**Means**

1. $E(aX + bY) = aE(X) + bE(Y)$

2. $E\left(\sum\limits_{i=1}^{n} a_i X_i\right) = \sum\limits_{i=1}^{n} a_i E(X_i) = a_1 E(X_1) + a_2 E(X_2) + \cdots + a_n E(X_n)$

3. $E\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \sum_{i=1}^{n} \frac{\mu}{n} = \frac{n}{n}\mu = \mu$

$X_i$'s all have mean $\mu$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \implies E(\bar{X}) = \mu$

**Variances**

1. $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X,Y)$

2. $Var(\bar{X}) = Var\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

where $X_i$'s are independent. Thus, if we have independent RV's $X_1, \ldots, X_n$ all with $\mu, \sigma^2$, then $E(\bar{X}) = \mu$, $Var(\bar{X}) = \frac{\sigma^2}{n}$

$\boxed{\sigma/\sqrt{n} \rightarrow \text{ std error of the mean}}$

**Note:**

$$
\begin{aligned}
Cov(X,X) &= E[XX] - E[X]E[X] \\
&= E[X^2] - (E[X])^2 \\
&= Var(X)
\end{aligned}
$$

$\implies Corr(X,X) = 1$

**Covariances**

$$Cov(aX + bY, cZ + dW) = ac Cov(X,Z) + ad Cov(X,W) + bc Cov(Y,Z) + bd Cov(Y,W)$$

## 9.6 Linear Combinations of Normal Random Variables

**Claim:** If $X_i \sim N(\mu_i, \sigma_i^2)$ for $i = 1, 2, \ldots, n$ are random variables, then

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

$$\boxed{X_i \sim N(\mu, \sigma^2) \implies \bar{X} \sim N(\mu, \sigma^2/n)}$$

**Example**

Weight of a cat $C \sim N(4.1, 1.6^2)$, weight of a dog $D \sim (9.4, 3.6^2)$. Find the probability that a cat weighs more than a dog.

**Solution.**

$P(C > D) \implies P(C - D > 0) \rightarrow C - D \sim N(4.1 - 9.4, 1.6^2 + (-3.6)^2)$

$$
\begin{aligned}
&= P\left(\frac{C - D - (-5.3)}{\sqrt{15.52}} > \frac{0 - (-5.3)}{\sqrt{15.52}}\right) \\
&= P(Z > 1.35) \\
&= 1 - 0.91149 \\
&= 0.08851
\end{aligned}
$$

**Example**

Heights of cats are $N(24, 1.5^2)$. Find probability a cat has height within 1cm of average.

**Solution.**

$$P(23 < X < 25) = P\left(\frac{23 - 24}{1.5} < \frac{X - 24}{1.5} < \frac{25 - 24}{1.5}\right)$$
$$= P(-0.67 < Z < 0.67)$$
$$= 2(0.74857) - 1$$
$$= 0.49714$$

**Example**

Find the probability the **average height** of 5 cats is within 1cm of average.

**Solution.**

$$\bar{X} = \sum_{i=1}^{5} X_i \sim N(24, {}^{1.5^2}/_5)$$

$$P\left(\left|\bar{X} - 24\right| < 1\right) = P(23 < \bar{X} < 25)$$
$$= P\left(\frac{23 - 24}{1.5/\sqrt{5}} < Z < \frac{25 - 25}{1.5/\sqrt{5}}\right)$$
$$= P(-1.49 < Z < 1.49)$$
$$= 0.86378$$

---

## LECTURE 32

**Today**

1. Quiz #3 - Nov. 29th, 7-8pm [Sec. 9.2-9.7 except 9.3 (Markov Chains)]
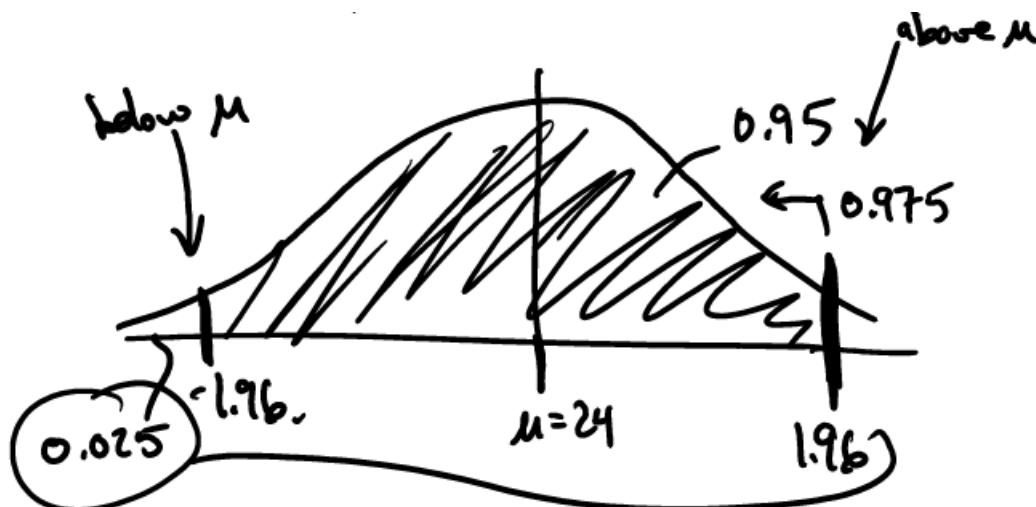
2. Cat exercise

3. Indicator variables

**Cat Exercise**

How many cats would you need to have for a $0.95$ probability that the average height is within 1cm of the true average?

**Solution.** Let $X$ be the height of the cat.

$$X \sim N(24, 1.5^2)$$

(finding the middle 0.95, tails are $0.025$ each; in the table we look for $0.975$)

$$P\left(\left|\bar{X} - 24\right| < 1\right) = 0.95$$

$$P(0.975) = 1.96$$

$$Z = \pm 1.96 \implies Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \implies 1.96 = \frac{25 - 24}{1.5/\sqrt{n}}$$

$n = 8.64 = 9$

## 9.7 Indicator Variables

A tool you can use to evaluate more complicated distributions.

### 9.7.1 Definition (Indicator Variable)

An *indicator variable* (Bernoulli variables)

$$I_A = \begin{cases} 1, \text{ if } A \text{ occurs} \\ 0, \text{ if } A \text{ does not occur} \end{cases}$$

$E(I_A) = 1P(A) + 0(1 - P(A)) = P(A)$

$E(I_A^2) = 1^2 P(A) + 0^2(1 - P(A)) = P(A)$

$Var(I_A) = E(I_A^2) - (E(I_A))^2 = P(A) - (P(A))^2 = P(A)(1 - P(A))$

$$I_B = \begin{cases} 1, \text{ if } B \text{ occurs} \\ 0, \text{ if } B \text{ does not occur} \end{cases}$$

$Cov(I_A, I_B)$

| $I_B \backslash I_A$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

$$I_A I_B = \begin{cases} 0 \\ 0 \\ 0 \\ 1, \text{ if } A \text{ and } B \text{ occur} \end{cases} \text{otherwise}$$

$E(A) = P(A)$

$E(B) = P(B)$

$E(I_A I_B) = P(AB)$

$$Cov(I_A, I_B) = P(AB) - P(A)P(B)$$

**REMARK 9.7.1.** If $A$ and $B$ are independent, $I_A$ and $I_B$ will be uncorrelated.

1. Let $X \sim \text{Binomial}(n, p)$ use indicator variables to find $\mu$ and $\sigma^2$

Let

$$X_i = \begin{cases} 1, \text{ if trial } i \text{ is a success} \\ 0, \text{ if trial } j \text{ is a failure} \end{cases}$$

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

$E(X_i) = p$

$E(X) = E(\sum\limits_{i=1}^{n} X_i) = \sum\limits_{i=1}^{n} E(X_i) = np = \mu$

$Var(X) = \sum\limits_{i=1}^{n} Var(X_i) = np(1-p) = \sigma^2$

2. $X \sim \text{Hyp}(N, r, n)$; reminder: ($N$ trials, $r$ S's, $n$ selections) Let

$$X_i = \begin{cases} 1, \text{ if object } i \text{ is a success (S)} \\ 0, \text{ if object } j \text{ is a failure (F)} \end{cases}$$

$E(X_i) = P(\text{select object is an S}) = {}^r/_N \text{ [from } E(I_A) = P(A)]$

$Var(X_i) = {}^r/_N(1 - {}^r/_N) \text{ [from } Var(I_A) = P(A) - (1 - P(A))]$

$$Cov(X_i, X_j) = P(\text{objects } i \text{ and } j \text{ are S's}) - \left(\frac{r}{N}\right)\left(\frac{r}{N}\right)$$

$$= \frac{\binom{r}{2}}{\binom{N}{2}} - \left(\frac{r}{N}\right)^2$$

$$= \frac{r(r-1)}{N(N-1)} - \frac{r^2}{N^2}$$

$$= -\frac{r(N-r)}{N^2(N-1)} < 0$$

$X = \sum\limits_{i=1}^{n} X_i \implies E(X) = \sum\limits_{i=1}^{n} {}^r/_N = {}^{nr}/_N$

$Var(X) = \sum\limits_{i=1}^{n} Var(X_i) + 2 \sum\limits_{i<j} Cov(X_i, X_j)$

where $Var(X)$ comes from properties in 9.5, the first term has $n$ terms, the second term has $\binom{n}{2}$ terms.

$$= \frac{nr}{N}\left(1 - \frac{r}{N}\right) + 2\left[\frac{n(n-1)}{2}\right]\left[-\frac{r(N-r)}{N^2(N-1)}\right] = \frac{nr}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

where $\frac{nr}{N}\left(1 - \frac{r}{N}\right)$ is $\text{Binomial}(n, r/N)$ and the term $\left(\frac{N-n}{N-1}\right)$ reduces variance because we are sampling without replacement.

**Example**

$N$ messages come to a server which randomly gives one message to each intended recipient. Find the mean and variance of the # of correctly delivered messages.

**Solution.**

Let

$$X_i = \begin{cases} 1, & \text{if msg } i \text{ is correct} \\ 0, & \text{otherwise} \end{cases}$$

then $X = \sum\limits_{i=1}^{N} X_i$

$$\left.\begin{array}{l} E(X_i) = P(\text{msg } i \text{ is correct}) = \dfrac{1}{N} \\[2mm] Var(X_i) = \dfrac{1}{N}\left(1 - \dfrac{1}{N}\right) \end{array}\right\} \text{properties of indicator variables}$$

$$E(X_i X_j) = P(i \text{ correct})P(j \text{ correct}|i \text{ correct})$$
$$= \frac{1}{N}\frac{1}{N-1}$$
$$= \frac{1}{N(N-1)}$$

$Cov(X_i, X_j) = \frac{1}{N(N-1)} - \left(\frac{1}{N}\right)^2 = \frac{1}{N^2(N-1)} > 0$

$E(X) = \sum\limits_{i=1}^{N} E(X_i) = \sum\limits_{i=1}^{N} \frac{1}{N} = N\frac{1}{N} = 1$

$Var(X) = 1 \rightarrow$ for Diana

---

**LECTURE 33**