BTRY 4080 / STSCI 4080, Fall 2009

Theory of Probability

Instructor: Ping Li

Department of Statistical Science

Cornell University

General Information

- Lectures: Tue, Thu 10:10-11:25 am, Malott Hall 253
- Section 1: Mon 2:55 - 4:10 pm, Warren Hall 245
- Section 2: Wed 1:25 - 2:40 pm, Warren Hall 261
- Instructor: Ping Li, pingli@cornell.edu,
  Office Hours: Tue, Thu 11:25 am -12 pm, 1192, Comstock Hall
- TA: Xiao Luo, lx42@cornell.edu. Office hours:
  (1) Mon, 4:10 - 5:10pm Warren Hall 245;
  (2) Wed, 2:40 - 3:40pm, Warren Hall 261.
- Prerequisites: Math 213 or 222 or equivalent; a course in statistical methods is recommended
- Textbook: Ross, A First Course in Probability, 7th edition, 2006

- Exams: Locations TBD
  - Prelim 1: In Class, September 29, 2009
  - Prelim 2: In Class, October 20, 2009
  - Prelim 3: In Class, TBD
  - Final Exam: Wed. December 16, 2009, from 2pm to 4:30pm.
  - Policy: Close book, close notes

- Homework: Weekly
  - Please turn in your homework either in class or to BSCB front desk (Comstock Hall, 1198).
  - No late homework will be accepted.
  - Before computing your overall homework grade, the assignment with the lowest grade (if $\geq 25\%$) will be dropped, the one with the second lowest grade (if $\geq 50\%$) will also be dropped.
  - It is the students' responsibility to keep copies of the submitted homework.

- Grading:

  1. The homework: 30%
  2. The prelim with the lowest grade: 5%
  3. The other two prelims: 15% each
  4. The final exam 35%

- Course Letter Grade Assignment

  A $\approx 90\%$ (in the absolute scale)

  C $\approx 60\%$ (in the absolute scale)

In borderline cases, participation in section and class interactions will be used as a determining factor.

Material

Topic	Textbook
Combinatorics	1.1 - 1.6
Axioms of Probability	2.1 - 2.5
Conditional Probability and Independence	3.1 - 3.5
Discrete Random Variables	4.1 - 4.9
Continuous Random Variables	5.1 - 5.7
Jointly Distributed Random Variables	6.1 - 6.7
Expectation	7.1 - 7.9

As time permits: Limit Theorems 8.1 - 8.6,

Poisson Processes and Markov Chains 9.1 - 9.2 Simulation 10.1 - 10.4

Probability Is Increasingly Popular & Truly Useful

An Example: Probabilistic Counting

Consider two sets $S_1, S_2 \in \Omega = \{0, 1, 2, ..., D-1\}$, for example

$$D = 2000$$

$$S_1 = \{0, 15, 31, 193, 261, 495, 563, 919\}$$

$$S_2 = \{5, 15, 19, 193, 209, 325, 563, 1029, 1921\}$$

Q: What is the size of intersection $a = |S_1 \cap S_2|$?

A: Easy! $a = |\{15, 193, 563\}| = 3$.

Challenges

- What if the size of space $D = 2^{64}$?
- What if there are 10 billion of sets, instead of just 2?
- What if we only need approximate answers?
- What if we need the answers really quickly?
- What if we are mostly interested in sets that are highly similar?
- What if the sets are dynamic and frequently updated?
- ...

Search Engines

Google/Yahoo!/Bing all collected at least 10^{10} (10 billion) web pages.

Two examples (among many tasks)

- **(Very quickly) Determine if two Web pages are nearly duplicates**

  A document can be viewed as a collection (set) of words or contiguous words.

  This is a counting of set intersection problem.

- **(Very quickly) Determine the number of co-occurrences of two words**

  A word can be viewed as a set of documents containing that word

  This is also a counting of set intersection problem.

Term Document Matrix

	Doc 1	Doc 2						Doc m
Word 1	1	0	0	1	0	0	0	1
Word 2	0	1	0	1	0	0	1	0
Word 3								
Word 4								
Word n								

$\geq 10^{10}$ documents (Web pages).

$\geq 10^{5}$ common English words. (What about total number of contiguous words?)

One Method for Probabilistic Counting of Intersections

Consider two sets $S_1, S_2 \subset \Omega = \{0, 1, 2, ..., D-1\}$

$$S_1 = \{0, 15, 31, 193, 261, 495, 563, 919\}$$

$$S_2 = \{5, 15, 19, 193, 209, 325, 563, 1029, 1921\}$$

One can first generate a **random permutation** π:

$$\pi: \quad \Omega = \{0, 1, 2, ..., D-1\} \longrightarrow \{0, 1, 2, ..., D-1\}$$

Apply π to sets S_1 and S_2 (and all other sets). For example

$$\pi\left(S_1\right) = \{13, 139, 476, 597, 698, 1355, 1932, 2532\}$$

$$\pi\left(S_2\right) = \{13, 236, 476, 683, 798, 1456, 1968, 2532, 3983\}$$

Store only the minimums

$$\min\{\pi(S_1)\} = 13, \qquad \min\{\pi(S_2)\} = 13$$

Suppose we can theoretically calculate the probability

$$P = \mathbf{Pr}\left(\min\{\pi(S_1)\} = \min\{\pi(S_2)\}\right)$$

S_1 and S_2 are more similar $\Longrightarrow$ larger P (i.e., closer to 1).

Generate (small) k random permutations and every times store the minimums.

Count the number of matches to estimate the similarity between S_1 and S_2.

Later, we can easily show that

$$P = \mathbf{Pr}\left(\min\{\pi\left(S_1\right)\} = \min\{\pi\left(S_2\right)\}\right) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = \frac{a}{f_1 + f_2 - a}$$

where $a = |S_1 \cap S_2|, \quad f_1 = |S_1|, \quad f_2 = |S_2|$.

Therefore, we can estimate P as a **binomial probability** from k permutations.

Combinatorial Analysis

Ross: Chapter 1, Sections 1.1 - 1.6

Basic Principles of Counting, Section 1.2

1. **The multiplicative principle**

   A project requires two sub-tasks, A and B.

   There are m ways to complete A, and there are n ways to complete B.

   Q: How many different ways to complete this project?

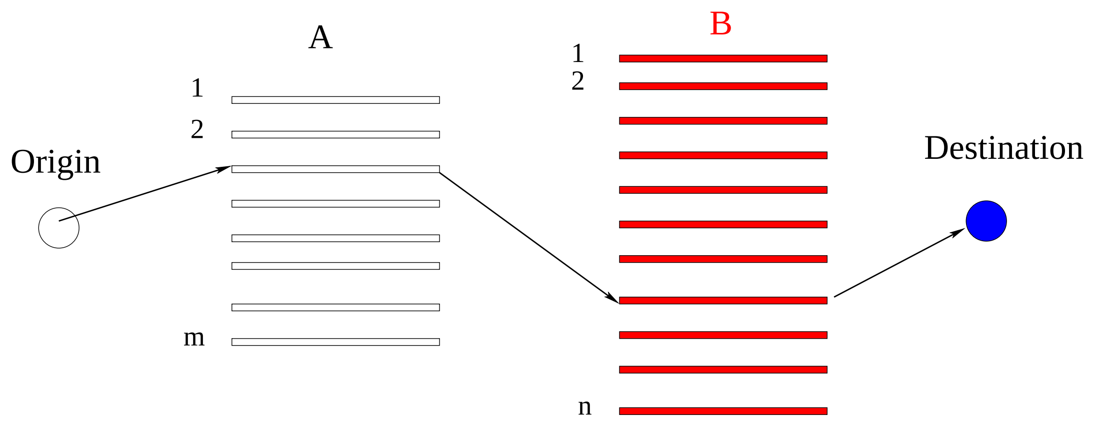
2. **The additive principle**

   A project can be completed either by doing task A, or by doing task B.

   There are m ways to complete A, and there are n ways to complete B.

   Q: How many different ways to complete this project?

The Multiplicative Principle



Q: How many different routes from the origin to reach the destination?

Route 1: $\{$ A:1 + B:1 $\}$

Route 2: $\{$ A:1 + B:2 $\}$

...

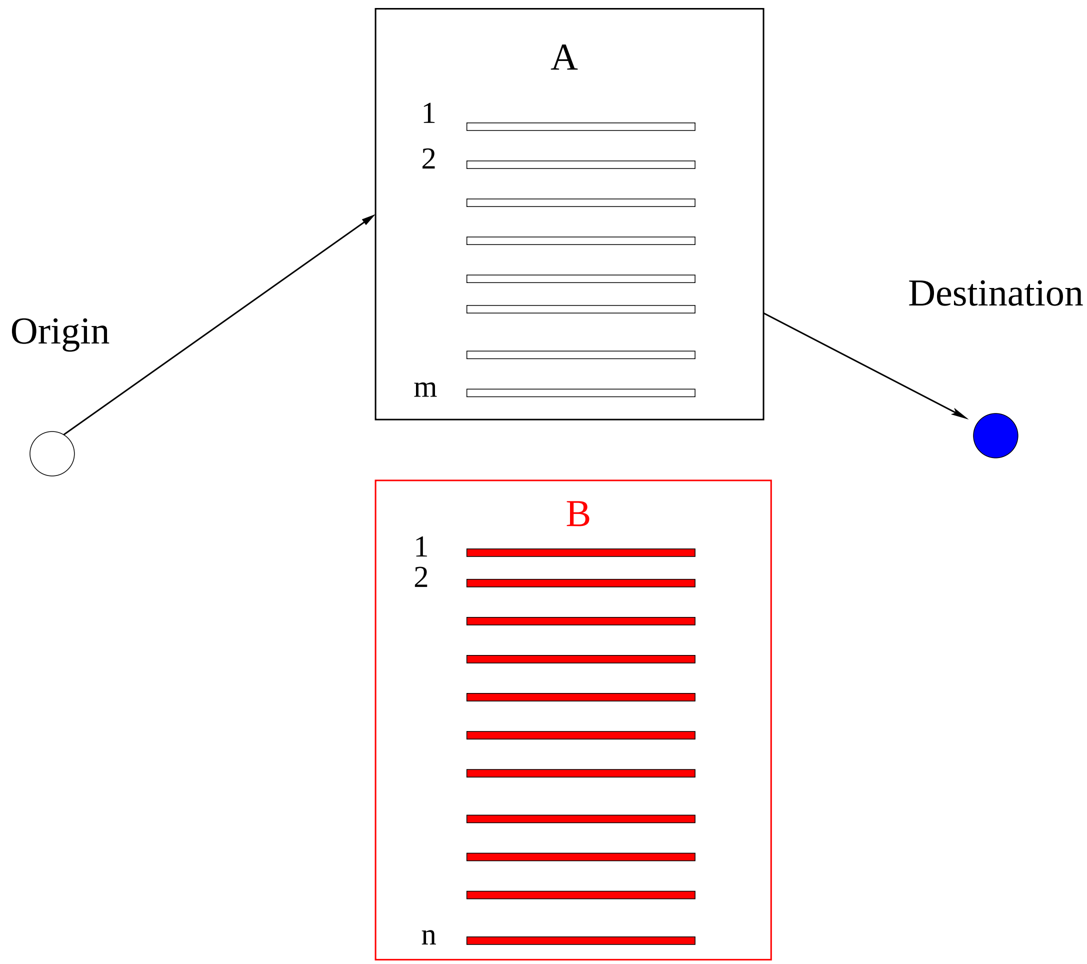
Route n: $\{$ A:1 + B:n $\}$

Route n+1: $\{$ A:2 + B:1 $\}$

...

Route ? $\{$ A:m + B:n $\}$

The additive Principle



Q: How many different routes from the origin to reach the destination?

Example:

A college planning committee consists of 3 freshmen, 4 sophomores, 5 juniors, and 2 seniors. A subcommittee of 4, consisting of 1 person from each class, is to be chosen.

Q: How many different subcommittees are possible?

Multiplicative or additive?

Example:

In a 7-place license plate, the first 3 places are to be occupied by letters and the final 4 by numbers, e.g., ITH-0827.

Q: How many different 7-place license plates are possible?

Multiplicative or additive?

Example:

How many license plates would be possible if repetition among letters or numbers were prohibited?

Permutations and Combinations, Section 1.3, 1.4

Need to select 3 letters from 26 English letter, to make a license plate.

Q: How many different arrangements are possible?

The answer differs if

- Order of 3 letters matters $\Longrightarrow$ Permutation

  e.g., ITH, TIH, HIT, are all different.

- Order of 3 letters does not matter $\Longrightarrow$ Combination

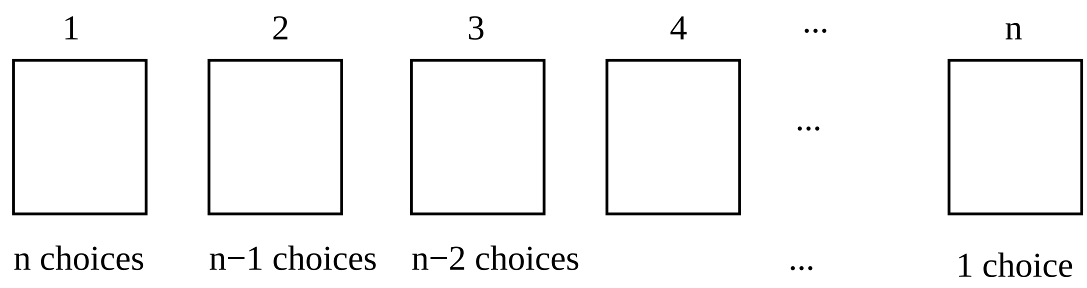
Permutations

Basic formula:

Suppose we have n objects. How many different permutations are there?

$$n \times (n-1) \times (n-2) \times ... \times 1 = n!$$

Why?



Which counting principle?: Multiplicative? additive?

Example:

A baseball team consists of 9 players.

Q: How many different batting orders are possible?

Example:

A class consists of 6 men and 4 women. The students are ranked according to their performance in the final, assuming no ties.

Q: How many different rankings are possible?

Q: If the women are ranked among themselves and the men among themselves how many different rankings are possible?

Example:

How many different letter arrangements can be formed using P,E,P ?

The brute-force approach

- Step 1: Permutations assuming P_1 and P_2 are different.

  P_1 P_2 E　　P_1 E P_2

  P_2 P_1 E　　P_2 E P_1

  E P_1 P_2　　E P_2 P_1

- Step 2: Removing duplicates assuming P_1 and P_2 are the same.

  P P E　　P E P　　E P P

Q: How many duplicates?

Example:

How many different letter arrangements can be formed using P,E,P,P,E,R ?

Answer:

$$\frac{6!}{3! \times 2! \times 1!} = 60$$

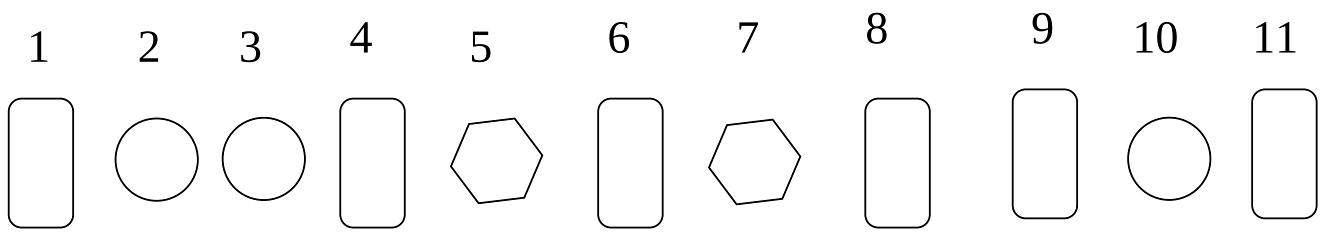
Why?

A More General Example of Permutations:

How many different permutations of n objects, of which n_1 are identical, n_2 are identical, ..., n_r are identical?

- Step 1: Assuming all n objects are distinct, there are $n!$ permutations.
- Step 2: For each permutation, there are $n_1! \times n_2! \times ... \times n_r!$ ways to switch objects and still remain the same permutation.
- Step 3: The total number of (unique) permutations are

$$\frac{n!}{n_1! \times n_2! \times ... \times n_r!}$$



How many ways to switch the elements and remain the same permutation?

Example:

A signal consists of 9 flags hung in a line, which can be made from a set of 4 white flags, 3 red flags, and 2 blue flags. All flags of the same color are identical.

Q: How many different signals are possible?

Combinations

Basic question of combination: The number of different groups of r objects that could be formed from a total of n objects.

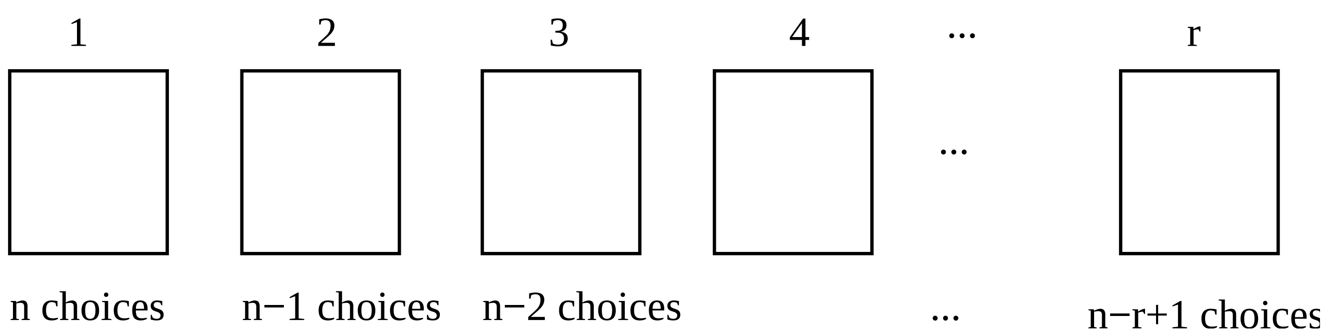
Basic formula of combination:

$$\binom{n}{r} = \frac{n!}{(n-r)! \times r!}$$

Pronounced as "n-choose-r"

Basic Formula of Combination

Step 1: When orders are relevant, the number of different groups of r objects formed from n objects $= n \times (n-1) \times (n-2) \times ... \times (n-r+1)$



Step 2: When orders are NOT relevant, each group of r objects can be formed by $r!$ different ways.

Step 3: Therefore, when orders are NOT relevant, the number of different groups of r objects would be

$$\frac{n \times (n-1) \times ... \times (n-r+1)}{r!}$$

$$=\frac{n \times (n-1) \times ... \times (n-r+1) \times (n-r) \times (n-r-1) \times ... \times 1}{r! \times (n-r) \times (n-r-1) \times ... \times 1}$$

$$=\frac{n!}{(n-r)! \times r!}$$

$$=\binom{n}{r}$$

Example:

A committee of 3 is to be formed from a group of 20 people.

Q: How many different committees are possible?

Example:

How many different committees consisting of 2 women and 3 men can be formed, from a group of 5 women and 7 men?

Example (1.4.4c):

Consider a set of n antennas of which m are defective and $n-m$ are functional.

Q: How many linear orderings are there in which no two defectives are consecutive?

A: $\binom{n-m+1}{m}$.

Line up $n-m$ functional antennas. Insert (at most) one defective antenna between two functional antennas. Considering the boundaries, there are in total $n-m+1$ places to insert defective antennas.

Useful Facts about Binomial Coefficient $\binom{n}{r}$

1. $\binom{n}{0} = 1$, and $\binom{n}{n} = 1$. Recall $0! = 1$.

2. $\binom{n}{r} = \binom{n}{n-r}$

3. $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$

Proof 1 by algebra

$$\binom{n-1}{r-1} + \binom{n-1}{r} = \frac{(n-1)!}{(n-r)!(r-1)!} + \frac{(n-1)!}{(n-r-1)!r!} = \frac{(n-1)!(r+n-r)}{(n-r)!r!}$$

Proof 2 by a combinatorial argument

A selection of r objects either contains object 1 or does not contain object 1.

4. Some computational tricks

- Avoid factorial $n!$, which is easily very large. $170! = 7.2574 \times 10^{306}$

- Take advantage of $\binom{n}{r} = \binom{n}{n-r}$

- Convert products to additions

$$\binom{n}{r} = \frac{n(n-1)(n-2)...(n-r+1)}{r!}$$
$$= \exp\left\{\sum_{j=0}^{r-1}\log(n-j) - \sum_{j=1}^{r}\log j\right\}$$

- Use accurate approximations of the log factorial function ($\log n!$). Check `http://en.wikipedia.org/wiki/Factorial`.

The Binomial Theorem

The Binomial Theorem:

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

Example for $n = 2$:

$$(x+y)^2 = x^2 + 2xy + y^2, \qquad \binom{2}{0} = \binom{2}{2} = 1, \binom{2}{1} = 2$$

Proof of the Binomial Theorem by Induction

Two key components in proof by induction:

1. Verify a base case, e.g., $n = 1$.
2. Assume the case $n - 1$ is true, prove for the case n.

Proof:

The base case is trivially true (but always check ! !)

Assume

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

Then

$$
\begin{aligned}
&\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} = \sum_{k=1}^{n-1}\binom{n}{k}x^k y^{n-k} + x^n + y^n\\
=&\sum_{k=1}^{n-1}\binom{n-1}{k}x^k y^{n-k} + \sum_{k=1}^{n-1}\binom{n-1}{k-1}x^k y^{n-k} + x^n + y^n\\
=&y\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-1-k} - y^n + \sum_{k=1}^{n-1}\binom{n-1}{k-1}x^k y^{n-k} + x^n + y^n\\
=&y(x+y)^{n-1} + \sum_{j=0}^{n-2}\binom{n-1}{j}x^{j+1}y^{n-1-j} + x^n, \qquad \text{(let } j = k-1)\\
=&y(x+y)^{n-1} + x\sum_{j=0}^{n-1}\binom{n-1}{j}x^{j}y^{n-1-j} - x^n + x^n\\
=&y(x+y)^{n-1} + x(x+y)^{n-1} = (x+y)^n
\end{aligned}
$$

Multinomial Coefficients

A set of n distinct items is to be divided into r distinct groups of respective size n_1, n_2, ..., n_r, where $\sum_{i=1}^{r} n_i = n$.

Q: How many different divisions are possible?

Answer:

$$\binom{n}{n_1, n_2, ..., n_r} = \frac{n!}{n_1!n_2!...n_r!}$$

Proof

Step 1: There are $\binom{n}{n_1}$ ways to form the first group of size n_1.

Step 2: There are $\binom{n-n_1}{n_2}$ ways to form the second group of size n_2.

...

Step r: There are $\binom{n-n_1-n_2-...-n_{r-1}}{n_r}$ ways to form the last group of size n_r.

Therefore, the total number of divisions should be

$$\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times ... \times \binom{n-n_1-n_2-...-n_{r-1}}{n_r}$$

$$= \frac{n!}{(n-n_1)!n_1!} \times \frac{(n-n_1)!}{(n-n_1-n_2)!n_2!} \times ... \times \frac{(n-n_1-...-n_{r-1})!}{(n-n_1-...n_{r-1}-n_r)!n_r!}$$

$$= \frac{n!}{n_1!n_2!...n_r!}$$

Example:

A police department in a small city consists of 10 officers. The policy is to have 5 patrolling the streets, 2 working full time at the station, and 3 on reserve at the station.

Q: How many different divisions of 10 officers into 3 groups are possible?

The Number of Integer Solutions of Equations

Suppose x_1, x_2, ..., x_r, are positive, and $x_1 + x_2 + ... + x_r = n$.

Q: How many distinct vectors $(x_1, x_2, ..., x_r)$ can satisfy the constraints.

A:

$$\binom{n-1}{r-1}$$

Line up n 1's. Draw $r-1$ lines from in total $n-1$ places to form r groups (numbers).

Suppose x_1, x_2, ..., x_r, are non-negative, and $x_1 + x_2 + ... + x_r = n$.

Q: How many distinct vectors $(x_1, x_2, ..., x_r)$ can satisfy the constraints.

A:

$$\binom{n+r-1}{r-1},$$

by padding r zeros after n one's.

One can also let $y_i = x_i + 1$, $i = 1$ to r. Then the problem becomes finding positive integers $(y_1, ..., y_r)$ so that $y_1 + y_2 + ... + y_r = n + r$.

Example:

8 identical blackboards are to be divided among 4 schools.

Q (a): How many divisions are possible?

Q (b): How many, if each school must receive at least 1 blackboard?

More Examples for Chapter 1

From a group of 8 women and 6 men a committee consisting of 3 men and 3 women is to be formed.

(a): How many different committees are possible?

(b): How many, if 2 of the men refuse to serve together?

(c): How many, if 2 of the women refuse to serve together?

(d): How many, if 1 man and 1 woman refuse to serve together?

Solution:

(a): How many different committees are possible?

$$\binom{8}{3} \times \binom{6}{3} = 56 \times 20 = 1120$$

(b): How many, if 2 of the men refuse to serve together?

$$\binom{8}{3} \times \left[\binom{4}{3} + \binom{2}{1}\binom{4}{2}\right] = 56 \times 16 = 896$$

(c): How many, if 2 of the women refuse to serve together?

$$\left[\binom{6}{3} + \binom{2}{1}\binom{6}{2}\right] \times \binom{6}{3} = 50 \times 20 = 1000$$

(d): How many, if 1 man and 1 woman refuse to serve together?

$$\binom{7}{3}\binom{5}{3} + \binom{7}{3}\binom{5}{2} + \binom{7}{2}\binom{5}{3} = 350 + 350 + 210 = 910$$

Or $\binom{8}{3}\binom{6}{3} - \binom{7}{2}\binom{5}{2} = 910$

Fermat's Combinatorial Identity

Theoretical Exercise 11.

$$\binom{n}{k} = \sum_{i=k}^{n} \binom{i-1}{k-1}, \qquad n \geq k$$

There is an interesting proof (different from the book) using the result for the number of integer solutions of equations, by a recursive argument.

Proof:

Let $f(m, r)$ be the number of positive integer solutions, i.e.,
$x_1 + x_2 + ... + x_r = m$.

We know $f(m, r) = \binom{m-1}{r-1}$.

Suppose $m = n + 1$ and $r = k + 1$, Then

$$f(n+1, k+1) = \binom{m-1}{r-1} = \binom{n}{k}$$

x_1 can take values from 1 to $(n + 1) - k$, because $x_i \geq 1$, $i = 1$ to $k + 1$.

Decompose the problem into $(n + 1) - k$ sub-problems by fixing a value for x_1.

If $x_1 = 1$, then $x_2 + x_3 + ... + x_{k+1} = m - 1 = n$.
The total number of positive integer solutions is $f(n, k)$.

If $x_1 = 1$, there are still $f(n,k)$ possibilities, i.e., $\binom{n-1}{k-1}$.

If $x_1 = 2$, there are still $f(n-1,k)$ possibilities, i.e., $\binom{n-2}{k-1}$

...

If $x_1 = (n+1) - k$, there are still $f(k,k)$ possibilities, i.e., $\binom{k-1}{k-1}$

Thus, by additive counting principle,

$$f(n+1,k+1) = \sum_{j=1}^{n+1-k} f(n+1-j,k) = \sum_{j=1}^{n+1-k} \binom{n-j}{k-1}$$

But, is this what we want? **Yes!**

By additive counting principle,

$$f(n+1,k+1) = \sum_{j=1}^{n+1-k} f(n+1-j,k)$$

$$= \sum_{j=1}^{n+1-k} \binom{n-j}{k-1}$$

$$= \binom{n-1}{k-1} + \binom{n-2}{k-1} + ... + \binom{k-1}{k-1}$$

$$= \binom{k-1}{k-1} + \binom{k}{k-1} + ... + \binom{n-1}{k-1}$$

$$= \sum_{i=k}^{n} \binom{i-1}{k-1}$$

Example: A company has 50 employees, 10 of whom are in management positions.

Q: How many ways are there to choose a committee consisting of 4 people, if there must be at least one person of management rank?

A:

$$\binom{50}{4} - \binom{40}{4} = 138910 = \sum_{k=1}^{4}\binom{10}{k}\binom{40}{4-k}$$

——–-

Can we also compute the answer using $\binom{10}{1}\binom{49}{3}$ = 184240? **No**.

It is not correct because the two sub-tasks, $\binom{10}{1}$ and $\binom{49}{3}$, are overlapping; and hence we can not simply apply the multiplicative principle.

Axioms of Probability

Chapter 2, Sections 2.1 - 2.5, 2.7

Section 2.1

In this chapter we introduce the concept of the probability of an event and then show how these probabilities can be computed in certain situations. As a preliminary, however, we need the concept of the sample space and the events of an experiment.

Sample Space and Events, Section 2.2

Definition of Sample Space:

The sample space of an experiment, denoted by S, is the set of all possible outcomes of the experiment.

Example:

$S = \{\text{girl}, \text{boy}\}$, if the output of an experiment consists of the determination of the gender of a newborn.

Example:

If the experiment consists of flipping two coins (H/T), the the sample space

$S = \{(H, H), (H, T), (T, H), (T, T)\}$

Example:

If the outcome of an experiment is the order of finish in a race among 7 horses having post positions, 1, 2, 3, 4, 5, 6, 7, then

$S = \{\text{all } 7! \text{ permutations of } (1, 2, 3, 4, 5, 6, 7)\}$

Example:

If the experiment consists of tossing two dice, then

$S = \{(i, j),\ \ i, j = 1, 2, 3, 4, 5, 6\}$, What is the size $|S|$?

Example:

If the experiment consists of measuring (in hours) the lifetime of a transistor, then

$S = \{t : \ 0 \leq t \leq \infty\}$.

Definition of Events:

Any subset E of the sample space S is an event.

Example:

$S = \{(H,H),(H,T),(T,H),(T,T)\}$,

$E = \{(H,H)\}$ is an event. $E = \{(H,T),(T,T)\}$ is also an event

Example:

$S = \{t: \ 0 \leq t \leq \infty\}$,

$E = \{t: \ 1 \leq t \leq 5\}$ is an event,

But $\{t: \ -10 \leq t \leq 5\}$ is NOT an event.

Set Operations on Events

Let A, B, C be events of the same sample space S

$A \subset S$, $B \subset S$, $C \subset S$. Notation for subset: $\subset$

Set Union:

$A \cup B$ = the event which occurs if either A OR B occurs.

Example:

$S = \{1, 2, 3, 4, 5, ...\}$ = the set of all positive integers.

$A = \{1\}, \qquad B = \{1, 3, 5\}, \qquad C = \{2, 4, 6\},$

$A \cup B = \{1, 3, 5\}, \quad B \cup C = \{1, 2, 3, 4, 5, 6\},$

$A \cup B \cup C = \{1, 2, 3, 4, 5, 6\}$

Set Intersection:

$A \cap B = AB$ = the event which occurs if both A AND B occur.

Example:

$S = \{1, 2, 3, 4, 5, ...\} =$ the set of all positive integers.

$A = \{1\}, \qquad B = \{1, 3, 5\}, \qquad C = \{2, 4, 6\},$

$$A \cap B = \{1\}, \qquad B \cap C = \{\} = \emptyset, \qquad A \cap B \cap C = \emptyset$$

Mutually Exclusive:

If $A \cap B = \emptyset$, then A and B are mutually exclusive.

Set Complement:

$A^c = S - A$ = the event which occurs only if A does not occur.

Example:

$S = \{1, 2, 3, 4, 5, ...\} =$ the set of all positive integers.

$A = \{1, 3, 5, ...\}$ = the set of all positive odd integers.

$B = \{2, 4, 6, ...\}$ = the set of all positive even integers.

$$A^c = S - A = B, \qquad B^c = S - B = A.$$

Laws of Set Operations

Commutative Laws:

$$A \cup B = B \cup A \qquad A \cap B = B \cap A$$

Associative Laws:

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

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

Distributive Laws:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C) = AC \cup BC$$

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

These laws can be verified by Venn diagrams.

DeMorgan's Laws

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

Proof of $\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$

Let $A = \left(\bigcup_{i=1}^{n} E_i\right)^c$ and $B = \bigcap_{i=1}^{n} E_i^c$

1. Prove $A \subseteq B$

   Suppose $x \in A$. $\Longrightarrow$ x does not belong to any E_i.
   $\Longrightarrow$ x belongs to every E_i^c. $\Longrightarrow$ $x \in \bigcap_{i=1}^{n} E_i^c = B$.

2. Proof $B \subseteq A$

   Suppose $x \in B$. $\Longrightarrow$ x belongs to every E_i^c.
   $\Longrightarrow$ x does not belong to any E_i. $\Longrightarrow$ $x \in \left(\bigcup_{i=1}^{n} E_i\right)^c = A$.

3. Consequently $A = B$

Proof of $\left(\bigcap_{i=1}^n E_i\right)^c = \bigcup_{i=1}^n E_i^c$

Let $F_i = E_i^c$.

We have proved $\left(\bigcup_{i=1}^n F_i\right)^c = \bigcap_{i=1}^n F_i^c$

$$\left(\bigcap_{i=1}^n E_i\right)^c = \left(\bigcap_{i=1}^n F_i^c\right)^c$$

$$= \left(\left(\bigcup_{i=1}^n F_i\right)^c\right)^c = \bigcup_{i=1}^n F_i = \bigcup_{i=1}^n E_i^c$$

Axioms of Probability, Section 2.3

For each event E in the sample space S, there exists a value $P(E)$, referred to as the probability of E.

$P(E)$ satisfies three axioms.

Axiom 1: $0 \leq P(E) \leq 1$

Axiom 2: $P(S) = 1$

Axiom 3: For any sequence of mutually exclusive events, E_1, E_2, ..., (that is, $E_iE_j = \emptyset$ whenever $i \neq j$), then

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

Consequences of Three Axioms

Consequence 1: $P(\emptyset) = 0$.

$$P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset) \Longrightarrow P(\emptyset) = 0$$

Consequence 2:

If $S = \bigcup_{i=1}^{n} E_i$ and all mutually-exclusive events E_i are equally likely, meaning $P(E_1) = P(E_2) = ... = P(E_n)$, then

$$P(E_i) = \frac{1}{n}, \qquad i = 1 \text{ to } n$$

Some Simple Propositions, Section 2.4

Proposition 1: $P(E^c) = 1 - P(E)$

Proof: $S = E \cup E^c \implies 1 = P(S) = P(E) + P(E^c)$

Proposition 2: If $E \subset F$, then $P(E) \leq P(F)$

Proof:

$E \subset F \implies F = E \cup (E^c \cap F)$

$\implies P(F) = P(E) + P(E^c \cap F) \geq P(E)$

Proposition 3: $$P(E \cup F) = P(E) + P(F) - P(EF)$$

Proof:

$E \cup F = (E - EF) \cup (F - EF) \cup EF$, union of three mutually exclusive sets.

$$\begin{aligned}
\Longrightarrow P(E \cup F) &= P(E - EF) + P(F - EF) + P(EF) \\
&= P(E) - P(EF) + P(F) - P(EF) + P(EF) \\
&= P(E) + P(F) - P(EF)
\end{aligned}$$

Note that $E = EF \cup (E - EF)$, union of two mutually exclusive sets.

A Generalization of Proposition 3:

$$
\begin{aligned}
&P(E \cup F \cup G) \\
=&P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)
\end{aligned}
$$

Proof:

Let $B = F \cup G$. Then $P(E \cup B) = P(E) + P(B) - P(EB)$

$$P(B) = P(F \cup G) = P(F) + P(G) - P(FG)$$

$$P(EB) = P(E(F \cup G)) = P(EF \cup EG) =$$
$$P(EF) + P(EG) - P(EF\ EG) = P(EF) + P(EG) - P(EFG)$$

Proposition 4: $P(E_1 \cup E_2 \cup ... \cup E_n)$ = ?

Read this proposition and its proof yourself.

We will temporarily skip this proposition and relevant examples 5m, 5n, and 5o. However, we will come back to those material at a later time.

Example:

A total of 36 members of a club play tennis, 28 play squash, and 18 play badminton. Furthermore, 22 of the members play both tennis and squash, 12 play both tennis and badminton, 9 play both squash and badminton, and 4 play all three sports.

Q: How many members of this club play at least one of these sports?

Solution:

$A = \{$ members playing tennis $\}$

$B = \{$ members playing squash $\}$

$C = \{$ members playing badminton $\}$

$|A \cup B \cup C| =??$

Sample Space Having Equally Likely Outcomes, Sec. 2.5

Assumption: In an experiment, all outcomes in the sample space S are equally likely to occur.

For example, $S = \{1, 2, 3, ..., N\}$. Assuming equally likely outcomes, then

$$P(\{i\}) = \frac{1}{N}, \quad i = 1 \text{ to } N$$

For any event E:

$$P(E) = \frac{\text{Number of outcomes in } E}{\text{Number of outcomes in } S} = \frac{|E|}{|S|}$$

Example: If two dice are rolled, what is the probability that the sum of the upturned faces will equal 7?

$$S = \{(i,j), i,j = 1,2,3,4,5,6\}$$

$$E = \{??\}$$

Example: If 3 balls are randomly drawn from a bag containing 6 white and 5 black balls, what is the probability that one of the drawn balls is white and the other two black?

$$S = \{??\}$$

$$E = \{??\}$$

Example: An urn contains n balls, of which one is special. If k of these balls are withdrawn randomly, which is the probability that the special ball is chosen?

Revisit Probabilistic Counting of Set Intersections

Consider two sets $S_1, S_2 \subset \Omega = \{0, 1, 2, ..., D-1\}$. For example,

$$S_1 = \{0, 15, 31, 193, 261, 495, 563, 919\}$$

$$S_2 = \{5, 15, 19, 193, 209, 325, 563, 1029, 1921\}$$

One can first generate a **random permutation** π:

$$\pi: \quad \Omega = \{0, 1, 2, ..., D-1\} \longrightarrow \{0, 1, 2, ..., D-1\}$$

Apply π to sets S_1 and S_2. For example

$$\pi\left(S_1\right) = \{13, 139, 476, 597, 698, 1355, 1932, 2532\}$$

$$\pi\left(S_2\right) = \{13, 236, 476, 683, 798, 1456, 1968, 2532, 3983\}$$

Store only the minimums

$$\min\{\pi(S_1)\} = 13, \qquad \min\{\pi(S_2)\} = 13$$

Suppose we can theoretically calculate the probability

$$P = \mathbf{Pr}\left(\min\{\pi(S_1)\} = \min\{\pi(S_2)\}\right)$$

S_1 and S_2 are more similar $\Longrightarrow$ larger P (i.e., closer to 1).

Generate (small) k random permutations and every times store the minimums.

Count the number of matches to estimate the similarity between S_1 and S_2.

Now, we can easily (**How?**) show that

$$P = \mathbf{Pr}\left(\min\{\pi\left(S_1\right)\} = \min\{\pi\left(S_2\right)\}\right) = \frac{|S_1 \cap S_2|}{|S_1 \cup S_2|} = \frac{a}{f_1 + f_2 - a}$$

where $a = |S_1 \cap S_2|, \quad f_1 = |S_1|, \quad f_2 = |S_2|$.

Therefore, we can estimate P as a **binomial probability** from k permutations.

How many permutations (value of k) we need?:

Depend on the variance of this binomial distribution.

Poker Example (2.5.5f): A poker hand consists of 5 cards. If the cards have distinct consecutive values and are not all of the same suit, we say that the hand is a straight. What is the probability that one is dealt a straight?

A 2 3 4 5 6 7 8 9 10 J Q K A

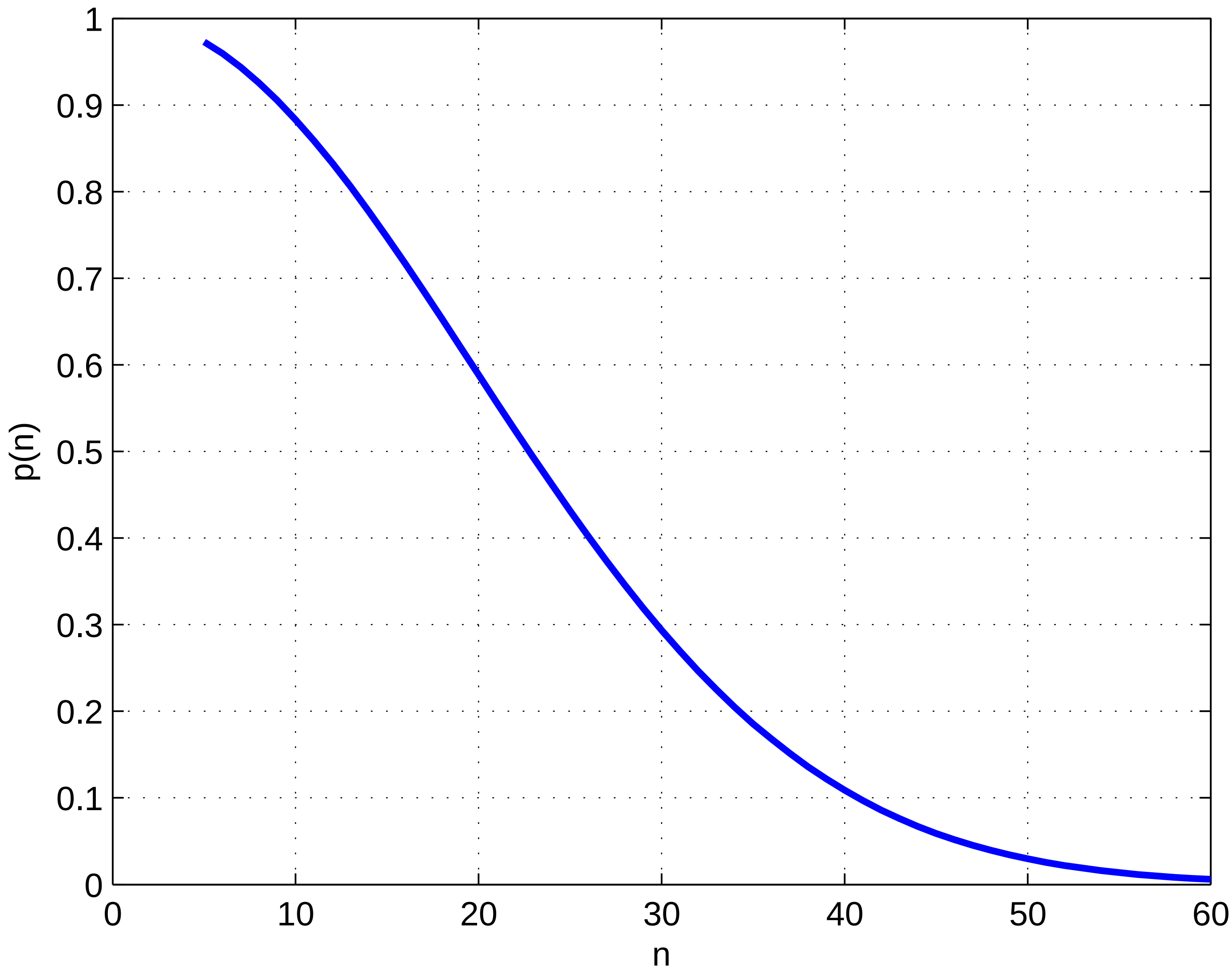
Example: A 5-card poker hand is a full house if it consists of 3 cards of the same denomination and 2 cards of the same denomination. (That is, a full house is three of a kind plus a pair.) What is the probability that one is dealt a full house?

Example (The Birthday Attack): If n people are present in a room, what is the probability that no two of them celebrate their birthday on the same day of the year (365 days) ? How large need n be so that this probability is less than $\frac{1}{2}$?

Example (The Birthday Attack): If n people are present in a room, what is the probability that no two of them celebrate their birthday on the same day of the year? How large need n be so that this probability is less than $\frac{1}{2}$?

$$p(n) = \frac{365 \times 364 \times ... \times (365 - n + 1)}{365^n}$$
$$= \frac{365}{365} \times \frac{364}{365} \times ... \times \frac{365 - n + 1}{365}$$

$p(23) = 0.4927 < \frac{1}{2}$, $\qquad p(50) = 0.0296$



p(n)
n
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60

Matlab Code

```
n = 5:60;
for i = 1:length(n)
    p(i) = 1;
    for j = 2:n(i);
        p(i) = p(i) * (365-j+1)/365;
    end;
end;
figure; plot(n,p,'linewidth',2); grid on; box on;
xlabel('n'); ylabel('p(n)');
```

How to improve the efficiency for computing many n's?

Example: A football team consists of 20 offensive and 20 defensive players. The players are to be paired into groups of 2 for the purpose of determining roommates. The pairing is done at random.

(a) What is the probability that there are no offensive-defensive roommate pairs?

(b) What is the probability that there are $2i$ offensive-defensive roommate pairs?

Denote the answer to (b) by P_{2i}. The answer to (a) is then P_0.

Solution Steps:

1. Total ways of dividing the 40 players into 20 unordered pairs.
2. Ways of pairing $2i$ offensive-defensive pairs.
3. Ways of pairing $20 - 2i$ remaining players within each group.
4. Applying multiplicative principle and equally likely outcome assumption.
5. Obtaining numerical probability values.

Solution Steps:

1. Total ways of dividing the 40 players into 20 unordered pairs.

ordered $$\binom{40}{2,2,...,2} = \frac{(40)!}{(2!)^{20}}$$

unordered $$\frac{(40)!}{(2!)^{20}(20)!}$$

2. Ways of pairing $2i$ offensive-defensive pairs.

$$\binom{20}{2i} \times \binom{20}{2i} \times (2i)!$$

3. Ways of pairing $20 - 2i$ remaining players within each group.

$$\left[\frac{(20-2i)!}{2^{10-i}(10-i)!}\right]^2$$

4. Applying multiplicative principle and equally likely outcome assumption.

$$P_{2i} = \frac{\left[\binom{20}{2i}\right]^2 (2i)! \left[\frac{(20-2i)!}{2^{10-i}(10-i)!}\right]^2}{\frac{40!}{2^{20}20!}}$$

5. Obtaining numerical probability values.

Stirling's approximation: $n! \approx n^{n+1/2}e^{-n}\sqrt{2\pi}$

Be careful!

- Better approximations are available.
- $n!$ could easily overflow, so could $\binom{n}{k}$.
- The probability never overflows, $P \in [0, 1]$.
- Looking for possible cancelations.
- Rearranging terms in numerator and denominator to avoid overflow.

Conditional Probability and Independence

Chapter 3, Sections 3.1 - 3.5

Introduction, Section 3.1

Conditional Probability: one of the most important concepts in probability:

- Calculating probabilities when partial (side) information is available.
- Computing desired probabilities may be easier.

A Famous Brain-Teaser

- There are 3 doors.
- Behind one door is a prize.
- Behind the other 2 is nothing.
- First, you pick one of the 3 doors at random.
- One of the doors that you did not pick will be opened to reveal nothing.
- Q: Do you stick with the door you chose in the first place, or do you switch?

A Joke: Tom & Jerry Conversation

- Jerry: I've heard that you will take a plane next week. Are you afraid that there might be terrist activities?
- Tom: Yes. That's exactly why I am prepared.
- Jerry: How?
- Tom: It is reported that the chance of having a bomb in a plane is $\frac{1}{10^6}$. According to what I learned in my probability class (4080), the chance of having two bombs in the plane will be $\frac{1}{10^{12}}$.
- Jerry: So?
- Tom: Therefore, to dramatically reduce the chance from $\frac{1}{10^6}$ to $\frac{1}{10^{12}}$, I will bring one bomb to the plane myself!
- Jerry: AH...

Section 3.2, Conditional Probabilities

Example: Suppose we toss a fair dice twice.

- What is the probability that the sum of the 2 dice is 8?

  Sample space $S = \{??\}$
  Event $E = \{??\}$

- What if we know the first dice is 3?

  (Reduced) $S = \{??\}$ (Reduced) $E = \{??\}$

The Basic Formula of Conditional Probability

Two events $E,\ F \subseteq S$.

$P(E|F)$: conditional probability that E occurs given that F occurs.

If $P(F) > 0$, then

$$P(E|F) = \frac{P(EF)}{P(F)}$$

The Basic Formula of Conditional Probability

$$P(E|F) = \frac{P(EF)}{P(F)}$$

How to understand this formula? —- Venn diagram

Given that F occurs,

- What is the new (ie, reduced) sample space S'?
- What is the new (ie, reduced) event E'?
- $P(E|F) = \frac{|E'|}{|S'|} = \frac{|E'|/S}{|S'|/S} = ??$

Example (3.2.2c):

In the card game bridge the 52 cards are dealt out equally to 4 players — called East, West, North, and South. If North and South have a total of 8 spades among them, what is the probability that East has 3 of the remaining 5 spades?

The reduced sample space $S' = \{??\}$

The reduced event $E' = \{??\}$

A transformation of the basic formula

$$P(E|F) = \frac{P(EF)}{P(F)}$$

$\Longrightarrow$

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

$P(E|F)$ may be easier to compute than $P(F|E)$, or vice versa.

Example:

A total of n balls are sequentially and randomly chosen, without replacement, from an urn containing r red and b blue balls ($n \leq r + b$).

Given that k of the n balls are blue, what is the probability that the first ball chosen is blue?

Two solutions:

1. Working with the reduced sample space — Easy!

   Basically a problem of choosing one ball from n balls randomly.

2. Working with the original sample space and basic formula — More difficult!

Working with original sample space:

Event B: the first ball chosen is blue.

Event B_k: a total of k blue balls are chosen.

Desired probability:

$$P(B|B_k) = \frac{P(BB_k)}{P(B_k)} = \frac{P(B_k|B)P(B)}{P(B_k)}$$

$P(B) =??$ $\quad P(B_k) =??$ $\quad P(B_k|B) =??$

Desired probability:

$$P(B|B_k) = \frac{P(BB_k)}{P(B_k)} = \frac{P(B_k|B)P(B)}{P(B_k)}$$

$$P(B) = \frac{b}{r+b}, \qquad P(B_k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{r+b}{n}}, \qquad P(B_k|B) = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}}$$

Therefore,

$$P(B|B_k) = \frac{P(B_k|B)P(B)}{P(B_k)} = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}} \times \frac{b}{r+b} \times \frac{\binom{r+b}{n}}{\binom{b}{k}\binom{r}{n-k}} = \frac{k}{n}$$

Example (3.2.2f):

Suppose an urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement. Assume that at each draw each ball in the urn is equally likely to be chosen.

Q: what is probability that both balls drawn are red?

Two solutions:

1. The direct approach
2. The conditional probability approach

Two solutions:

1. The direct approach

$$\frac{\binom{8}{2}}{\binom{12}{2}} = \frac{14}{33}$$

2. The conditional probability approach

$$\frac{8}{12} \times \frac{7}{11} = \frac{14}{33}$$

The Multiplication Rule

An extremely important formula in science

$$P\left(E_1E_2E_3...E_n\right) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)...P(E_n|E_1E_2...E_{n-1})$$

Proof?

Example (3.2.2g):

An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 cards each.

Q: Compute the probability that each pile has exactly 1 ace.

A: Two solutions:

1. Direct approach.
2. Conditional probability (as in the textbook).

Solution by the direct approach

- The size of sample space)

$$\binom{52}{13}\binom{39}{13}\binom{26}{13} = \frac{52!}{13! \times 13! \times 13! \times 13!} = \frac{52!}{(13!)^4}$$

- Fix 4 aces in four piles

- Only need to select 12 cards for each pile (from 48 cards total)

$$\binom{48}{12}\binom{36}{12}\binom{24}{12} = \frac{48!}{(12!)^4}$$

- Consider $4!$ permutations

- The desired probability

$$\frac{\binom{48}{12}\binom{36}{12}\binom{24}{12}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}} \times 4! = \frac{48!}{52!}\frac{(13!)^4}{(12!)^4}4! = 0.1055$$

Solution using the conditional probability

Define

- E_1 = $\{$The ace of spades is in any one of the piles $\}$
- E_2 = $\{$The ace of spades and the ace of hearts are in different piles $\}$
- E_3 = $\{$The ace of spades, hearts, and diamonds are all in different piles $\}$
- E_4 = $\{$All four aces are in different piles$\}$

The desired probability

$$
\begin{aligned}
&P\left(E_1E_2E_3E_4\right)\\
=&P\left(E_1\right)P\left(E_2|E_1\right)P\left(E_3|E_1E_2\right)P\left(E_2|E_1\right)P\left(E_4|E_1E_2E_3\right)
\end{aligned}
$$

- $P(E_1) = 1$
- $P(E_2|E_1) = \frac{39}{51}$
- $P(E_3|E_1E_2) = \frac{26}{50}$
- $P(E_4|E_1E_2E_3) = \frac{13}{49}$
- $P\left(E_1E_2E_3E_4\right) = \frac{39}{51}\frac{26}{50}\frac{13}{49} = 0.1055.$

Section 3.3, Baye's Formula

The basic Baye's formula

$$P(E) = P(E|F)P(F) + P(E|F^c)\left(1 - P(F)\right)$$

Proof:

- By Venn diagram. Or
- By algebra using the conditional probability formula

Example (3.3.3a):

An accident-prone person will have an accident within a fixed one-year period with probability 0.4. This probability decreases to 0.2 for a non-accident-prone person. Suppose 30% of the population is accident prone.

Q 1: What is the probability that a random person will have an accident within a fixed one-year period?

Q 2: Suppose a person has an accident within a year. What is the probability that he/she is accident prone?

Example:

Let p be the probability that the student knows the answer and $1-p$ the probability that the student guesses. Assume that a student who guesses will be correct with probability $\frac{1}{m}$, where m is the total number of multiple-choice alternatives.

Q: What is the conditional probability that a student knew the answer, given that he/she answered it correctly?

Solution:

Let C = event that the student answers the question correctly.

Let K = event that he/she actually knows the answer.

$P(K|C) = ??$

Solution:

Let C = event that the student answers the question correctly.

Let K = event that he/she actually knows the answer.

$$\begin{aligned}
P(K|C) = \frac{P(KC)}{P(C)} &= \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)} \\
&= \frac{1 \times p}{1 \times p + \frac{1}{m}(1-p)} \\
&= \frac{mp}{1 + (m-1)p}
\end{aligned}$$

What happens when m is large?

Example (3.3.3d):

A lab blood test is 95% effective in detecting a certain disease when it is represent. However, the test also yields a false positive result for 1% of the healthy tested. Assume 0.5% of the population actually has the disease.

Q: What is the probability that a person has the disease given that the test result is positive?

Solution:

Let D = event that the tested person has the disease.

Let E = event that the test result is positive.

$P(D|E)$ = ??

The Generalized Baye's Formula

Suppose F_1, F_2, ..., F_n are mutually exclusive events such as $\bigcup_{i=1}^n F_i = S$, then

$$P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i),$$

which is easy to understand from the Venn diagram.

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)}$$

Example (3.3.3m):

A bin contains 3 different types of disposable flashlights. The probability that a type 1 flashlight will give over 100 hours of use is 0.7, with the corresponding probabilities for type 2 and type 3 flashlights being 0.4 and 0.3 respectively. Suppose that 20% of the flashlights in the bin are type 1, and 30% are type 2, and 50% are type 3.

Q 1: What is the probability that a randomly chosen flashlight will give more than 100 hours of use?

Q 2: Given the flashlight lasted over 100 hours, what is the conditional probability that it was type j?

Q 1: What is the probability that a randomly chosen flashlight will give more than 100 hours of use?

$$0.7 \times 0.2 + 0.4 \times 0.3 + 0.3 \times 0.5 = 0.14 + 0.12 + 0.15 = 0.41$$

Q 2: Given the flashlight lasted over 100 hours, what is the conditional probability that it was type j?

Type 1: $\frac{0.14}{0.41} = 0.3415$
Type 2: $\frac{0.12}{0.41} = 0.3927$
Type 3: $\frac{0.15}{0.41} = 0.3659$

Independent Events

Definition:

Two events E and F are said to be independent if

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

Consequently, if E and F are independent, then

$$P(E|F) = P(E), \qquad\qquad P(F|E) = P(F).$$

Independent is different from mutually exclusive!

Check Independence by Verifying $P(EF) = P(E)P(F)$

Example: Two coins are flipped and all 4 outcomes are equally likely. Let E be the event that the first lands heads and F the event that the second lands tails.

Verify that E and F are independent.

Example: Suppose we toss 2 fair dice. Let E denote the event that the sum of two dice is 6 and F denote the event that the first die equals 4.

Verify that E and F are not independent.

Proposition 4.1:

If E and F are independent, then so are E and F^c.

Proof:

Given $P(EF) = P(E)P(F)$

Need to show $P(EF^c) = P(E)P(F^c)$.

Which formula contains both $P(EF)$ and $P(EF^c)$?

Independence of Three Events

Definition:

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

$$P(EFG) = P(E)P(F)P(G)$$

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

$$P(EG) = P(E)P(G)$$

$$P(FG) = P(F)P(G)$$

Independence of More Than Three Events

Definition:

The events, E_i, $i = 1$ to n, are said to be independent if, for every E_1, E_2, ..., E_r, $r \leq n$, of these events,

$$P(E_1E_2...E_r) = P(E_1)P(E_2)...P(E_r)$$

Example (3.4.4f):

An infinite sequence of independent trials is to be performed. Each trial results in a success with probability p and a failure with probability $1-p$.

Q: What is the probability that

1. at least one success occurs in the first n trials.
2. exactly k successes occur in the first n trials.
3. all trials result in successes?

Solution:

Q: What is the probability that

1. at least one success occurs in the first n trials.

   $$1-(1-p)^n$$

2. exactly k successes occur in the first n trials.

   $$\binom{n}{k}p^k(1-p)^{n-k}$$

   How is this formula related to the binomial theorem?

3. all trials result in successes?

   $p^n \rightarrow 0$ if $p < 1$.

Example (3.4.4h):

Independent trials, consisting of rolling a pair of fair dice, are performed. What is the probability that an outcome of 5 appears before an outcome of 7 when the outcome is the sum of the dice?

Solution 1:

Let E_n denote the event that no 5 or 7 appears on the first $n-1$ trials and a 5 appears on the nth trial, then the desired probability is

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

Why E_i and E_j are mutually exclusive?

$P(E_n) = P($ no 5 or 7 on the first $n-1$ trials AND a 5 on the nth trial)

By independence

$P(E_n) = P($ no 5 or 7 on the first $n-1$ trials)$\times P($ a 5 on the nth trial)

$P($ a 5 on the nth trial) = P(a 5 on any trial)

P(no 5 or 7 on the first $n-1$ trials) = $[P(\text{no 5 or 7 on any trial})]^{n-1}$

$P(\text{no 5 or 7 on any trial}) = 1 - P(\text{either a 5 OR a 7 on any trial})$

$P(\text{either a 5 OR a 7 on any trial}) = P(\text{a 5 on any trial}) + P(\text{a 7 on any trial})$

$P(E_n) = P($ no 5 or 7 on the first $n-1$ trials)$\times P($ a 5 on the nth trial)

$$P(E_n) = \left(1 - \frac{4}{36} - \frac{6}{36}\right)^{n-1} \frac{4}{36} = \frac{1}{9}\left(\frac{13}{18}\right)^{n-1}$$

The desired probability

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n) = \frac{1}{9}\frac{1}{1-\frac{13}{18}} = \frac{2}{5}.$$

Solution 2 (using a conditional argument):

Let E = the desired event.

Let F = the event that the first trial is a 5.

Let G = the event that the first trial is a 7.

Let H = the event that the first trial is neither a 5 nor 7.

$$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H)$$

A General Result:

If E and F are mutually exclusive events of an experiment, then when independent trials of this experiment are performed, E will occur before F with probability

$$\frac{P(E)}{P(E)+P(F)}$$

The Gambler's Ruin Problem (Example 3.4.4k)

Two gamblers, A and B, bet on the outcomes of successive flips of a coin. On each flip, if the coin comes up head, A collects 1 unit from B, whereas if it comes up tail, A pays 1 unit to B. They continue until one of them runs out of money.

If it is assumed that the successive flips of the coin are independent and each flip results in a head with probability p, what is the probability that A ends up with all the money if he starts with i units and B starts with $N - i$ units?

Intuition: What is this probability if the coin is fair (i.e., $p = 1/2$)?

Solution:

E = event that A ends with all the money when he starts with i.

H = event that first flip lands heads.

$$P_i = P(E) = P(E|H)P(H) + P(E|H^c)P(H^c)$$

$P_0 =??$

$P_N =??$

$P(E|H) =??$

$P(E|H^c) =??$

$$
\begin{aligned}
P_i = P(E) =& P(E|H)P(H) + P(E|H^c)P(H^c) \\
=& pP_{i+1} + (1-p)P_{i-1} = pP_{i+1} + qP_{i-1}
\end{aligned}
$$

$$
\Longrightarrow P_{i+1} = \frac{P_i - qP_{i-1}}{p} \Longrightarrow P_{i+1} - P_i = \frac{q}{p}\left(P_i - P_{i-1}\right)
$$

$$
\begin{aligned}
&P_2 - P_1 = \frac{q}{p}P_1 \\
&P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1 \\
&\ldots \\
&P_i - P_{i-1} = \left(\frac{q}{p}\right)^{i-1} P_1
\end{aligned}
$$

$$P_i = P_1 \left[1 + \left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + ... + \left(\frac{q}{p}\right)^{i-1}\right]$$

$$= P_1 \left[\frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)}\right] \quad \text{if } q \neq p$$

Because $P_N = 1$, we know

$$P_1 = \frac{1 - \left(\frac{q}{p}\right)}{1 - \left(\frac{q}{p}\right)^N} \quad \text{if } q \neq p$$

$$P_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N} \quad \text{if } q \neq p$$

(Do we really have to worry about $q = p = \frac{1}{2}$?)

The Matching Problem (Example 3.5.5d)

At a party n men take off their hats. The hats are then mixed up, and each man randomly selects one. We say that a match occurs if a man selects his own hat.

Q: What is the probability of

(a) no matches;

(b) exactly k matches?

Solution (a): using recursion and conditional probabilities.

1. Let E = event that no matches occurs. Denote $P_n = P(E)$.

   Let M = event that the first man selects his own hat.

2. $P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c)$

3. Express P_n as a function a P_{n-1}, P_{n-2}, etc.

4. Identify the boundary conditions, eg $P_1 =?$, $P_2 =?$.

5. Prove P_n, eg, by induction.

Solution (a): using recursion and conditional probabilities.

1. Let E = event that no matches occurs. Denote $P_n = P(E)$.

   Let M = event that the first man selects his own hat.

2. $P_n = P(E) = P(E|M)P(M) + P(E|M^c)P(M^c)$

   $$P(E|M) = 0, \qquad P(M^c) = \frac{n-1}{n}, \qquad P_n = \frac{n-1}{n}P(E|M^c)$$

3. Express P_n as a function a P_{n-1}, P_{n-2}, etc.

   $$P(E|M^c) = \frac{1}{n-1}P_{n-2} + P_{n-1}, \qquad P_n = \frac{n-1}{n}P_{n-1} + \frac{1}{n}P_{n-2}.$$

4. Identify the boundary conditions.

   $P_1 = 0$, $P_2 = 1/2$.

5. Prove P_n by induction.

   $P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - ... + \frac{(-1)^n}{n!}$.

   (What is a good approximation of P_n?)

Solution (b):

1. Fix a group of k men.

2. The probability that these k men select their own hats.

3. The probability that none of the remaining $n - k$ men select their own hats.

4. Multiplicative counting principle.

5. Adjustment for selecting k men.

Solution (b):

1. Fix a group of k men.
2. The probability that these k men select their own hats.
   $\frac{1}{n}\frac{1}{n-1}\frac{1}{n-2}\ldots\frac{1}{n-k+1} = \frac{(n-k)!}{n!}$
3. The probability that none of the remaining $n-k$ men select their own hats.
   P_{n-k}
4. Multiplicative counting principle.
   $\frac{(n-k)!}{n!} \times P_{n-k}$
5. Adjustment for selecting k men.

$$\frac{(n-k)!}{n!} P_{n-k} \binom{n}{k}$$

$P(.|F)$ Is A Probability

Conditional probabilities satisfy all the properties of ordinary probabilities.

Proposition 5.1

- $0 \leq P(E|F) \leq 1$
- $P(S|F) = 1$
- If E_i, $i = 1, 2, ...$, are mutually exclusive events, then

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

Conditionally Independent Events

Definition:

Two events, E_1 and E_2, are said to be conditionally independent given event F, if

$$P(E_1|E_2F) = P(E_1|F)$$

or, equivalently

$$P(E_1E_2|F) = P(E_1|F)P(E_2|F)$$

Laplace's Rule of Succession (Example 3.5.5e)

There are $k+1$ coins in a box. The ith coin will, when flipped, turn up heads with probability i/k, $i=0$, 1, ..., k. A coin is randomly selected from the box and is then repeatedly flipped.

Q: If the first n flip all result in heads, what is the conditional probability that the $(n+1)$st flip will do likewise?

Key: Conditioning on that the ith coin is selected, the outcomes will be conditionally independent.

Solution:

Let C_i = event that the ith coin is selected.

Let F_n = event the first n flips all result in heads.

Let H = event that the $(n+1)$st flip is a head.

The desired probability:

$$P(H|F_n) = \sum_{i=0}^{k} P(H|F_nC_i)P(C_i|F_n)$$

By conditional independence $P(H|F_nC_i) = ??$

Solve $P(C_i|F_n)$ using Baye's formula.

The desired probability:

$$P(H|F_n) = \sum_{i=0}^{k} P(H|F_nC_i)P(C_i|F_n)$$

By conditional independence $P(H|F_nC_i) = P(H|C_i) = \frac{i}{k}$

Using Baye's formula

$$P(C_i|F_n) = \frac{P(C_iF_n)}{P(F_n)} = \frac{P(F_n|C_i)P(C_i)}{\sum_{j=0}^{k} P(F_n|C_j)P(C_j)} = \frac{\left(\frac{i}{k}\right)^n \frac{1}{k+1}}{\sum_{j=0}^{k} \left(\frac{j}{k}\right)^n \frac{1}{k+1}}$$

The final answer:

$$P(H|F_n) = \frac{\sum_{j=0}^{k} \left(\frac{j}{k}\right)^{n+1}}{\sum_{j=0}^{k} \left(\frac{j}{k}\right)^n}$$

Chapter 4: (Discrete) Random Variables

Definition:

A random variable is a function from the sample space S to the real numbers.

A discrete random variable is a random variable that takes only on a finite (or at most a countably infinite) number of values.

Example (1a):

Suppose the experiment consists of tossing 3 fair coins. Let Y denote the number of heads appearing, then Y is a discrete random variable taking on one of the values 0, 1, 2, 3 with probabilities

$$P(Y=0) = P\{(T,T,T)\} = \frac{1}{8}$$

$$P(Y=1) = P\{(H,T,T),\ (T,H,T),\ (T,T,H)\} = \frac{3}{8}$$

$$P(Y=2) = P\{(T,H,H),\ (H,T,H),\ (H,H,T)\} = \frac{3}{8}$$

$$P(Y=3) = P\{(H,H,H)\} = \frac{1}{8}$$

Check

$$P\left(\bigcup_{i=0}^{3}\{Y=i\}\right) = \sum_{i=0}^{3} P\{Y=i\} = 1.$$

Example (1c)

Independent trials, consisting of the flipping of a coin having probability p of coming up heads, are continuously performed until either a head occurs or a total of n flips is made. If we let X denote the number of times the coin is flipped, then X is a random variable taking on one of the the values $1, 2, ..., n$.

$$
\begin{aligned}
&P(X=1) = P\{H\} = p \\
&P(X=2) = P\{(T,H)\} = (1-p)p \\
&P(X=3) = P\{(T,T,H)\} = (1-p)^2 p \\
&... \\
&P(X=n-1) = (1-p)^{n-2} p \\
&P(X=n) = (1-p)^{n-1}
\end{aligned}
$$

Check $\sum_{i=1}^{n} P(X=i) = 1??$

Example (1d)

3 balls are randomly chosen from an urn containing 3 white, 3 red, and 5 black balls. Suppose that we win $1 for each white ball selected and lose $1 for each red ball selected. Let X denote our total winnings from the experiments.

Q 1: What possible values can X take on?

Q 2: What are the respective probabilities?

Solution 1: What possible values can X take on?

$X \in \{3, 2, 1, 0, -1, -2, -3\}$.

Solution 2: What are the respective probabilities?

$$P(X = 0) = \frac{\binom{5}{3} + \binom{3}{1}\binom{3}{1}\binom{5}{1}}{\binom{11}{3}} = \frac{55}{165}$$

$$P(X = 3) = P(X = -3) = \frac{\binom{3}{3}}{\binom{11}{3}} = \frac{1}{165}$$

The Coupon Collection Problem (Example 1e)

There are N distinct types of coupons and each time one obtains a coupon equally likely from one of the N types, independent of prior selections.

A random variable T = number of coupons that needs to be collected until one obtains a complete set of at least one of each type.

Q: What is $P(T = n)$?

Solution:

1. Suffice to study $P(T > n)$
   because $P(T = n) = P(T > n) - P(T > n - 1)$

2. Define event A_j = **no** type j coupon is collected among first n.

3. $P(T > n) = P\left(\bigcup_{j=1}^{N} A_j\right)$

Then what?

$$P\left(\bigcup_{j=1}^{N} A_j\right) = \sum_{j} P(A_j) - \sum_{j_1<j_2} P(A_{j_1}A_{j_2})$$

$$+ \sum_{j_1<j_2<j_3} P(A_{j_1}A_{j_2}A_{j_3}) + ... + (-1)^{N+1}P(A_1A_2...A_N)$$

$$P(A_j) = \left(\frac{N-1}{N}\right)^n$$

$$P(A_{j_1}A_{j_2}) = \left(\frac{N-2}{N}\right)^n$$

$$\ldots$$

$$P(A_{j_1}A_{j_2}...A_{j_k}) = \left(\frac{N-k}{N}\right)^n$$

$$\ldots$$

$$P(A_{j_1}A_{j_2}...A_{j_{N-1}}) = \left(\frac{1}{N}\right)^n$$

$$P(A_1A_2...A_N) = 0$$

$$P(T>n) = P\left(\bigcup_{j=1}^{N} A_j\right) = \sum_j P(A_j) - \sum_{j_1<j_2} P(A_{j_1}A_{j_2})$$

$$+ \sum_{j_1<j_2<j_3} P(A_{j_1}A_{j_2}A_{j_3}) + ... + (-1)^{N+1}P(A_1A_2...A_N)$$

$$= N\left(\frac{N-1}{N}\right)^n - \binom{N}{2}\left(\frac{N-2}{N}\right)^n + \binom{N}{3}\left(\frac{N-3}{N}\right)^n$$

$$+ ... + (-1)^N \binom{N}{N-1}\left(\frac{1}{N}\right)^n$$

$$= \sum_{j=1}^{N-1}\binom{N}{j}\left(\frac{N-j}{N}\right)^n (-1)^{j+1}$$

Cumulative Distribution Function

For a random variable X, its cumulative distribution function F is defined as

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

$F(x)$ is a **non-decreasing** function:

If $x \leq y$, then $F(x) \leq F(y)$.

Section 4.2: Discrete Random Variables

Discrete random variable A random variable X that take on at most a countable number of possible values.

The probability mass function $p(a) = P\{X = a\}$

If X takes one of the values x_1, x_2, ..., then

$$p(x_i) \geq 0 \qquad \sum_{i=1}^{\infty} p(x_i) = 1$$

Example (4.2.2a): The probability mass function of a random variable X is given by $p(i) = \frac{c\lambda^i}{i!}$, $i = 0, 1, 2, ...$

Q: Find

- the relation between c and λ.
- $P\{X = 0\}$;
- $P\{X > 2\}$;

Solution:

- The relation between c and λ.

$$\sum_{i=0}^{\infty} p(i) = 1 \Longrightarrow c\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1 \Longrightarrow ce^{\lambda} = 1 \Longrightarrow c = e^{-\lambda}.$$

- $P\{X = 0\}$;

$$P\{X = 0\} = c = e^{-\lambda}.$$

- $P\{X > 2\}$;

$$\begin{aligned} P\{X > 2\} =& 1 - P\{X = 0\} - P\{X = 1\} - P\{X = 2\} \\ =& 1 - c - c\lambda - c\frac{\lambda^2}{2} \\ =& 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2 e^{-\lambda}}{2}. \end{aligned}$$

Section 4.3: Expected Value

Definition:

If X is a discrete random variable having a probability mass function $p(x)$, the expected value of X is defined by

$$E(X) = \sum_{x;p(x)>0} xp(x)$$

The expected value is a weighted average of possible values that X can take on.

Example Find $E(x)$ where X is the outcome when we roll a fair die.

$$E(X) = \sum_{i=1}^{6} i\frac{1}{6} = \frac{1}{6}\sum_{i=1}^{6} i = \frac{7}{2}$$

Example Let I be an indicator variable for event A, if

$$I = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A^c \text{ occurs} \end{cases}$$

$E(X) =??$

Example (3d) A school class of 120 students are driven in 3 buses to a symphony. There are 36 students in one of the buses, 40 in another, and 44 in the third bus. When the buses arrive, one of the 120 students is randomly chosen. Let X denote the number of students on the bus of that randomly chosen student, find $E(X)$.

Solution:

$X = \{??\}$

$P(X) = ??$

Example (4a)

$$P\{X=-1\}=0.2, \qquad P\{X=0\}=0.5, \qquad P\{X=1\}=0.2.$$

Find $E(X^2)$.

Solution: Let $Y=X^2$.

$$P\{Y=0\}=P\{X=0\}=0.5$$

$$\begin{aligned}P\{Y=1\}=&P\{X=1\ OR\ X=-1\}\\=&P\{X=1\}+P\{X=-1\}\\=&0.2+0.3=0.5\end{aligned}$$

Therefore

$$E(X^2)=E(Y)=0\times 0.5+1\times 0.5=0.5$$

Verify

$$
\begin{aligned}
&\sum_i P\{X = x_i\} x_i^2 \\
=&P\{X=-1\}(-1)^2 + P\{X=0\}(0)^2 + P\{X=1\}(1)^2 \\
=&0.5 \overset{??}{=} E(X^2)
\end{aligned}
$$

Section 4.4: Expectation of a Function of Random Variable

Proposition 4.1 If X is a discrete random variable that takes on one of the values x_i, $i \geq 1$, with respective probabilities $p(x_i)$, then for any real-valued function g

$$E\left(g(X)\right) = \sum_i g(x_i)p(x_i)$$

Proof: Group the i's with the same $g(x_i)$.

Proof:

Group the i's with the same $g(x_i)$.

$$
\begin{aligned}
\sum_i g(x_i)p(x_i) =& \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p(x_i)\\
=& \sum_j y_j \sum_{i:g(x_i)=y_j} p(x_i)\\
=& \sum_j y_j P\{g(X) = y_j\}\\
=& E(g(X)).
\end{aligned}
$$

Corollary 4.1 If a and b are constants, then

$$E(aX+b) = aE(X) + b.$$

Proof Use the definition of $E(X)$.

The nth moment

$$E(X^n) = \sum_{x;p(x)>0} x^n p(x)$$

Section 4.5: Variance

The expectation $E(X)$ measures the average of possible values of X.

The variance $Var(X)$ measures the variation, or spread of these values.

Definition If X is a random variable with mean μ, then the variance of X is defined by

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

An alternative formula for $Var(X)$

$$Var(X) = E(X^2) - (E(X))^2 = E(X^2) - \mu^2$$

Proof:

$$\begin{aligned} Var(X) =& E\left[(X-\mu)^2\right] \\ =& \sum_x (x-\mu)^2 p(x) \\ =& \sum_x (x^2 - 2\mu x + \mu^2) p(x) \\ =& \sum_x x^2 p(x) - 2\mu \sum_x x p(x) + \mu^2 \sum_x p(x) \\ =& E(X^2) - 2\mu^2 + \mu^2 \\ =& E(X^2) - \mu^2. \end{aligned}$$

Two important identities

If a and b are constants, then

$$Var(a) = Var(b) = 0$$

$$Var(aX + b) = a^2 Var(X)$$

Proof: Use the definitions.

$$E(aX+b) = aE(x) + b$$

$$\begin{aligned}
Var(aX+b) =& E\left((aX+b) - E(aX+b)\right)^2 \\
=& E\left((aX+b) - aEX - b\right)^2 \\
=& E\left(aX - aEX\right)^2 \\
=& a^2 E\left(X - EX\right)^2 \\
=& a^2 Var(X)
\end{aligned}$$

Section 4.6: Bernoulli and Binomial Random Variables

Bernoulli random variable $X \sim Bernoulli(p)$

$$P\{X=1\} = p(1) = p,$$

$$P\{X=0\} = p(0) = 1-p.$$

Properties:

$$E(X) = p$$

$$Var(X) = p(1-p)$$

Binomial random variable

Definition A random variable X presents the number of total successes that occur in n independent trials. Each trial results in a success with probability p and a failure with probability $1-p$.

$$X \sim Binomial(n,p).$$

The probability mass function

$$p(i) = P\{X=i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

- Proof?
- $\sum_{i=0}^{n} p(i) = 1$?

A very useful representation:

$X \sim binomial(n, p)$ can be viewed as the sum of n independent Bernoulli random variables with parameter p.

$$X = \sum_{i=1}^{n} Y_i, \qquad\qquad Y_i \sim Bernoulli(p)$$

Example: Five fair coins are flipped. If the outcomes are assumed independent, find the probability mass function of the numbers of heads obtained.

Solution:

- X = number of heads obtained.
- $X \sim binomial(n, p)$. n=??, p=??
- $P\{X = i\} =$?

Properties of $X \sim binomial(n, p)$

Expectation: $E(X) = np$.

Variance: $Var(X) = np(1 - p)$.

Proof:??

$$
\begin{aligned}
E(X) &= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=0}^{n} i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} i \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= np \sum_{i=1}^{n} \frac{(n-1)!}{(i-1)!((n-1)-(i-1))!} p^{i-1} (1-p)^{(n-1)-(i-1)} \\
&= np \sum_{j=0}^{n-1} \frac{(n-1)!}{j!((n-1)-j)!} p^j (1-p)^{(n-1)-j} \\
&= np
\end{aligned}
$$

$$E(X^2) = \sum_{i=0}^{n} i^2 \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^{n} i^2 \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

$$= np \sum_{i=1}^{n} i \frac{(n-1)!}{(i-1)!((n-1)-(i-1))!} p^{i-1} (1-p)^{(n-1)-(i-1)}$$

$$= np \sum_{j=0}^{n-1} (j+1) \frac{(n-1)!}{j!((n-1)-j)!} p^j (1-p)^{(n-1)-j}$$

$$= npE(Y+1) \qquad Y \sim binomial(n-1, p)$$

$$= np((n-1)p+1)$$

$$Var(X) = E(X^2) - (E(X))^2 = np((n-1)p+1) - (np)^2 = np(1-p)$$

What Happens When $n \to \infty$ and $np \to \lambda$?

$X \sim binomial(n, p)$, $\quad n \to \infty$ and $np \to \lambda$

$$
\begin{aligned}
P\{X = i\} =& \frac{n!}{(n-i)!i!} p^i (1-p)^{n-i} \\
=& \frac{(1-p)^n}{i!} \frac{p^i}{(1-p)^i} n(n-1)(n-2)...(n-i+1) \\
=& \left[\frac{(1-p)^n (np)^i}{i!}\right] \left[\frac{n(n-1)(n-2)...(n-i+1)}{(1-p)^i n^i}\right] \\
\approx& \frac{e^{-\lambda}\lambda^i}{i!}
\end{aligned}
$$

$$\lim_{n\rightarrow\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

For fixed (small) i,

$$\lim_{n\rightarrow\infty}\frac{n(n-1)(n-2)...(n-i+1)}{(n-\lambda)^i} = 1$$

Section 4.7: The Poisson Random Variable

Definition: A random variable X, taking on one of the values 0, 1, 2, ..., is a Poisson random variable with parameter λ, if for some $\lambda > 0$,

$$p(i) = P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \qquad i = 0, 1, 2, 3, ...$$

Check

$$\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{\infty} e^{-\lambda}\frac{\lambda^i}{i!} = 1??$$

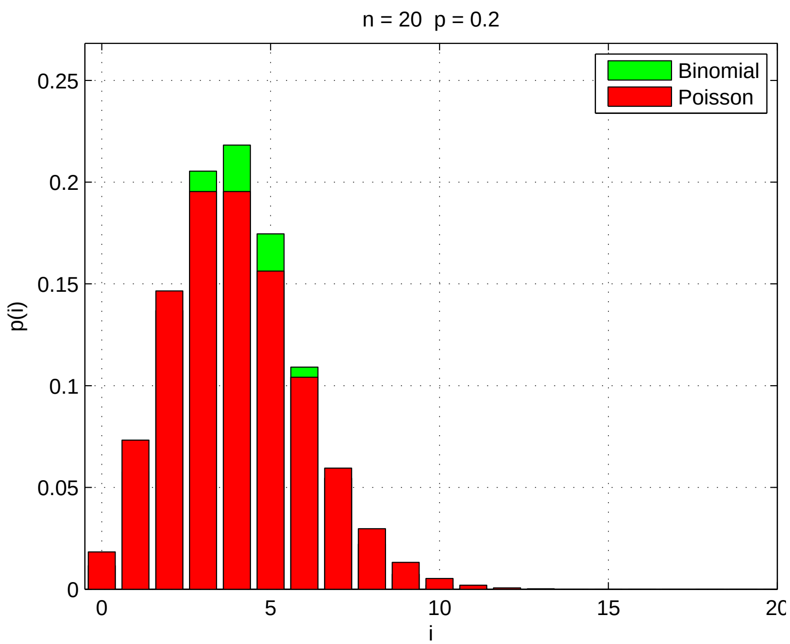
Facts about Poisson Distribution

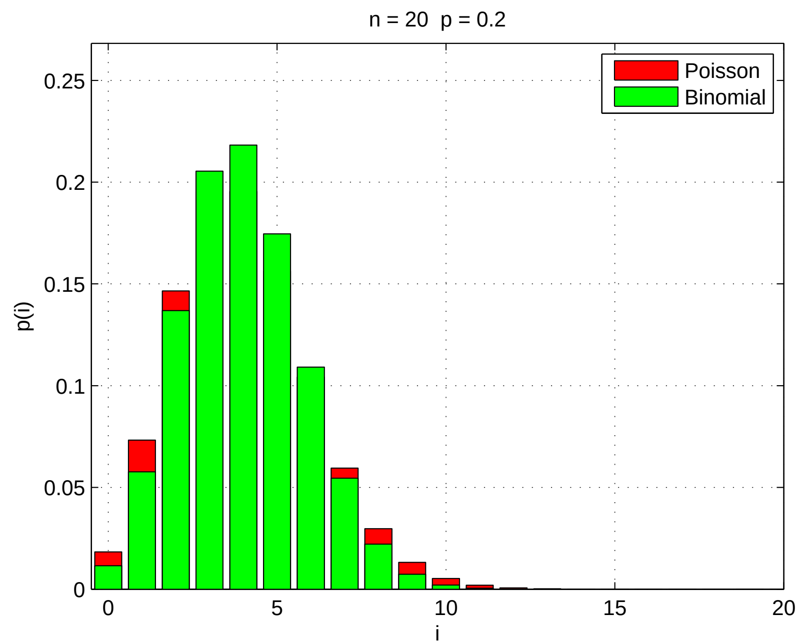
- Poisson(λ) approximates binomial (n, p), when n large and $np \approx \lambda$.

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

- Poisson is widely used for modeling rare ($p \approx 0$) events (but not too rare).
  - The number of misprints on a page of a book
  - The number of people in a community living to 100 years of age
  - ...

Compare Binomial with Poisson





Poisson approximation is good

- n does not have to be too large.
- p does not have to be too small.

Matlab Code

```
function comp_binom_poisson(n,p);

i=0:n; P1=binopdf(i,n,p); P2=poisspdf(i,n*p);
figure; bar(i,P1,'g'); hold on; grid on;
bar(i,P2,'r'); legend('Binomial','Poisson');
xlabel('i'); ylabel('p(i)');
axis([-0.5 n 0 max(P1)+0.05]);
title(['n = ' num2str(n) '  p = ' num2str(p)]);

figure; bar(i,P2,'r'); hold on; grid on;
bar(i,P1,'g'); legend('Poisson','Binomial');
xlabel('i'); ylabel('p(i)');
axis([-0.5 n 0 max(P1)+0.05]);
title(['n = ' num2str(n) '  p = ' num2str(p)]);
```

Example Suppose that the number of typographical errors on a single page of a book has a Poisson distribution with parameter $\lambda = \frac{1}{2}$.

Q: Calculate the probability that there is at least one error on this page.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/2} = 0.393.$$

Expectation and Variance of Poisson

$X \sim Poisson(\lambda)$ $\qquad$ $Y \sim binomial(n, p)$

X approximates Y when (1): n large, (2): p small, and (3): $np \approx \lambda$

$$E(Y) = np, \qquad Var(Y) = np(1-p)$$

Guess $E(X)$ = ??, $Var(X)$ =??

$X \sim Poisson(\lambda)$

$$E(X) = \sum_{i=1}^{\infty} i e^{-\lambda}\frac{\lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!} = \lambda \sum_{j=0}^{\infty} e^{-\lambda}\frac{\lambda^j}{j!} = \lambda$$

$$E(X^2) = \sum_{i=1}^{\infty} i^2 e^{-\lambda}\frac{\lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} i e^{-\lambda}\frac{\lambda^{i-1}}{(i-1)!}$$

$$= \lambda \sum_{j=0}^{\infty} (j+1) e^{-\lambda}\frac{\lambda^j}{j!} = \lambda\left(E(X)+1\right) = \lambda^2 + \lambda$$

$Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$

Section 4.8.1: Geometric Random Variable

Definition: Suppose that independent trials, each having probability p of being a success, are performed until a success occurs. X is the number of trials required. Then

$$P\{X = n\} = (1-p)^{n-1}p, \qquad n = 1, 2, 3, ...$$

Check $\sum_{n=1}^{\infty} P\{X=n\} = \sum_{n=1}^{\infty}(1-p)^{n-1}p = 1??$

Expectation $X \sim Geometric(p)$.

$$E(X) = \sum_{n=1}^{\infty} n(1-p)^{n-1}p \qquad\qquad (q = 1-p)$$

$$= p + 2qp + 3q^2p + 4q^3p + 5q^4p + ...$$

$$= p\left(1 + 2q + 3q^2 + 4q^3 + 5q^4 + ...\right) = ??$$

We know

$$1 + q + q^2 + q^3 + q^4 + ... = \frac{1}{1-q}$$

How about

$$I = 1 + 2q + 3q^2 + 4q^3 + 5q^4 + ... = ??$$

$$I = 1 + 2q + 3q^2 + 4q^3 + 5q^4 + ... =??$$

First trick

$$q \times I = q + 2q^2 + 3q^3 + 4q^4 + 5q^5 + ...$$

$$I - qI = 1 + q + q^2 + q^3 + q^4 + ... = \frac{1}{1-q}$$

$$\Longrightarrow I = \frac{1}{(1-q)^2}$$

$$\Longrightarrow E(X) = pI = \frac{p}{(1-q)^2} = \frac{1}{p}$$

$$I = 1 + 2q + 3q^2 + 4q^3 + 5q^4 + ... = ??$$

Second trick

$$I = \frac{d}{dq}\left(q + q^2 + q^3 + q^4 + q^5 + ...\right)$$

$$I = \frac{d}{dq}\frac{1}{1-q} = \frac{1}{(1-q)^2}$$

Variance $X \sim Geometric(p)$.

$$E(X^2) = \sum_{n=1}^{\infty} n^2(1-p)^{n-1}p \qquad (q = 1-p)$$
$$= p\left(1 + 2^2q + 3^2q^2 + 4^2q^3 + 5^2q^4 + ...\right)$$

$$J = 1 + 2^2q + 3^2q^2 + 4^2q^3 + 5^2q^4 + ... =??$$

$$J = 1 + 2^2 q + 3^2 q^2 + 4^2 q^3 + 5^2 q^4 + ... =??$$

Trick 1

$$qJ = q + 2^2 q^2 + 3^2 q^3 + 4^2 q^4 + 5^2 q^5 + ...$$

$$J(1-q) = 1 + (2^2 - 1^2)q + (3^2 - 2^2)q^2 + (4^2 - 3^2)q^3 + ...$$

$$= 1 + 3q + 5q^2 + 7q^3 + ...$$

$$J(1-q)q = q + 3q^2 + 5q^3 + 7q^4 + ...$$

$$J(1-q)^2 = 1 + 2q + 2q^2 + 2q^3 + ... = -1 + \frac{2}{1-q} = \frac{1+q}{1-q}$$

$$J = \frac{1+q}{(1-q)^3} \qquad E(X^2) = \frac{p(2-p)}{p^3} = \frac{2-p}{p^2} \qquad Var(X) = \frac{1-p}{p^2}$$

$$J = 1 + 2^2 q + 3^2 q^2 + 4^2 q^3 + 5^2 q^4 + ... = ??$$

Trick 2

$$\begin{aligned}
J = & \sum_{n=1}^{\infty} n^2 q^{n-1} = \sum_{n=1}^{\infty} \frac{d}{dq} n q^n \\
= & \frac{d}{dq}\left[\sum_{n=1}^{\infty} n q^n\right] = \frac{d}{dq}\left[\frac{q}{p}E(X)\right] \\
= & \frac{q}{p}\frac{d}{dq}\frac{1}{1-q} = \frac{q}{p}\frac{1}{(1-q)^2}
\end{aligned}$$

Section 4.8.2: Negative Binomial Distribution

Definition: Independent trials, each having probability p, of being a success are performed until a total of r successes is accumulated. Let X denote the number of trials required. Then X is a negative binomial random variable

$$X \sim NB(r, p)$$

$X \sim Geometric(p)$ is the same as $X \sim NB(1, p)$.

What is probability mass function of $X \sim NB(r, p)$?

$$P\{X = n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \qquad n = r, r+1, r+2, ...$$

$$E(X^k) = \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

$$= \sum_{n=r}^{\infty} n^k \frac{(n-1)!}{(r-1)!(n-r)!} p^r (1-p)^{n-r}$$

Now what?

$$
\begin{aligned}
E(X^k) =& \sum_{n=r}^{\infty} n^{k-1} \frac{n!}{(r-1)!(n-r)!} p^r (1-p)^{n-r} \\
=& \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \frac{n!}{r!(n-r)!} p^{r+1} (1-p)^{n-r}, \qquad (m = n+1) \\
=& \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \frac{(m-1)!}{r!(m-1-r)!} p^{r+1} (1-p)^{m-1-r} \\
=& \frac{r}{p} E[(Y-1)^{k-1}], \qquad\qquad Y \sim NB(r+1, p)
\end{aligned}
$$

$E(X) =??$, $E(X^2) =??$, $Var(X) =??$

$$E(X^k) = \frac{r}{p}E[(Y-1)^{k-1}], \qquad Y \sim NB(r+1,p)$$

Let $k=1$. Then

$$E(X) = \frac{r}{p}E[(Y-1)^0] = \frac{r}{p}$$

Let $k=2$. Then

$$E(X^2) = \frac{r}{p}E[(Y-1)] = \frac{r}{p}\left(\frac{r+1}{p} - 1\right)$$

$$Var(X) = E(X^2) - E^2X = \frac{r(1-p)}{p^2}$$

Example (4.8.8a) An urn contains N white and M black balls. Balls are randomly selected, one at a time, until a black one is obtained. Each selected ball is replaced before the next one is drawn.

Q(a): What is the probability that exactly n draws are needed?

Q(b): What is the expected number of draws needed?

Q(c): What is the probability that at least k draws are needed?

Solution: This is a geometric random variable with $p = \frac{M}{M+N}$.

Q(a): What is the probability that exactly n draws are needed?

$$P(X=n) = (1-p)^{n-1}p = \left(\frac{N}{M+N}\right)^{n-1}\frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}$$

Q(b): What is the expected number of draws needed?

$$E(X) = \frac{1}{p} = \frac{M+N}{M} = 1 + \frac{N}{M}$$

Q(c): What is the probability that at least k draws are needed?

$$P(X \geq k) = \sum_{n=k}^{\infty}(1-p)^{n-1}p = \frac{p(1-p)^{k-1}}{1-(1-p)} = \left(\frac{N}{M+N}\right)^{k-1}$$

Example (4.8.8e): A pipe-smoking mathematician carries 2 matchboxes, 1 in the left-hand pocket and 1 in his right-hand pocket. Each time he needs a match he is equal likely to take it from either pocket. Consider the moment when the mathematician first discovers that one of his matchboxes is empty. If it is assumed that both matchboxes initially contained N matches, what is the probability that there are exactly k matches in the other box, $k = 0, 1, ..., N$?

Solution:

1. Assume the left-hand pocket is empty.

2. X = total matches taken at the moment when left-hand pocket is empty.

3. View X as a random variable $X \sim NB(r, p)$. What is r, p?

4. Remember there are k matches remaining in the right-hand pocket.

5. The desired probability = $P\{X = n\}$. What is n?

6. Adjustment by considering either left-hand or right-hand can be empty.

Solution:

1. Assume the left-hand pocket is empty.
2. X = total matches taken at the moment when left-hand pocket is empty.
3. View X as a random variable $X \sim NB(r, p)$. What is r, p?

   $$p = \frac{1}{2}, r = N + 1.$$

4. Remember there are k matches remaining in the right-hand pocket.
5. The desired probability = $P\{X = n\}$. What is n?

   $$n = 2N - k + 1.$$

6. Adjustment by considering either left-hand or right-hand can be empty.

   $$2\binom{2N-k}{N}\left(\frac{1}{2}\right)^{2N-k+1}.$$

Section 4.8.3: Hypergeometric Distribution

Definition A sample of size n is chosen randomly without replacement from an urn containing N balls, of which m are white and $N-m$ are black. Let X denote the number of white balls selected, then $X \sim HG(N,m,n)$

$$P\{X=i\} = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}, \qquad i = 0, 1, ..., n.$$

$$n-(N-m) \leq i \leq \min(n,m)$$

If the n balls are chosen with replacement, then it is $binomial\left(n, p = \frac{m}{N}\right)$.

when m and N are large relative to n.

$$X \sim HG(N, m, n) \approx binomial\left(n, \frac{m}{N}\right)$$

$$E(X) \overset{??}{\approx} np = \frac{nm}{N} \;??$$

$$Var(X) \overset{??}{\approx} np(1-p) = \frac{nm}{N}\left(1 - \frac{m}{N}\right) \;??$$

$$
\begin{aligned}
E(X^k) &= \sum_{i=1}^{n} i^k \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \\
&= \sum_{i=1}^{n} i^{k-1} \frac{m!}{(i-1)!(m-i)!} \frac{(N-m)!}{(n-i)!(N-m-n-i)!} \frac{n!(N-n)!}{N!} \\
&= \sum_{j=0}^{n-1} (j+1)^{k-1} \frac{m!}{j!(m-1-j)!} \frac{(N-m)!}{(n-1-j)!(N-m-n-1-j)!} \frac{n!(N-n)!}{N!} \\
&= \frac{nm}{N} E\left[(Y+1)^{k-1}\right], \qquad\qquad Y \sim HG(N-1, m-1, n-1)
\end{aligned}
$$

Section 4.9: Properties of Cumulative Distribution Function

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

1. $F(x)$ is a non-decreasing function: If $a < b$, then $F(a) \leq F(b)$
2. $\lim_{x\rightarrow\infty} = 1$
3. $\lim_{x\rightarrow-\infty} = 0$
4. $F(x)$ is right continuous.

Chapter 5: Continuous Random Variables

Definition: X is a continuous random variable if there exists a non-negative function f, defined for all real $x \in (-\infty,\ \infty)$, having the property that for any set B of real numbers,

$$P\{X \in B\} = \int_B f(x)dx$$

The function f is called the probability density function of the random variable X.

Some Facts

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

- $P\{X \in (a, b)\} = P\{a \leq X \leq b\} = \int_a^b f(x)dx$

- $P\{X = a\} = \int_a^a f(x)dx = 0$

- $F(a) = P\{X \leq a\} = P\{X < a\} = \int_{-\infty}^{a} f(x)dx$

- $F(\infty) = 1, \qquad\qquad F(-\infty) = 0$

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

Example 5.1.1a: Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Q(a): What is the value of C?

Q(b): Find $P\{X > 1\}$.

Example 5.1.1d: Denote F_X and f_X the cumulative function and the density function of a continuous random variable X, respectively. Find the density function of $Y = 2X$.

Example 5.1.1d: Denote F_X and f_X the cumulative function and the density function of a continuous random variable X, respectively. Find the density function of $Y = 2X$.

Solution:

$$F_Y(y) = P\{Y \leq y\} = P\{2X \leq y\} = P\{X \leq y/2\} = F_X(y/2)$$

$$f_Y(x) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(y/2) = \frac{1}{2}f_X(y/2)$$

Derivatives of Common Functions

http://en.wikipedia.org/wiki/Derivative

- Exponential and logarithmic functions

$\frac{d}{dx}e^x = e^x$, $\frac{d}{dx}a^x = a^x \log a$, (Notation: $\log x = \log_e x$)

$\frac{d}{dx}\log x = \frac{1}{x}$, $\frac{d}{dx}\log_a x = \frac{1}{x \log a}$

- Trigonometric functions

$\frac{d}{dx}\sin x = \cos x$, $\frac{d}{dx}\cos x = -\sin x$, $\frac{d}{dx}\tan x = \sec^2 x$

- Inverse trigonometric functions

$\frac{d}{dx}\arcsin x = \frac{1}{\sqrt{1-x^2}}$, $\frac{d}{dx}\arccos x = -\frac{1}{\sqrt{1-x^2}}$,

$\frac{d}{dx}\arctan x = \frac{1}{1+x^2}$,

Integration by Parts

$$\int_a^b f(x)dx = f(x)x|_a^b - \int_a^b xf'(x)dx$$

$$\int_a^b f(x)d[g(x)] = f(x)g(x)|_a^b - \int_a^b g(x)f'(x)dx$$

$$\int_a^b f(x)h(x)dx = \int_a^b f(x)d[H(x)] = f(x)H(x)|_a^b - \int_a^b H(x)f'(x)dx$$

where $h(x) = H'(x)$.

Examples:

$$\begin{aligned}\int_a^b \log x dx =& x\log x|_a^b - \int_a^b x\left[\log x\right]' dx\\ =& b\log b - a\log a - \int_a^b x\frac{1}{x}dx\\ =& b\log b - a\log a - b + a.\end{aligned}$$

$$\begin{aligned}\int_a^b x\log x dx =& \frac{1}{2}\int_a^b \log x d[x^2] = \frac{1}{2}x^2\log x|_a^b - \frac{1}{2}\int_a^b x^2\left[\log x\right]' dx\\ =& \frac{1}{2}b^2\log b - \frac{1}{2}a^2\log a - \frac{1}{2}\int_a^b x dx\\ =& \frac{1}{2}b^2\log b - \frac{1}{2}a^2\log a - \frac{1}{4}\left(b^2 - a^2\right)\end{aligned}$$

$$\int_a^b x^n e^x dx = \int_a^b x^n d[e^x] = x^n e^x|_a^b - n\int_a^b e^x x^{n-1} dx$$

$$\int_a^b x^n \cos x dx = \int_a^b x^n d[\sin x] = x^n \sin x|_a^b - n\int_a^b \sin x\, x^{n-1} dx$$

$$\begin{aligned}\int_a^b \sin^n x \sin 2x dx =& 2\int_a^b \sin^n x \sin x \cos x dx = 2\int_a^b \sin^{n+1} x d[\sin x]\\ =& \frac{2}{n+2}\sin^{n+2} x\Big|_a^b\end{aligned}$$

Expectation

Definition: Given a continuous random variable X with density $f(x)$, its density is defined as

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Example: The density of X is given by

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find $E\left[e^X\right]$.

Two solutions:

1. • Find the density function of $Y = e^X$, denoted by $f_Y(y)$.
   • $E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy$.

2. Apply the formula

$$E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Solution 1: $Y = e^X$.

$$F_Y(y) = P\{Y \leq y\} = P\left\{e^X \leq y\right\}$$
$$= P\{X \leq \log y\} = \int_0^{\log y} 2y dy = \log^2 y$$

Therefore,

$$f_Y(y) = \frac{2}{y}\log y, \qquad 1 \leq y \leq e$$
$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_1^e 2\log y dy$$
$$= 2y\log y|_1^e - \int_1^e 2y\frac{1}{y}dy = 2$$

Solution 2:

$$
\begin{aligned}
E\left[e^X\right] &= \int_0^1 2xe^x dx = \int_0^1 2x de^x \\
&= 2xe^x\big|_0^1 - 2\int_0^1 e^x dx \\
&= 2e - 2(e-1) = 2
\end{aligned}
$$

Lemma 5.2.2.1 For a non-negative random variable X,

$$E(X) = \int_0^\infty P\{X > x\}dx$$

Lemma 5.2.2.1 For a non-negative random variable X,

$$E(X) = \int_0^\infty P\{X > x\}dx$$

Proof:

$$\begin{aligned}\int_0^\infty P\{X > x\}dx &= \int_0^\infty \int_x^\infty f_X(y)dydx \\ &= \int_0^\infty \int_0^y f_X(y)dxdy \\ &= \int_0^\infty f_X(y)\left(\int_0^y dx\right)dy \\ &= \int_0^\infty f_X(y)ydy = E(X)\end{aligned}$$

X is a continuous random variable.

- $E(aX + b) = aE(X) + b$, where a and b are constants.

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

- $Var(aX + b) = a^2 Var(X)$

Section 5.3: The Uniform Random Variable

Definition: X is a uniform random variable on the interval (a,b) if its probability density function is given by

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

Denote $X \sim Uniform(a,b)$, or simply $X \sim U(a,b)$.

Cumulative function

$$F(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a} & \text{if } a < x < b \\ 1 & \text{if } x \geq b \end{cases}$$

If $X \sim U(0,1)$, then $F(x) = x$, if $0 \leq x \leq 1$.

Expectation

$$E(X) = \frac{a+b}{2}$$

Variance

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

Section 5.4: Normal Random Variable

Definition: X is a normal random variable, $X \sim N\left(\mu, \sigma^2\right)$ if its density is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad -\infty < x < \infty$$

- If $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu$ and $Var(X) = \sigma^2$.
- The density function is a "bell-shaped" curve.
- It was first introduced by French Mathematician Abraham DeMoivre in 1733, to approximate the binomial when n was large.
- The distribution became truly famous because of Karl F. Gauss and hence its another common name is "Gaussian distribution."
- It is called "normal distribution" (probably started by Karl Pearson) because "most people" believed that it was "normal" for any well-behaved data set to follow this distribution.

Check that

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

Note that, by letting $\boxed{z = \frac{x-\mu}{\sigma}}$,

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = I.$$

It suffices to show that $I^2 = 1$.

$$\begin{aligned}
I^2 =& \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw \\
=& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{z^2+w^2}{2}} dzdw \\
=& \frac{1}{2\pi} \int_{0}^{\infty} \int_{0}^{2\pi} e^{-\frac{r^2}{2}} rd\theta dr = \frac{1}{2\pi} \int_{0}^{2\pi} d\theta \int_{0}^{\infty} e^{-\frac{r^2}{2}} d\left[\frac{r^2}{2}\right] = 1.
\end{aligned}$$

Expectation and Variance

If $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu$ and $Var(X) = \sigma^2$.

$$
\begin{aligned}
E(X) =& \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
=& \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \int_{-\infty}^{\infty} \mu \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
=& \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d[x-\mu] + \mu \\
=& \int_{-\infty}^{\infty} z \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} dz + \mu \\
=& \mu.
\end{aligned}
$$

$$
\begin{aligned}
Var(X^2) =& E\left[(X-\mu)^2\right] \\
=& \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d[x-\mu] \\
=& \int_{-\infty}^{\infty} z^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} dz \\
=& \sigma^2 \int_{-\infty}^{\infty} \frac{z^2}{\sigma^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}} d\left[\frac{z}{\sigma}\right] \\
=& \sigma^2 \int_{-\infty}^{\infty} t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} d\left[t\right] \\
=& \sigma^2 \int_{-\infty}^{\infty} t \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} d\left[\frac{t^2}{2}\right] \\
=& \sigma^2 \int_{-\infty}^{\infty} -t \frac{1}{\sqrt{2\pi}} d\left[e^{-\frac{t^2}{2}}\right] \\
=& \sigma^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \sigma^2 t e^{-\frac{t^2}{2}} \Big|_{-\infty}^{\infty} = \sigma^2
\end{aligned}
$$

Why $\lim_{t\to\infty} te^{-\frac{t^2}{2}} = 0??$

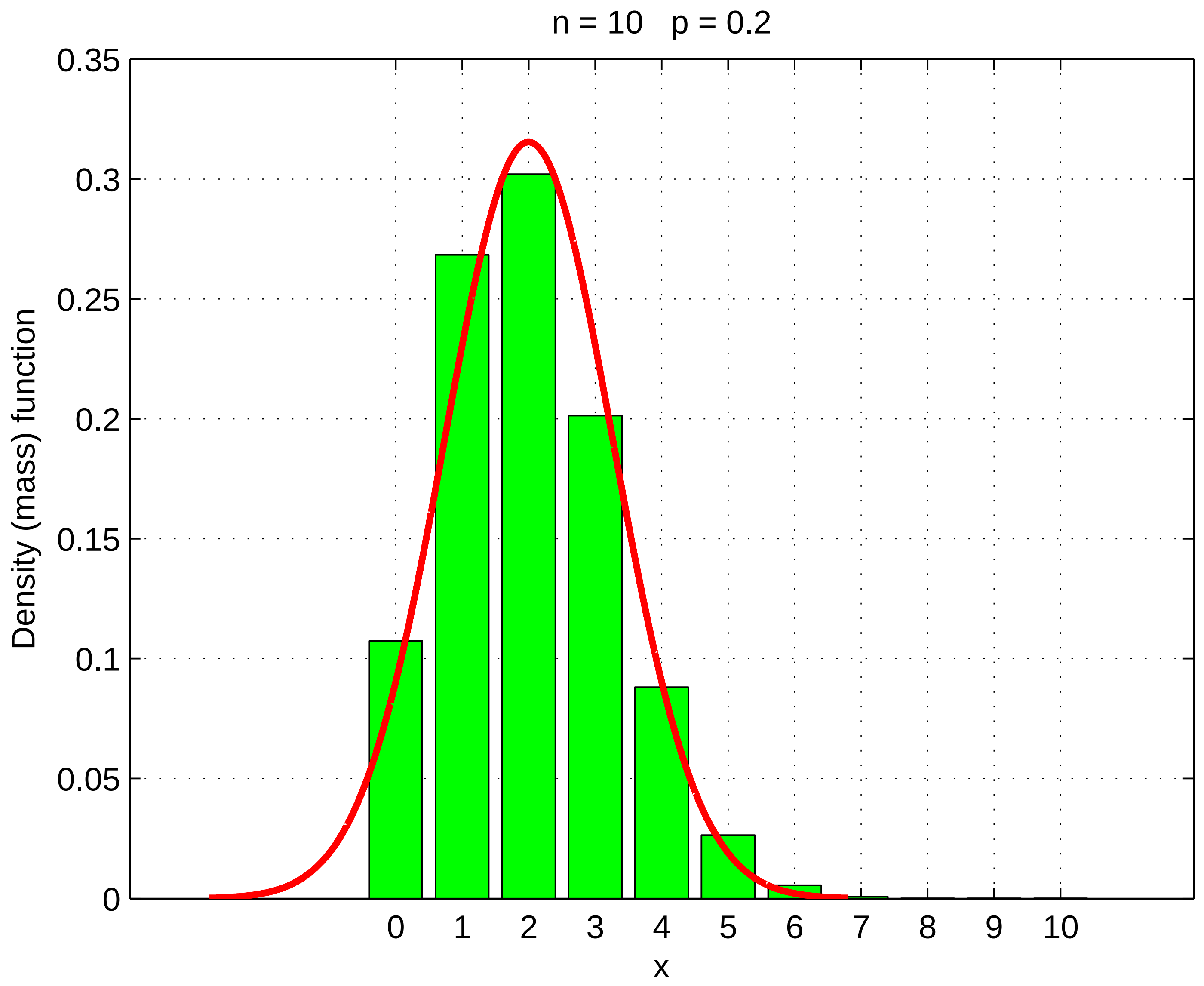
$$\lim_{t\to\infty} te^{-\frac{t^2}{2}} = \lim_{t\to\infty} \frac{t}{e^{t^2/2}} = \lim_{t\to\infty} \frac{[t]'}{\left[e^{t^2/2}\right]'} = \lim_{t\to\infty} \frac{1}{te^{t^2/2}} = 0$$

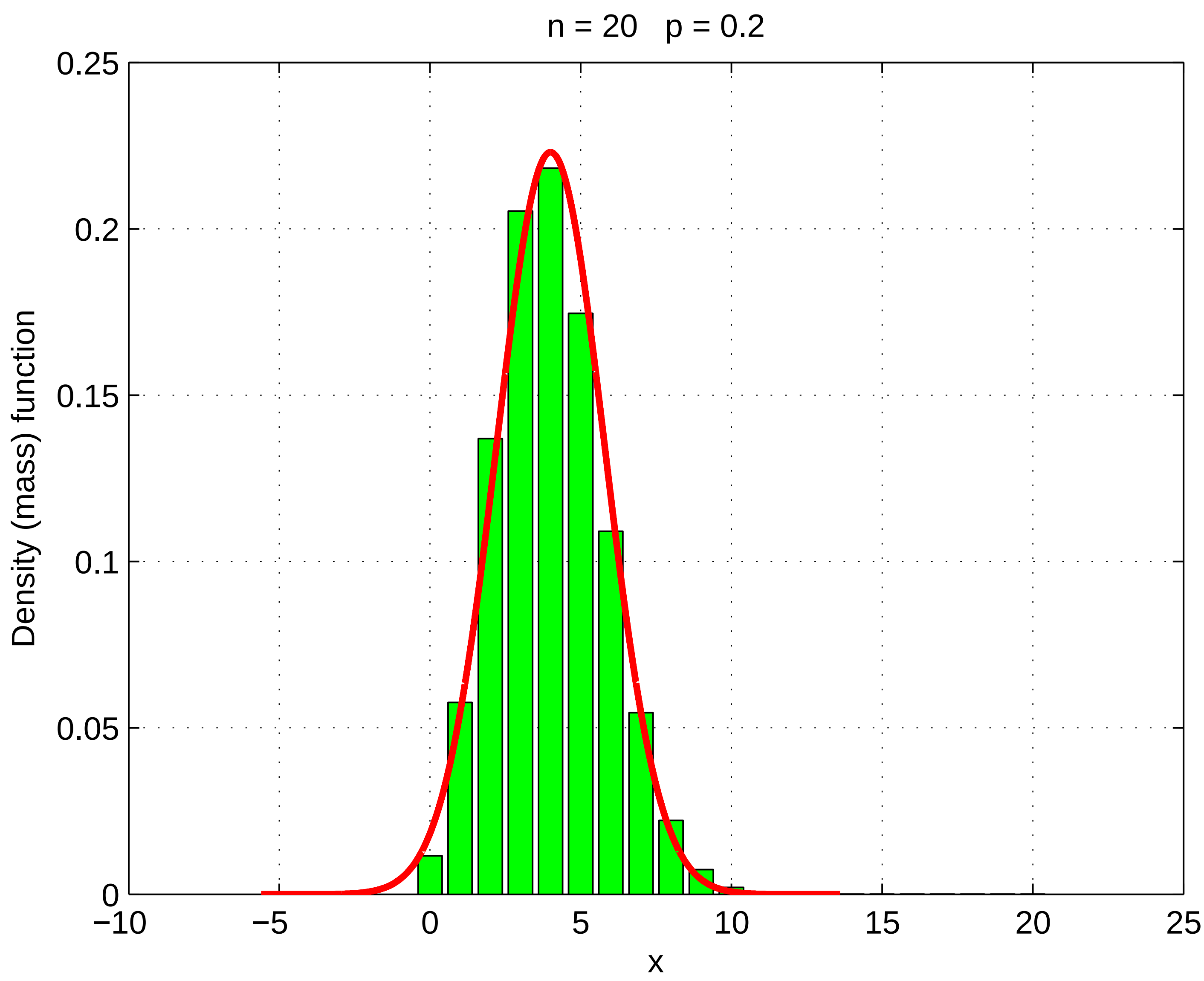
Normal Distribution Approximates Binomial

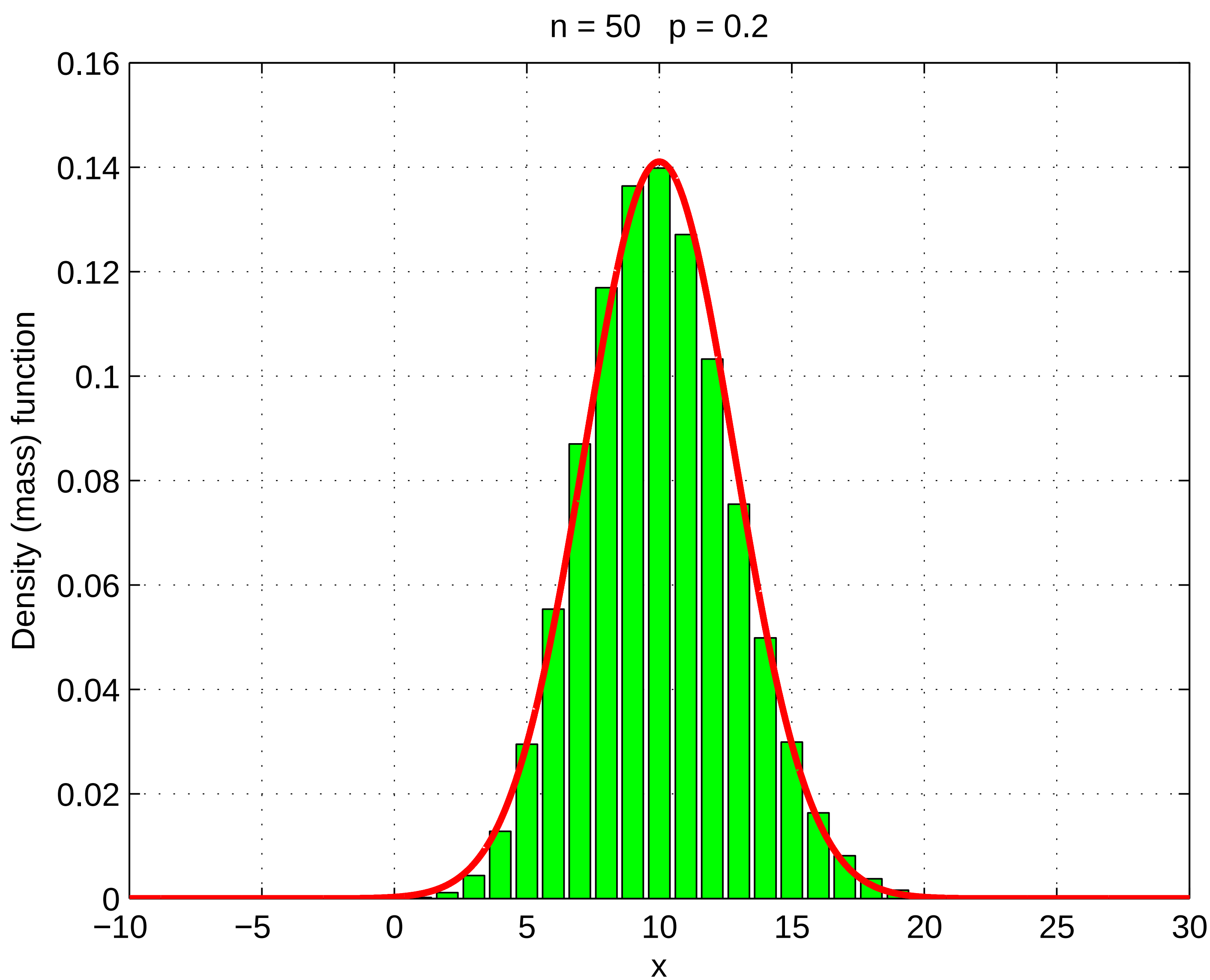
Suppose $X \sim Binomial(n, p)$. For fixed p, as $n \rightarrow \infty$

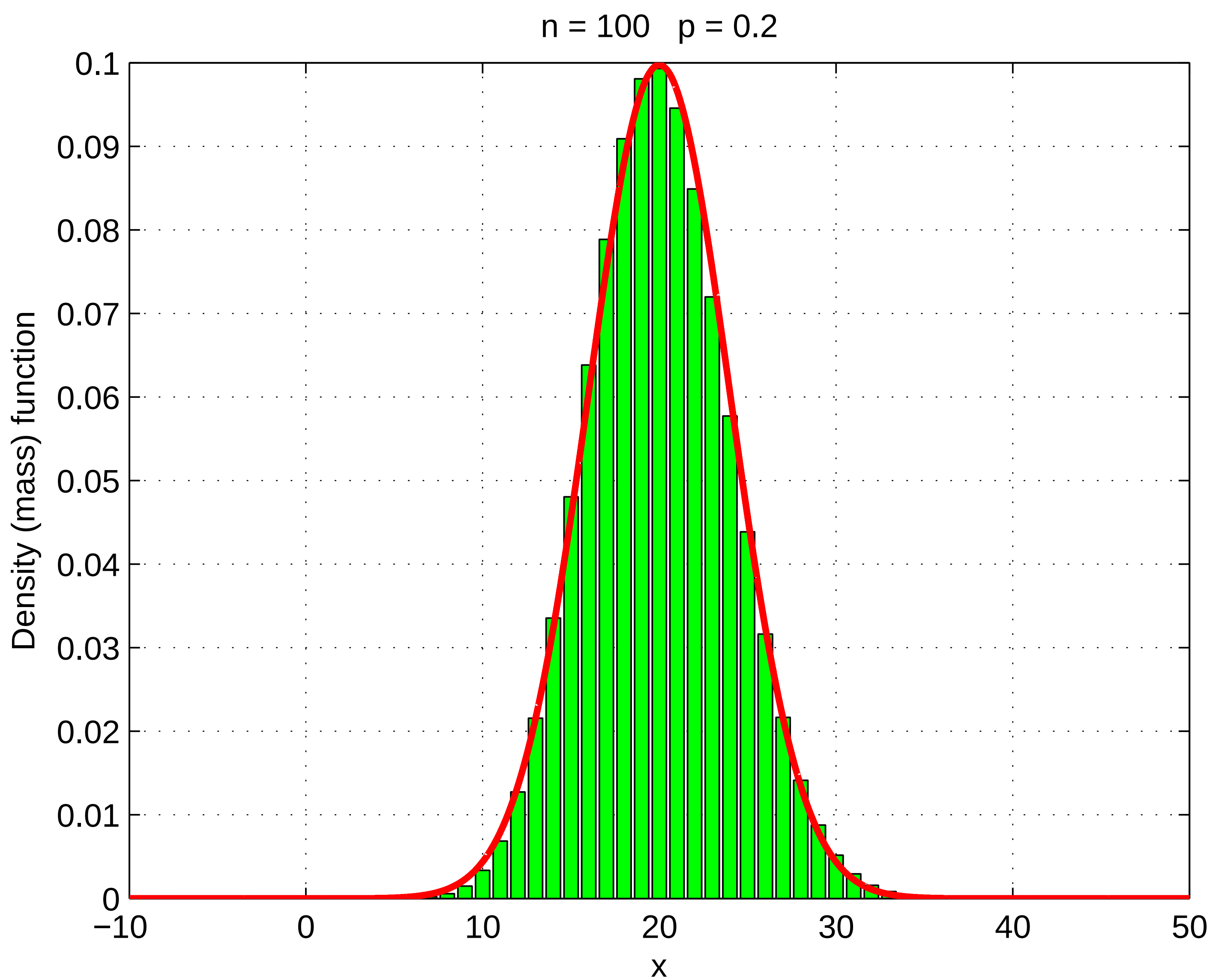
$$Binomial(n, p) \approx N(\mu, \sigma^2),$$

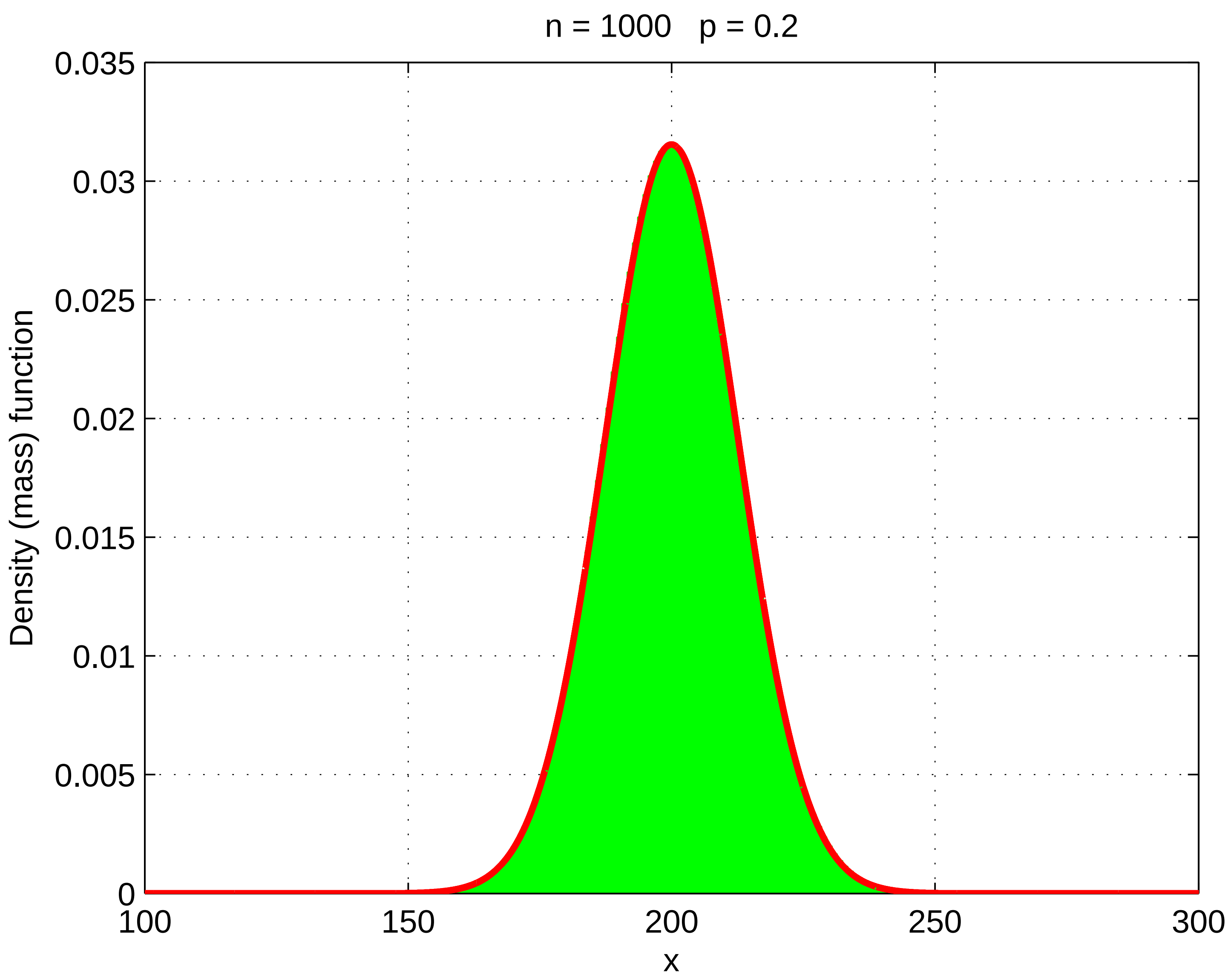
$$\mu = np, \quad \sigma^2 = np(1-p).$$


n = 10 p = 0.2
Density (mass) function
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
1
2
3
4
5
6
7
8
9
10
x


n = 20 p = 0.2
Density (mass) function
x
0.25
0.2
0.15
0.1
0.05
0
−10
−5
0
5
10
15
20
25


n = 50 p = 0.2
Density (mass) function
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
−10
−5
0
5
10
15
20
25
30
x


n = 100 p = 0.2
Density (mass) function
x
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
−10
0
10
20
30
40
50


n = 1000 p = 0.2
Density (mass) function
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
100
150
200
250
300
x

Matlab code

```
function NormalApprBinomial(n,p);

mu = n*p; sigma2 = n*p*(1-p);

figure;
bar((0:n),binopdf(0:n,n,p),'g');hold on; grid on;
x = mu - 3*sigma2:0.001:mu+3*sigma2;
plot(x,normpdf(x,mu,sqrt(sigma2)),'r-','linewidth',2);
xlabel('x');ylabel('Density (mass) function');
title(['n = ' num2str(n) '   p = ' num2str(p)]);
```

The Standard Normal Distribution

$Z \sim N(0,1)$ is called the standard normal.

- $E(Z) = 0, Var(Z) = 1.$

- If $X \sim N\left(\mu, \sigma^2\right)$, then $Y = \frac{x-\mu}{\sigma} \sim N(0,1)$.

- $Z \sim N(0,1)$. Denote density function $f_Z(z)$ by $\phi(z)$ and cumulative function $F_Z(z)$ by $\Phi(z)$.

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \qquad \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

- $\phi(-x) = \phi(x), \quad \Phi(-x) = 1 - \Phi(x).$ Numerical table on Page 222.

Example 5.4.4b: $X \sim N(\mu = 3, \sigma^2 = 9)$. Find

- $P\{2 < X < 5\}$
- $P\{X > 0\}$
- $P\{|X - 3| > 6\}$.

Solution:

Let $Y = \frac{X-\mu}{\sigma} = \frac{X-3}{3}$. Then $Y \sim N(0, 1)$.

$$
\begin{aligned}
P\{2 < X < 5\} =& P\left\{\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right\} \\
=& P\left\{-\frac{1}{3} < Y < \frac{2}{3}\right\} = \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \\
=& \Phi\left(0.6667\right) - 1 + \Phi\left(0.3333\right) \\
\approx& 0.7475 - 1 + 0.6306 \approx 0.3781
\end{aligned}
$$

Section 5.5: Exponential Random Variable

Definition: An exponential random variable X has the probability density function given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \qquad (\lambda > 0)$$

$X \sim EXP(\lambda)$.

Cumulative function:

$$F(x) = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}, \qquad (x \geq 0).$$

$F(\infty) = ??$

Moments:

$$
\begin{aligned}
E(X^n) &= \int_0^\infty x^n \lambda e^{-\lambda x} dx \\
&= -\int_0^\infty x^n de^{-\lambda x} \\
&= \int_0^\infty e^{-\lambda x} n x^{n-1} dx \\
&= \frac{n}{\lambda} E(X^{n-1})
\end{aligned}
$$

$E(X) = \frac{1}{\lambda}$, $E(X^2) = \frac{2}{\lambda^2}$, $Var(X) = \frac{1}{\lambda^2}$, $E(X^n) = \frac{n!}{\lambda^n}$,

Applications of Exponential Distribution

- The exponential distribution often arises as being the distribution of the amount of time until some specific event occurs.

- For instance, the amount of time (starting from now) until an earthquake occurs, or until a new war breaks out, etc.

- The exponential distribution is memoryless.

$$P\{X > s + t | X > t\} = P\{X > s\}, \quad \text{for all } \; s, t \geq 0$$

Example 5b: Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = \frac{1}{10}$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait

- (a) more than 10 minutes;
- (b) between 10 and 20 minutes.

Solution:

- $P\{X > 10\} = e^{-\lambda 10} = e^{-1} = 0.368.$
- $P\{10 < X < 20\} = e^{-1} - e^{-2} = 0.233$

Section 5.6.1: The Gamma Distribution

Definition: A random variable has a gamma distribution with parameters (α, λ), where $\alpha, \lambda > 0$, if its density function is given by

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Denote $X \sim Gamma(\alpha, \lambda)$.

$\int_0^\infty f(x)dx = 1$??

The Gamma function is defined as

$$\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy = (\alpha-1)\Gamma(\alpha-1).$$

$\Gamma(1) = 1.$ $\Gamma(0.5) = \sqrt{\pi}.$ $\Gamma(n) = (n-1)!.$

Connection to exponential distribution

- When $\alpha = 1$, the gamma distribution reduces to an exponential distribution. $Gamma(\alpha = 1, \lambda) = EXP(\lambda)$.

- A gamma distribution $Gamma(n, \lambda)$ often arises as the distribution of the amount of time one has to wait until a total of n events has occurred.

$X \sim Gamma(\alpha, \lambda)$.

Expectation and Variance:

$E(X) = \frac{\alpha}{\lambda}, \qquad Var(X) = \frac{\alpha}{\lambda^2}.$

$$\begin{aligned}
E(X) =& \int_0^\infty x \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} dx \\
=& \int_0^\infty \frac{e^{-\lambda x}(\lambda x)^{\alpha}}{\Gamma(\alpha)} dx \\
=& \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{1}{\lambda} \int_0^\infty \lambda \frac{e^{-\lambda x}(\lambda x)^{\alpha}}{\Gamma(\alpha+1)} dx \\
=& \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{1}{\lambda} \\
=& \frac{\alpha}{\lambda}.
\end{aligned}$$

Section 5.6.4: The Beta Distribution

Definition: A random variable X has a beta distribution if its density is given by

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

Denote $X \sim Beta(a,b)$.

The Beta function

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1}dx$$

$X \sim Beta(a, b)$.

Expectation and Variance:

$$E(X) = \frac{a}{a+b},$$

$$Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

Sec. 5.7: Distribution of a Function of a Random Variable

Given the density function $f_X(x)$ and cumulative function $F_X(x)$ for a random variable X, what is the general formula for the density function $f_Y(y)$ of the random variable $Y = g(X)$?

Example 5.7.7a: $X \sim Uniform(0,1)$. $Y = g(X) = X^n$.

$$F_Y(y) = P\{Y \leq y\} = P\{X^n \leq y\} = P\{X \leq y^{1/n}\} = F_X(y^{1/n}) = y^{1/n}$$

$$\begin{aligned} f_Y(y) =& \frac{1}{n} y^{1/n-1}, \qquad 0 \leq y \leq 1 \\ =& f_X\left(g^{-1}(y)\right) \left| \frac{d}{dy} g^{-1}(y) \right| \end{aligned}$$

$y = g(x) = x^n \implies g^{-1}(y) = y^{1/n}$.

Theorem 7.1: Let X be a continuous random variable having probability density function f_X. Suppose $g(x)$ is a strictly monotone (increasing or decreasing), differentiable function of x. Then the random variable $Y = g(X)$ has a probability density function given by

$$f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\left|\frac{d}{dy}g^{-1}(y)\right| & \text{if } y = g(x) \text{ for some } x \\ 0 & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

Example 7b: $Y = X^2$

$$
\begin{aligned}
F_Y(y) = P\{Y \leq y\} = P\left\{X^2 \leq y\right\} \\
=& P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\
=& F_X(\sqrt{y}) - F_X(-\sqrt{y})
\end{aligned}
$$

$$
f_Y(y) = \frac{1}{2\sqrt{y}}\left(f_X(\sqrt{y}) + f_X(-\sqrt{y})\right)
$$

Does this example violate Theorem 7.1?

Example (Cauchy distribution): Suppose a random variable X has density function

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty$$

Q: Find the distribution of $\frac{1}{X}$.

One can check $\int_{-\infty}^{\infty} f(x) = 1$, but $E(X) = \infty$.

Solution: Denote the cumulative function of X by $F_X(x)$.

Let $Z = 1/X$. Assume $z > 0$,

$$\begin{aligned}
F_Z(z) =& P\{1/X \leq z\} = P\{X \leq 0\} + P\{X \geq 1/z, X \geq 0\} \\
=& \frac{1}{2} + \int_{1/z}^{\infty} \frac{1}{\pi(1+x^2)} dx \\
=& \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(z)\Big|_{1/z}^{\infty} = \frac{1}{2} + \frac{1}{\pi}\left(\frac{\pi}{2} - \tan^{-1}(1/z)\right)
\end{aligned}$$

Therefore, for $z > 0$,

$$f_Z(z) = \frac{1}{\pi}\left[\frac{\pi}{2} - \tan^{-1}(1/z)\right]' = \frac{1}{\pi(1+z^2)}.$$

Assume $z < 0$,

$$\begin{aligned} F_Z(z) =& P\{1/X \leq z\} = P\{X \geq 1/z, X \leq 0\} \\ =& \int_{1/z}^{0} \frac{1}{\pi(1+x^2)} dx \\ =& \frac{1}{\pi} \tan^{-1}(z)\Big|_{1/z}^{0} = \frac{1}{\pi}\left(0 - \tan^{-1}(1/z)\right) \end{aligned}$$

Therefore, for $z < 0$,

$$f_Z(z) = \frac{1}{\pi}\left[0 - \tan^{-1}(1/z)\right]' = \frac{1}{\pi(1+z^2)}.$$

Thus, Z and X have the same density function.

Chapter 6: Jointly Distributed Random Variables

Definition: For any two random variables X and Y, the joint cumulative probability distribution is defined as

$$F(x,y) = P\left\{X \leq x, Y \leq y\right\}, \qquad -\infty < x, y < \infty$$

$$F(x,y) = P\{X \leq x, Y \leq y\}, \qquad -\infty < x, y < \infty$$

The distribution of X is

$$\begin{aligned} F_X(x) =& P\{X \leq x\} \\ =& P\{X \leq x, Y \leq \infty\} \\ =& F(x, \infty) \end{aligned}$$

The distribution of Y is

$$F_Y(y) = P\{Y \leq y\} = F(\infty, y)$$

Example:

$$P\{X > x, Y > y\} = 1 - F_X(x) - F_Y(y) + F(x,y)$$

Jointly Discrete Random Variables

X and Y are both discrete random variables.

Joint probability mass function

$$p(x,y) = P\{X = x, Y = y\}, \qquad \sum_x \sum_y p(x,y) = 1$$

Marginal probability mass function

$$p_X(x) = P\{X = x\} = \sum_{y:p(x,y)>0} p(x,y)$$

$$p_Y(y) = P\{Y = y\} = \sum_{x:p(x,y)>0} p(x,y)$$

Example 1a: Suppose that 3 balls are randomly selected from an urn containing 3 red, 4 white, and 5 blue balls. Let X and Y denote, respectively, the number of red and white balls chosen.

Q: The joint probability mass function $p(x, y) = P\{X = x, Y = y\}$=??

$$p(0,0) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}$$

$$p(0,1) = \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220}$$

$$p(0,2) = \frac{\binom{4}{2}\binom{5}{1}}{\binom{12}{3}} = \frac{30}{220}$$

$$p(0,3) = \frac{\binom{4}{3}}{\binom{12}{3}} = \frac{4}{220}$$

$$p_X(0) = p(0,0) + p(0,1) + p(0,2) + p(0,3) = \frac{84}{220}$$

$$p_X(0) = \frac{\binom{9}{3}}{\binom{12}{3}} = \frac{84}{220}$$

Jointly Continuous Random Variables

X and Y are jointly continuous if there exists a function $f(x,y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers

$$P\{(X,Y) \in C\} = \iint_{(x,y)\in C} f(x,y)dxdy$$

The function $f(x,y)$ is called the joint probability density function.

The joint cumulative function

$$F(x,y) = P\{X \in (-\infty, x], Y \in (-\infty, y]\}$$
$$= \int_{-\infty}^{y}\int_{-\infty}^{x} f(x,y)dxdy$$

The joint probability density function

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$$

The marginal probability density functions

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Example 1c: The joint density function is given by

$$f(x,y) = \begin{cases} Ce^{-x}e^{-2y} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

- (a) Determine C.
- (b) $P\{X > 1, Y < 1\}$,
- (c) $P\{X < Y\}$,
- (d) $P\{X < a\}$.

Solution (a):

$$
\begin{aligned}
1 = & \int_0^\infty \int_0^\infty Ce^{-x}e^{-2y}dxdy \\
= & C\int_0^\infty e^{-x}dx \int_0^\infty e^{-2y}dy \\
= & C \times \left. e^{-x}\right|_\infty^0 \times \frac{1}{2}\left. e^{-2y}\right|_\infty^0 = \frac{C}{2}
\end{aligned}
$$

Therefore, $C = 2$.

Solution (b):

$$
\begin{aligned}
P\{X > 1, Y < 1\} = & \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}dxdy \\
= & \left. e^{-x}\right|_\infty^1 \left. e^{-2y}\right|_1^0 = e^{-1}(1 - e^{-2})
\end{aligned}
$$

Solution (c):

$$
\begin{aligned}
P\{X<Y\} = & \iint_{(x,y):x<y} 2e^{-x}e^{-2y}dxdy = \\
= & \int_0^\infty \int_0^y 2e^{-x}e^{-2y}dxdy \\
= & \int_0^\infty \left[\int_0^y e^{-x}dx\right] 2e^{-2y}dy \\
= & \int_0^\infty \left[1-e^{-y}\right] 2e^{-2y}dy \\
= & \int_0^\infty 2e^{-2y} - 2e^{-3y}dy \\
= & 1-\frac{2}{3} = \frac{1}{3}
\end{aligned}
$$

Solution (d):

$$\begin{aligned}
P\{X < a\} =& P\{X < a, Y < \infty\} \\
=& \int_0^a \int_0^\infty 2e^{-2y} e^{-x} dy dx \\
=& \int_0^a e^{-x} dx \\
=& 1 - e^{-a}
\end{aligned}$$

Example 1e: The joint density is given by

$$f(x,y) = \begin{cases} e^{-(x+y)} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Q: Find the density function of the random variable X/Y.

Solution: Let $Z = X/Y,\ 0 < Z < \infty$.

$$\begin{aligned}F_Z(z) =& P\{Z < z\} = P\{X/Y \leq z\} \\ =& \iint_{x/y \leq z} e^{-(x+y)} dx dy \\ =& \int_0^\infty \left[\int_0^{yz} e^{-(x+y)} dx\right] dy \\ =& \int_0^\infty e^{-y} \left[e^{-x}|_{yz}^0\right] dy \\ =& \int_0^\infty e^{-y} - e^{-y-yz} dy \\ =& e^{-y}|_\infty^0 + \left.\frac{e^{-y-yz}}{z+1}\right|_0^\infty = 1 - \frac{1}{z+1}\end{aligned}$$

Therefore, $f_Z(z) = \left[1 - \frac{1}{z+1}\right]' = \frac{1}{(z+1)^2}$.

$$
\begin{aligned}
F_Z(z) =& P\left\{Z < z\right\} = P\left\{X/Y \leq z\right\} \\
=& \int\int_{x/y \leq z} e^{-(x+y)} dx dy \\
=& \int_0^\infty \left[\int_{\frac{x}{z}}^\infty e^{-(x+y)} dy\right] dx \\
=& \int_0^\infty e^{-x} \left[e^{-y}|_{\infty}^{\frac{x}{z}}\right] dy \\
=& \int_0^\infty e^{-x-\frac{x}{z}} dy \\
=& \left.\frac{e^{-x-\frac{x}{z}}}{1+\frac{1}{z}}\right|_{\infty}^{0} = \frac{1}{1+\frac{1}{z}} = 1 - \frac{1}{z+1}
\end{aligned}
$$

Multiple Jointly Distributed Random Variables

Joint cumulative distribution function

$$F(x_1, x_2, x_3, ..., x_n) = P\left\{X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n\right\}$$

Suppose X_1, X_2, ..., X_n are discrete random variables.

Joint probability mass function

$$p(x_1, x_2, ..., x_n) = P\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\}$$

X_1, X_2, ..., X_n are jointly continuous if there exists a function $f(x_1, x_2, ..., f_n)$ such that for any set C in n-dimensional space

$$P\{(X_1, X_2, ..., X_n) \in C\} = \int\int_{(x_1,x_2,...,x_n)\in C} ... \int f(x_1, x_2, ..., x_n)dx_1dx_2...dx_n$$

$f(x_1, x_2, ..., x_n)$ is the **joint probability density function**.

Section 6.2: Independent Random Variables

Definition: The random variable X and Y are said to be independent if for **any** two sets of real numbers A and B

$$P\{X \in A, Y \in B\} = P\{X \in A\} \times P\{Y \in B\}$$

equivalently

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\} \times P\{Y \leq y\}$$

or, equivalently,

$$\begin{cases} p(x,y) = p_X(x)p_Y(y) & \text{for } \textbf{all } x, y \quad \text{when } X \text{ and } Y \text{ are discrete} \\ f(x,y) = f_X(x)f_Y(y) & \text{for } \textbf{all } x, y \quad \text{when } X \text{ and } Y \text{ are continous} \end{cases}$$

Example 2b: Suppose the number of people that enter a post office on a given day is a Poisson random variable with parameter λ.

Each person that enters the post office is male with probability p and a female with probability $1 - p$.

Let X and Y, respectively, be the number of males and females that enter the post office.

- **(a):** Find the joint probability mass function $P\{X = i, Y = j\}$.
- **(b):** Find the marginal probability mass functions $P\{X = i\}$, $P\{Y = j\}$.
- **(c):** Show that X and Y are independent.

Key: Suppose we know there are $N = n$ people that enter a post office. Then the number of males is a **binomial** with parameters (n, p).

$$[X|N = n] \sim Binomial\,(n,\ p)$$

$$[N = X + Y\,] \sim Poi(\lambda).$$

Therefore,

$$P\{X = i,\ Y =\} = \sum_{n=0}^{\infty} P\{X = i,\ Y = j|N = n\}P\{N = n\}$$

$$= P\{X = i,\ Y = j|N = i + j\}P\{N = i + j\}$$

because $P\{X = i,\ Y = j|N \neq i + j\} = 0.$

Solution (a) Conditional on $X + Y = i + j$:

$$P\{X = i, Y = j\}$$

$$= P\{X = i, Y = j | X + Y = i + j\} P\{X + Y = i + j\}$$

$$= \left[\binom{i+j}{i} p^i (1-p)^j\right] \left[e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}\right]$$

$$= \frac{p^i(1-p)^j e^{-\lambda} \lambda^i \lambda^j}{i!j!}$$

$$= \frac{p^i(1-p)^j e^{-\lambda(p+(1-p))} \lambda^i \lambda^j}{i!j!}$$

$$= \left[e^{-\lambda p} \frac{(\lambda p)^i}{i!}\right] \left[e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}\right]$$

Solution (b)

$$P\{X=i\} = \sum_{j=0}^{\infty} P\{X=i, Y=j\}$$

$$= \sum_{j=0}^{\infty} \left[e^{-\lambda p}\frac{(\lambda p)^i}{i!}\right]\left[e^{-\lambda(1-p)}\frac{(\lambda(1-p))^j}{j!}\right]$$

$$= \left[e^{-\lambda p}\frac{(\lambda p)^i}{i!}\right]\sum_{j=0}^{\infty}\left[e^{-\lambda(1-p)}\frac{(\lambda(1-p))^j}{j!}\right]$$

$$= \left[e^{-\lambda p}\frac{(\lambda p)^i}{i!}\right]$$

Example 2c: A man and a woman decide to meet at a certain location. Each person independently arrives at a time uniformly distributed between 12 noon and 1 pm.

Q: Find the probability that the first to arrive has to wait longer than 10 minutes.

Solution:

Let X = number of minutes by which one person arrives later than 12 noon.
Let Y = number of minutes by which the other person arrives later than 12 noon.
X and Y are independent.

The desired probability = $P\{|X - Y| \geq 10\}$.

$$
\begin{aligned}
P\{|X-Y|\geq 10\} =& \iint_{|x-y|>10} f(x,y)dxdy \\
=& \iint_{|x-y|>10} f_X(x)f_Y(y)dxdy \\
=&2\iint_{y<x-10} \frac{1}{60}\frac{1}{60}dxdy \\
=&2\int_{10}^{60}\int_{0}^{x-10} \frac{1}{3600}dydx \\
=&\frac{25}{36}.
\end{aligned}
$$

A Necessary and Suffice Condition for Independence

Proposition 2.1 The continuous (discrete) random variables X and Y are independent if and only if their joint probability density (mass) function can be expressed as

$$f_{X,Y}(x,y) = h(x)g(y), \qquad -\infty < x, y < \infty$$

Proof: By definition, X and Y are independent if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Example 2f:

(a): Two random variables X and Y with joint density given by

$$f(x,y) = 6e^{-2x}e^{-3y}, \qquad 0 < x, y < \infty$$

are independent.

(b): Two random variables X and Y with joint density given by

$$f(x,y) = 24xy \qquad 0 < x, y < 1,\ 0 < x+y < 1$$

are NOT independent. (Why?)

Independence of More Than Two Random Variables

Definition: The n random variables, X_1, X_2, ..., X_n, are said to be independent if, for all sets of real numbers, A_1, A_2, ..., A_n

$$P\{X_1 \in A_1, X_2 \in A_2, ..., X_n \in A_n\} = \prod_{i=1}^{n} P\{X_i \in A_i\}$$

equivalently,

$$P\{X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n\} = \prod_{i=1}^{n} P\{X_i \leq x_i\},$$

for all $x_1, x_2, ..., x_n$.

Example 2h: Let X, Y, Z be independent and uniformly distributed over $(0,\ 1)$. Compute $P\{X \geq YZ\}$.

Solution:

$$\begin{aligned}
P\{X \geq YZ\} =& \iiint_{x \geq yz} f(x,y,z) dxdydz \\
=& \int_0^1 \int_0^1 \int_{yz}^1 dxdydz \\
=& \int_0^1 \int_0^1 (1 - yz)\, dydz \\
=& \int_0^1 \left(1 - \frac{z}{2}\right) dz \\
=& \frac{3}{4}
\end{aligned}$$

Chapter 6.3: Sum of Independent Random Variables

Suppose that X and Y are independent. Let $Z = X + Y$.

$$
\begin{aligned}
F_Z(z) =& P\{X+Y\leq z\} = \iint_{x+y\leq z} f_X(x)f_Y(y)dxdy\\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f_X(x)f_Y(y)dxdy\\
=& \int_{-\infty}^{\infty} f_Y(y)\int_{-\infty}^{z-y} f_X(x)dxdy\\
=& \int_{-\infty}^{\infty} f_Y(y)F_X(z-y)dy
\end{aligned}
$$

$$
f_Z(z) = \frac{\partial}{\partial z}F_Z(z) = \int_{-\infty}^{\infty} f_Y(y)f_X(z-y)dy
$$

The discrete case: X and Y are two independent discrete random variables, then $Z = X + Y$ has probability mass function:

$$p_Z(z) = \sum_j p_Y(j) p_X(z-j)$$

Example: Suppose $X, Y \sim Bernouli(p)$. Then

$$Z = X + Y \sim Binomial(2, p)$$

Suppose $X, Y \sim Bernouli(p)$ are independent. Then $Z = X + Y \sim Binomial(2, p)$.

$$p_Z(0) = \sum_j p_Y(j)p_X(0-j) = p_Y(0)p_X(0) = (1-p)^2 = \binom{2}{0}(1-p)^2$$

$$p_Z(1) = \sum_j p_Y(j)p_X(1-j) = p_Y(0)p_X(1-0) + p_Y(1)p_X(1-1)$$

$$= (1-p)p + p(1-p) = \binom{2}{1}p(1-p)$$

$$p_Z(2) = \sum_j p_Y(j)p_X(2-j) = p_Y(1)p_X(2-1) = p^2 = \binom{2}{2}p^2$$

Suppose $X \sim Binomial(n,p)$ and $Y \sim Binomial(m,p)$ are independent. Then $Z = X+Y \sim Binomial(m+n,p)$.

$$p_Z(z) = \sum_j p_Y(j)p_X(z-j).$$

Note $0 \leq j \leq m$ and $z-n \leq j \leq z$. We can assume $m \leq n$.

There are three cases to consider:

1. Suppose $0 \leq z \leq m$, then $0 \leq j \leq z$
2. Suppose $m < z \leq n$, then $0 \leq j \leq m$.
3. Suppose $n < z \leq m+n$, then $z-n \leq j \leq m$.

Suppose $0 \leq z \leq m$, then $0 \leq j \leq z$

$$
\begin{aligned}
p_Z(z) = & \sum_j p_Y(j)p_X(z-j) = \sum_{j=0}^{z} p_Y(j)p_X(z-j) \\
= & \sum_{j=0}^{z} \binom{m}{j} p^j(1-p)^{m-j} \binom{n}{z-j} p^{z-j}(1-p)^{n-z+j} \\
= & p^z(1-p)^{m+n-z} \sum_{j=0}^{z} \binom{m}{j}\binom{n}{z-j} \\
= & \binom{m+n}{z} p^z(1-p)^{m+n-z}
\end{aligned}
$$

because we have seen (e.g., using a combinatorial argument)

$$
\sum_{j=0}^{z} \binom{m}{j}\binom{n}{z-j} = \binom{m+n}{z}.
$$

Sum of Two Independent Uniform Random Variables

$X \sim U(0,1)$ and $Y \sim U(0,1)$ are independent. Let $Z = X + Y$.

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)dy = \int_0^1 f_X(z-y)dy$$

Must have $0 \leq y \leq 1$ and $0 \leq z - y \leq 1$, i.e., $z - 1 \leq y \leq z$.

Two cases to consider

1. If $0 \leq z \leq 1$, then $0 \leq y \leq z$.

2. If $1 < z \leq 2$, then $z - 1 < y \leq 1$.

If $0 \leq z \leq 1$, then $0 \leq y \leq z$.

$$f_Z(z) = \int_0^1 f_X(z-y)dy = \int_0^z dy = z.$$

If $1 < z \leq 2$, then $z - 1 \leq y \leq 1$.

$$f_Z(z) = \int_0^1 f_X(z-y)dy = \int_{z-1}^1 dy = 2 - z.$$

$$f_Z(z) = \begin{cases} z & 0 < z \leq 1 \\ 2 - z & 1 < z < 2 \\ 0 & \text{otherwise} \end{cases}$$

Sum of Two Gamma Random Variables

$X \sim Gamma(t, \lambda)$ and $Y \sim Gamma(s, \lambda)$ are independent.

Let $Z = X + Y$.

Show that $Z \sim Gamma(s + t, \lambda)$.

For an intuition, see notes page 246.

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)dy = \int_0^z f_Y(y) f_X(z-y)dy$$

$$= \int_0^z \left[\frac{\lambda e^{-\lambda y}(\lambda y)^{s-1}}{\Gamma(s)}\right]\left[\frac{\lambda e^{-\lambda(z-y)}(\lambda(z-y))^{t-1}}{\Gamma(t)}\right] dy$$

$$= \frac{e^{-\lambda z}\lambda^2 \lambda^{s+t-2}}{\Gamma(s)\Gamma(t)} \int_0^z \left[e^{-\lambda y} y^{s-1}\right]\left[e^{\lambda y}(z-y)^{t-1}\right] dy$$

$$= \frac{\lambda e^{-\lambda z}\lambda^{s+t-1}}{\Gamma(s)\Gamma(t)} \int_0^z y^{s-1}(z-y)^{t-1} dy$$

$$
\begin{aligned}
f_Z(z) =& \frac{\lambda e^{-\lambda z}\lambda^{s+t-1}}{\Gamma(s)\Gamma(t)}\int_0^z y^{s-1}(z-y)^{t-1}dy \\
=& \left[\frac{\lambda e^{-\lambda z}(\lambda z)^{s+t-1}}{\Gamma(s+t)}\right]\left[\frac{\Gamma(s+t)}{z^{s+t-1}\Gamma(s)\Gamma(t)}\int_0^z y^{s-1}(z-y)^{t-1}dy\right] \\
=& \left[\frac{\lambda e^{-\lambda z}(\lambda z)^{s+t-1}}{\Gamma(s+t)}\right]\left[\frac{\Gamma(s+t)}{\Gamma(s)\Gamma(t)}\int_0^z \left(\frac{y}{z}\right)^{s-1}\left(1-\frac{y}{z}\right)^{t-1} d\frac{y}{z}\right] \\
=& \left[\frac{\lambda e^{-\lambda z}(\lambda z)^{s+t-1}}{\Gamma(s+t)}\right]\left[\frac{\Gamma(s+t)}{\Gamma(s)\Gamma(t)}\int_0^1 \theta^{s-1}\left(1-\theta\right)^{t-1} d\theta\right] \\
=& f_W(z)\left[\frac{\Gamma(s+t)}{\Gamma(s)\Gamma(t)}\int_0^1 \theta^{s-1}\left(1-\theta\right)^{t-1} d\theta\right]
\end{aligned}
$$

$W \sim Gamma(s+t,\lambda)$.

Because $f_Z(z)$ is also a density function, we must have

$$
\left[\frac{\Gamma(s+t)}{\Gamma(s)\Gamma(t)}\int_0^1 \theta^{s-1}\left(1-\theta\right)^{t-1} d\theta\right] = 1
$$

Recall the definition of Beta function in Section 5.6.4

$$B(s,t) = \int_0^1 \theta^{s-1}\left(1-\theta\right)^{t-1} d\theta$$

and its property (Eq. (6.3) in Section 5.6.4)

$$B(s,t) = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}.$$

Thus,

$$\left[\frac{\Gamma(s+t)}{\Gamma(s)\Gamma(t)}\int_0^1 \theta^{s-1}\left(1-\theta\right)^{t-1} d\theta\right] = \left[\frac{\Gamma(s+t)}{\Gamma(s)\Gamma(t)}\frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}\right] = 1$$

Generalization to more than two gamma random variables

If X_1, X_2, ..., X_n are independent gamma random variables with respective parameters (s_1, λ), (s_2, λ), ..., (s_n, λ), then

$$X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i \sim Gamma\left(\sum_{i=1}^{n} s_i, \lambda\right)$$

Sum of Independent Normal Random Variables

Proposition 3.2 If X_i, $i = 1$ to n, are independent random variables that are normally distributed with respective parameters u_i and σ_i^2, $i = 1$ to n. Then

$$X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n}\mu_i,\ \sum_{i=1}^{n}\sigma_i^2\right)$$

It suffices to show

$$X_1 + X_2 \sim N\left(\mu_1 + \mu_2,\ \sigma_1^2 + \sigma_2^2\right)$$

$X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$. Let $Z = X + Y$.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$$

$$= \int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(z-y-\mu_x)^2}{2\sigma_x^2}}\right] \left[\frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}\right] dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y} e^{-\frac{(z-y-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y} e^{-\frac{(z-y-\mu_x)^2\sigma_y^2 + (y-\mu_y)^2\sigma_x^2}{2\sigma_x^2\sigma_y^2}} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y} e^{-\frac{y^2(\sigma_x^2+\sigma_y^2) - 2y(\mu_y\sigma_x^2 - (\mu_x - z)\sigma_y^2) + (\mu_x - z)^2\sigma_y^2 + \mu_y^2\sigma_x^2}{2\sigma_x^2\sigma_y^2}} dy$$

Let $A = \sigma_x\sigma_y$, $B = \sigma_x^2\sigma_y^2$, $C = \sigma_x^2 + \sigma_y^2$,
$D = -2(\mu_y\sigma_x^2 - (\mu_x - z)\sigma_y^2)$,
$E = (\mu_x - z)^2\sigma_y^2 + \mu_y^2\sigma_x^2$.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_x\sigma_y}e^{-\frac{y^2(\sigma_x^2+\sigma_y^2)-2y(\mu_y\sigma_x^2-(\mu_x-z)\sigma_y^2)+(\mu_x-z)^2\sigma_y^2+\mu_y^2\sigma_x^2}{2\sigma_x^2\sigma_y^2}}dy$$

$$= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}A}e^{-\frac{Cy^2+Dy+E}{2B}}dy = \sqrt{\frac{B}{CA^2}}e^{\frac{D^2-4EC}{8BC}}$$

because $\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}dy = 1$, for any μ, σ^2.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}A} e^{-\frac{Cy^2+Dy+E}{2B}} dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}A} e^{-\frac{y^2+\frac{D}{C}y+\frac{E}{C}}{2\frac{B}{C}}} dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}A} e^{-\frac{\left(y+\frac{D}{2C}\right)^2-\frac{D^2}{4C^2}+\frac{E}{C}}{2\frac{B}{C}}} dy$$

$$= e^{-\frac{-\frac{D^2}{4C^2}+\frac{E}{C}}{2\frac{B}{C}}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{\frac{B}{C}}\frac{A}{\sqrt{\frac{B}{C}}}} e^{-\frac{\left(y+\frac{D}{2C}\right)^2}{2\frac{B}{C}}} dy$$

$$= e^{-\frac{-\frac{D^2}{4C^2}+\frac{E}{C}}{2\frac{B}{C}}} \sqrt{\frac{B}{CA^2}} = \sqrt{\frac{B}{CA^2}} e^{\frac{D^2-4EC}{8BC}}$$

$$
\begin{aligned}
f_Z(z) =& \frac{1}{\sqrt{2\pi}}\sqrt{\frac{B}{CA^2}}e^{\frac{D^2-4EC}{8BC}}\\
=& \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_y^2}}e^{\frac{(\mu_y\sigma_x^2-(\mu_x-z)\sigma_y^2)^2-((\mu_x-z)^2\sigma_y^2+\mu_y^2\sigma_x^2)(\sigma_x^2+\sigma_y^2)}{2\sigma_x^2\sigma_y^2(\sigma_x^2+\sigma_y^2)}}\\
=& \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2+\sigma_y^2}}e^{-\frac{(z-\mu_x-\mu_y)^2}{2(\sigma_x^2+\sigma_y^2)}}
\end{aligned}
$$

Therefore, $Z = X + Y \sim N\left(\mu_x+\mu_y,\ \sigma_x^2+\sigma_y^2\right)$.

Sum of Independent Poisson Random Variables

If X_1, X_2, ..., X_n are independent Poisson random variables with respective parameters λ_i, $i = 1$ to n, then

$$X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i \sim Poisson\left(\sum_{i=1}^{n} \lambda_i\right)$$

Sum of Independent Binomial Random Variables

If X_1, X_2, ..., X_n are independent binomial random variables with respective parameters (m_i, p), $i = 1$ to n, then

$$X_1 + X_2 + ... + X_n = \sum_{i=1}^{n} X_i \sim Binomial\left(\sum_{i=1}^{n} m_i,\ p\right)$$

Sections 6.4, 6.5: Conditional Distributions

Recall $P(E|F) = \frac{P(EF)}{P(F)}$.

If X and Y are discrete random variables, then

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

If X and Y are continuous random variables, then

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

Example 4b: Suppose X and Y are independent Poisson random variables with respective parameters λ_x and λ_y.

Q: Calculate the conditional distribution of X, given that $X+Y=n$.

What is the distribution of $X+Y$??

Example 4b: Suppose X and Y are independent Poisson random variables with respective parameters λ_x and λ_y.

Q: Calculate the conditional distribution of X, given that $X+Y=n$.

Solution:

$$P\{X=k|X+Y=n\} = \frac{P\{X=k, X+Y=n\}}{P\{X+Y=n\}}$$

$$=\frac{P\{X=k, Y=n-k\}}{P\{X+Y=n\}} = \frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}}$$

$$=\left[\frac{e^{-\lambda_x}\lambda_x^k}{k!}\right]\left[\frac{e^{-\lambda_y}\lambda_y^{n-k}}{(n-k)!}\right]\left[\frac{n!}{e^{-\lambda_x-\lambda_y}(\lambda_x+\lambda_y)^n}\right]$$

$$=\binom{n}{k}\left(\frac{\lambda_x}{\lambda_x+\lambda_y}\right)^k\left(1-\frac{\lambda_x}{\lambda_x+\lambda_y}\right)^{n-k}$$

Therefore, $X \sim Binomial\left(n, p=\frac{\lambda_x}{\lambda_x+\lambda_y}\right)$.

Example 5b: The joint density of X and Y is given by

$$f(x,y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y} & 0 < x, y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Q: Find $P\{X > 1|Y = y\}$.

$$P\{X > 1|Y = y\} = \int_1^\infty f_{X|Y}(x|y)dx$$

Solution:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{e^{-x/y}e^{-y}}{y}}{\int_0^\infty \frac{e^{-x/y}e^{-y}}{y}dx} = \frac{e^{-x/y}}{\int_0^\infty e^{-x/y}dx} = \frac{1}{y}e^{-x/y}$$

$$P\{X > 1|Y = y\} = \int_1^\infty f_{X|Y}(x|y)dx = \int_1^\infty \frac{1}{y}e^{-x/y}dx = e^{-1/y}$$

The Bivariate Normal Distribution

The random variables X and Y have a bivariate normal distribution if, for constants, u_x, u_y, σ_x, σ_y, $-1 < \rho < 1$, their joint density function is given, for all $-\infty < x, y < \infty$, by

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]}$$

If X and Y are independent, then $\rho = 0$, and

$$\begin{aligned} f(x,y) =& \frac{1}{2\pi\sigma_x} e^{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]} \\ =& \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \times \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} \end{aligned}$$

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

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]}dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+t^2-2\rho\frac{(x-\mu_x)t}{\sigma_x}\right]}dt$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sqrt{1-\rho^2}} e^{-\frac{(x-\mu_x)^2+t^2\sigma_x^2-2\rho(x-\mu_x)t\sigma_x}{2(1-\rho^2)\sigma_x^2}}dt$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}A} e^{-\frac{Ct^2+Dt+E}{2B}}dt = \sqrt{\frac{B}{CA^2}}e^{\frac{D^2-4EC}{8BC}} = \frac{1}{\sqrt{2\pi}\sigma_x}e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$$

$$A = \sqrt{2\pi}\sqrt{1-\rho^2}\sigma_x, \qquad B = (1-\rho^2)\sigma_x^2, \qquad C = \sigma_x^2,$$

$$D = -2\rho(x-\mu_x)\sigma_x, \qquad E = (x-\mu_x)^2$$

Therefore,

$$X \sim N(\mu_x, \sigma_x^2), \qquad Y \sim N(\mu_y, \sigma_y^2)$$

The conditional density

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

A useful trick: Only worry about the terms that are functions of x; the rest terms are just normalizing constants.

Any (reasonable) non-negative function $g(x)$ can be essentially viewed as a probability density function because one can always normalize it.

$$f(x) = \frac{g(x)}{\int_{-\infty}^{\infty} g(x)dx} = Cg(x) \propto g(x), \qquad \text{where } C = \frac{1}{\int_{-\infty}^{\infty} g(x)dx}$$

is a probability density function. (One can verify $\int_{-\infty}^{\infty} f(x) = 1$).

C is just a (normalizing) constant.

When y is treated as a constant, $C(y)$ is still a constant.

$X|Y = y$ means y is a constant.

For example, $g(x) = e^{-\frac{(x-1)^2}{2}}$ is essentially the density function of $N(1,1)$.

Similarly, $g(x) = e^{-\frac{x^2-2x}{2}}$, $g(x) = e^{-\frac{x^2-2xy}{2}}$ are both (essentially) the density functions of normals.

$$\frac{1}{C} = \int_{-\infty}^{\infty} e^{-\frac{x^2-2xy}{2}}dx = \sqrt{2\pi}e^{\frac{y^2}{2}}$$

Therefore,

$$f(x) = Cg(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} \times g(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-y)^2}{2}}$$

is the density function of $N(y,1)$, where y is merely a constant.

(X, Y) is a bivariate normal.

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

$$\propto e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]}$$

$$\propto e^{-\frac{(x-\mu_x)^2 - 2\rho(x-\mu_x)(y-\mu_y)\frac{\sigma_x}{\sigma_y}}{2(1-\rho^2)\sigma_x^2}}$$

$$\propto e^{-\frac{\left(x-\mu_x-\rho(y-\mu_y)\frac{\sigma_x}{\sigma_y}\right)^2}{2(1-\rho^2)\sigma_x^2}}$$

Therefore,

$$X|Y \sim N\left(\mu_x + \rho(y-\mu_y)\frac{\sigma_x}{\sigma_y},\ (1-\rho^2)\sigma_x^2\right)$$

$$Y|X \sim N\left(\mu_y + \rho(x-\mu_x)\frac{\sigma_y}{\sigma_x},\ (1-\rho^2)\sigma_y^2\right)$$

Example 5d: Consider $n+m$ trials having a common probability of success. Suppose, however, that this success probability is not fixed in advance but is chosen from $U(0,1)$.

Q: What is the conditional distribution of this success probability given that the $n+m$ trials result in n successes?

Solution:

Let X = trial success probability. $X \sim U(0,1)$.

Let N = total number of successes. $N|X = x \sim Binomial(n+m, x)$.

The desired probability: $X|N = n$.

Solution:

Let X = trial success probability. $X \sim U(0,1)$.

Let N = total number of successes. $N|X=x \sim Binomial(n+m, x)$.

$$\begin{aligned}
f_{X|N}(x|n) =& \frac{P\{N=n|X=x\}f_X(x)}{P\{N=n\}}\\
=& \frac{\binom{n+m}{n}x^n(1-x)^m}{P\{N=n\}}\\
\propto& \, x^n(1-x)^m
\end{aligned}$$

Therefore, $X|N \sim Beta(n+1, m+1)$.

Here $X \sim U(0,1)$ is the prior distribution.

If $X|N \sim Beta(n+1, m+1)$, then

$$E(X|N) = \frac{n+1}{(n+1)+(m+1)}$$

Suppose we do not have a prior knowledge of the success probability X.

We observe n successes out of $n+m$ trials.

The most intuitive guess (estimate) of X should be

$$\hat{X} = \frac{n}{n+m}$$

Assuming a uniform prior on X leads to the **add-one smoothing**.

Section 6.6: Order Statistics

Let X_1, X_2, ..., X_n be n independent and identically distributed, continuous random variables having a common density f and distribution function F.

Define

$X_{(1)}$ = smallest of X_1, X_2, ..., X_n
$X_{(2)}$ = second smallest of X_1, X_2, ..., X_n
...
$X_{(j)}$ = jth smallest of X_1, X_2, ..., X_n
...
$X_{(n)}$ = largest of X_1, X_2, ..., X_n

The ordered values $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ are known as the order statistics corresponding to the random variables X_1, X_2, ..., X_n.

For example, the CDF of the largest (i.e., $X_{(n)}$) should be

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x)$$

because, if the largest is smaller than x, then everyone is smaller than x.

Then, by independence

$$\begin{aligned}F_{X_{(n)}}(x) =& P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x)\\ =& P(X_1 \leq x)P(X_2 \leq x)...P(X_n \leq x)\\ =& F(x)^n,\end{aligned}$$

and the density function is simply

$$f_{X_{(n)}}(x) = \frac{\partial F^n(x)}{\partial x} = nF^{n-1}(x)f(x)$$

The CDF of the smallest (i.e., $X_{(1)}$) should be

$$F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(X_{(1)} > x)$$
$$= 1 - P(X_1 > x, X_2 > x, ..., X_n > x)$$

because, if the smallest is larger than x, then everyone is larger than x.

Then, by independence

$$F_{X_{(1)}}(x) = 1 - P(X_1 > x)P(X_2 > x)...P(X_n > x)$$
$$= 1 - [1 - F(x)]^n$$

and the density function is simply

$$f_{X_{(1)}}(x) = \frac{\partial 1 - [1 - F(x)]^n}{\partial x} = n\,[1 - F(x)]^{n-1}\,f(x)$$

Joint density function of order statistics

for $x_1 \leq x_2 \leq ... \leq x_n$,

$$f_{X_{(1)},X_{(2)},...,X_{(n)}}(x_1, x_2, ..., x_n) = n! f(x_1) f(x_2) ... f(x_n),$$

Intuitively (and heuristically)

$$\begin{aligned}
&P\{X_{(1)} = x_1, X_{(2)} = x_2\}\\
=&P\{\{X_1 = x_1, X_2 = x_2\} \text{ OR } \{X_2 = x_1, X_1 = x_2\}\}\\
=&P\left\{X_1 = x_1, X_2 = x_2\right\} + P\left\{X_2 = x_1, X_1 = x_2\right\}\\
=&f(x_1)f(x_2) + f(x_1)f(x_2)\\
=&2! f(x_1)f(x_2)
\end{aligned}$$

Warning, this is a heuristic argument, not a rigorous proof. Please read the book.

Density function of the kth order statistics, $X_{(k)}$: $f_{X_{(k)}}(x)$.

- Select 1 from the n values X_1, X_2, ..., X_n. Let it be equal to x.
- Select $k-1$ from the remaining $n-1$ values. Let them be smaller than x.
- Let the rest be larger than x.
- Total number of choices is ??.
- For a given choice, the probability should be ??

- Select 1 from the n values X_1, X_2, ..., X_n. Let it be equal to x.
- Select $k-1$ from the remaining $n-1$ values. Let them be smaller than x.
- Let the rest be larger than x.
- Total number of choices is $\binom{n}{1,k-1,n-k}$.
- For a given choice, the probability should be $f(x)\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k}$.

Therefore,

$$f_{X_{(k)}}(x) = \binom{n}{1, k-1, n-k} f(x)\left[F(x)\right]^{k-1}\left[1-F(x)\right]^{n-k}$$

Cumulative function of the kth order statistics

- Via integrating the density function:

$$F_{X_{(k)}}(y) = \int_{-\infty}^{y} f_{X_{(k)}}(x)$$

$$= \int_{-\infty}^{y} \binom{n}{1, k-1, n-k} f(x) \left[F(x)\right]^{k-1} \left[1 - F(x)\right]^{n-k} dx$$

$$= \binom{n}{1, k-1, n-k} \int_{-\infty}^{y} f(x) \left[F(x)\right]^{k-1} \left[1 - F(x)\right]^{n-k} dx$$

$$= \binom{n}{1, k-1, n-k} \int_{0}^{F(y)} \left[z\right]^{k-1} \left[1 - z\right]^{n-k} dz$$

- Via a direct argument.

$$F_{X_{(k)}}(y) = P\{X_{(k)} \leq (y)\} = P\{\geq k \text{ of the } X_i\text{'s are } \leq y\}$$

$$\begin{aligned} F_{X_{(k)}}(y) &= P\{X_{(k)} \leq y\} \\ &= P\{\geq k \text{ of the } X_i\text{'s are } \leq y\} \\ &= \sum_{j=k}^{n} P\{j \text{ of the } X_i\text{'s are } \leq y\} \\ &= \sum_{j=k}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j} \end{aligned}$$

- A more intuitive solution by dividing y into non-overlapping segments.

$$\begin{aligned}\{X_{(k)} \leq y\} =& \{y \in [X_{(n)}, \infty)\} \bigcup \{y \in [X_{(n-1)}, X_{(n)})\} \\ & \bigcup ... \bigcup \{y \in [X_{(k)}, X_{(k+1)})\} \\ =& \sum_{j=k}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}\end{aligned}$$

$$P\{y \in [X_{(n)}, \infty)\} = \binom{n}{n} [F(y)]^n$$

$$P\{y \in [X_{(n-1)}, X_{(n)})\} = \binom{n}{n-1} [F(y)]^{n-1} [1 - F(y)]$$

$$P\{y \in [X_{(k)}, X_{(k+1)})\} = \binom{n}{k} [F(y)]^k [1 - F(y)]^{n-k}$$

An interesting identity by letting $X \sim U(0,1)$:

$$\sum_{j=k}^{n} \binom{n}{j} y^j \left[1-y\right]^{n-j}$$

$$=\binom{n}{1, k-1, n-k} \int_0^y x^{k-1} \left[1-x\right]^{n-k} dx, \qquad 0<y<1.$$

Joint Probability Distribution of Functions of Random Variables

X_1 and X_2 are jointly continuous random variables with probability density function f_{X_1,X_2}.

$$Y_1 = g_1(X_1, X_2), \qquad Y_2 = g_2(X_1, X_2).$$

The goal: Determine the density function f_{Y_1,Y_2}.

The formula to be remembered

$$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(x_1,x_2)|J(x_1,x_2)|^{-1}$$

where, the inverse functions:

$$x_1 = h_1(y_1,y_2), \qquad x_2 = h_2(y_1,y_2)$$

and the Jocobian:

$$J(x_1,x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_2}{\partial x_1}\frac{\partial g_1}{\partial x_2}$$

Two conditions

- $y_1 = g_1(x_1,x_2)$ and $y_2 = g_2(x_1,x_2)$ can be uniquely solved: $x_1 = h_1(y_1,y_2)$, and $x_2 = h_2(y_1,y_2)$
- The Jocobian $J(x_1,x_2) \neq 0$ for all points (x_1,x_2).

Example 7a: Let X_1 and X_2 be jointly continuous random variables with probability density f_{X_1,X_2}. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

Q: Find f_{Y_1,Y_2} in terms of f_{X_1,X_2}.

Solution:

$y_1 = g_1(x_1, x_2) = x_1 + x_2$, $y_2 = g_2(x_1, x_2) = x_1 - x_2$,
$\Longrightarrow$
$x_1 = h_1(y_1, y_2) = \frac{y_1+y_2}{2}$, $x_2 = h_2(y_1, y_2) = \frac{y_1-y_2}{2}$,

$$J(x_1, x_2) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2;$$

Therefore,

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) =& f_{X_1,X_2}(x_1, x_2)|J(x_1, x_2)|^{-1} \\ =& \frac{1}{2} f_{X_1,X_2}\left(\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}\right) \end{aligned}$$

Example 7b: The joint density in the polar coordinates.

$x = h_1(r,\theta) = r\cos\theta$, $y = h_2(r,\theta) = r\sin\theta$, i.e.,
$r = g_1(x,y) = \sqrt{x^2+y^2}$, $\theta = g_2(x,y) = \tan^{-1}\frac{y}{x}$.

Q (a): Express $f_{R,\Theta}(r,\theta)$ in terms of $f_{X,Y}(x,y)$.

Q (b): What if X and Y are independent standard normal random variables?

Solution: First, compute the Jocobian $J(x,y)$:

$$\frac{\partial g_1}{\partial x} = \frac{x}{\sqrt{x^2+y^2}} = \cos\theta$$

$$\frac{\partial g_1}{\partial y} = \frac{y}{\sqrt{x^2+y^2}} = \sin\theta$$

$$\frac{\partial g_2}{\partial x} = \frac{1}{1+(y/x)^2}\left(\frac{-y}{x^2}\right) = \frac{-y}{x^2+y^2} = \frac{-\sin\theta}{r}$$

$$\frac{\partial g_2}{\partial y} = \frac{1}{1+(y/x)^2}\left(\frac{1}{x}\right) = \frac{x}{x^2+y^2} = \frac{\cos\theta}{r}$$

$$J = \frac{\partial g_1}{\partial x}\frac{\partial g_2}{\partial y} - \frac{\partial g_1}{\partial y}\frac{\partial g_2}{\partial x} = \frac{\cos^2\theta}{r} + \frac{\sin^2\theta}{r} = \frac{1}{r}$$

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(x,y)|J|^{-1} = rf_{X,Y}(r\cos\theta, r\sin\theta)$$

where $0 < r < \infty$, $\quad 0 < \theta < 2\pi$.

If X and Y are independent standard normals, then

$$f_{R,\Theta}(r,\theta) = rf_{X,Y}(r\cos\theta, r\sin\theta) = \frac{r}{2\pi}e^{-\frac{x^2+y^2}{2}} = \frac{re^{-r^2/2}}{2\pi}$$

R and Θ are independent. $\Theta \sim U(0, 2\pi)$.

Example 7d: X_1, X_2, and X_3, are independent standard normal random variables. Let

$$Y_1 = X_1 + X_2 + X_3$$
$$Y_2 = X_1 - X_2$$
$$Y_3 = X_1 - X_3$$

Q: Compute the joint density function $f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3)$.

Solution:

Compute the Jocobian

$$J = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{vmatrix} = 3$$

Solve for X_1, X_2, and X_3

$$X_1 = \frac{Y_1 + Y_2 + Y_3}{3}, \qquad X_2 = \frac{Y_1 - 2Y_2 + Y_3}{3}, \qquad X_3 = \frac{Y_1 + Y_2 - 2Y_3}{3}$$

$$f_{Y_1,Y_2,Y_3}(y_1, y_2, y_3) = f_{X_1,X_2,X_3}(x_1, x_2, x_3)|J|^{-1}$$

$$= \frac{1}{3(2\pi)^{3/2}} e^{-\frac{x_1^2+x_2^2+x_3^2}{2}} = \frac{1}{3(2\pi)^{3/2}} e^{-\frac{1}{2}\left[\frac{y_1^2}{3} + \frac{2y_2^2}{3} + \frac{2y_3^2}{3} - \frac{2y_2y_3}{3}\right]}$$

More about Determinant

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh$$

$$\begin{vmatrix} x_1 & 0 & 0 & 0 & ... & 0 \\ \times & x_2 & 0 & 0 & ... & 0 \\ \times & \times & x_3 & 0 & ... & 0 \\ \times & \times & \times & x_4 & ... & 0 \\ \times & \times & \times & \times & ... & 0 \\ \times & \times & \times & \times & \times & x_n \end{vmatrix} = x_1 x_2 x_3 ... x_n$$

Chapter 7: Properties of Expectation

If X is a discrete random variable with mass function $p(x)$, then

$$E(X) = \sum_x xp(x)$$

If X is a continuous random variable with density $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

If $a \leq X \leq b$, then $a \leq E(X) \leq b$.

Proposition 2.1

If X and Y have a joint probability mass function $p(x, y)$, then

$$E[g(x, y)] = \sum_y \sum_x g(x, y)p(x, y)$$

If X and Y have a joint probability density function $f(x, y)$, then

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

What if $g(X, Y) = X + Y$?

Example 2a: An accident occurs at a point X that is uniformly distributed on a road of length L. At the time of the accident an ambulance is at a location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident.

Solution:

$X \sim U(0, L)$, $\quad Y \sim U(0, L)$. $\quad$ To find $E(|X - Y|)$.

$$E(|X - Y|) = \int\int |x - y| f(x, y) dx dy$$

Solution:

$X \sim U(0, L)$, $\quad Y \sim U(0, L)$. $\quad$ To find $E(|X - Y|)$.

$$\begin{aligned}
E(|X - Y|) &= \int\int |x - y| f(x, y) dx dy \\
&= \frac{1}{L^2}\int_0^L\int_0^L |x - y| dx dy \\
&= \frac{1}{L^2}\int_0^L\int_0^x (x - y)\, dy dx + \frac{1}{L^2}\int_0^L\int_x^L (y - x)\, dy dx \\
&= \frac{1}{L^2}\int_0^L xy - \frac{y^2}{2}\Big|_0^x dx + \frac{1}{L^2}\int_0^L \frac{y^2}{2} - xy\Big|_x^L dx \\
&= \frac{1}{L^2}\int_0^L \frac{x^2}{2} dx + \frac{1}{L^2}\int_0^L \frac{L^2}{2} - xL + \frac{x^2}{2} dx \\
&= \frac{L}{3}
\end{aligned}$$

Let X and Y be continuous random variables with density $f(x, y)$.
$g(X, Y) = X + Y$.

$$
\begin{aligned}
&E[g(x, y)] = E(X + Y) \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x + y)f(x, y)dxdy \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x, y)dxdy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf(x, y)dxdy \\
=&\int_{-\infty}^{\infty}x\left[\int_{-\infty}^{\infty}f(x, y)dy\right]dx + \int_{-\infty}^{\infty}y\left[\int_{-\infty}^{\infty}f(x, y)dx\right]dy \\
=&\int_{-\infty}^{\infty}xf_X(x)dx + \int_{-\infty}^{\infty}yf_Y(y)dy \\
=&E(X) + E(Y)
\end{aligned}
$$

Linearity of Expectations

$$E(X+Y) = E(X) + E(Y)$$

$$E(X+Y+Z) = E((X+Y)+Z)$$
$$=E(X+Y) + E(Z) = E(X) + E(Y) + E(Z)$$

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$$

Linearity is not affected by correlations among X_i's.

No assumption of independence is needed.

The Sample Mean

Let X_1, X_2, ..., X_n be identically distributed random variables having expected value μ. The quantity $\bar{X}$, defined by

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is called the sample mean.

$$
\begin{aligned}
E(\bar{X}) =& E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
=& \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] \\
=& \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\
=& \frac{1}{n}\sum_{i=1}^{n} \mu \\
=& \frac{1}{n} n\mu \\
=& \mu
\end{aligned}
$$

When μ is unknown, an unbiased estimator of μ is $\bar{X}$.

Expectation of a Binomial Random Variable

$X \sim Binomial(n,p)$

$$X = Z_1 + Z_2 + ... + Z_n$$

where Z_i, $i = 1$ to n, are independent Bernoulli with probability of success p.

Because $E(Z_i) = p$,

$$E(X) = E\left(\sum_{i=1}^{n} Z_i\right) = \sum_{i=1}^{n} E(Z_i) = np$$

Expectation of a Hypergeometric Random Variable

Suppose n balls are randomly selected without replacement, from an urn containing N balls of which m are white. Denote $X \sim HG(N, m, n)$.

Q: Find $E(X)$.

Solution 1: Let $X = X_1 + X_2 + ... + X_m$, where

$$X_i = \begin{cases} 1 & \text{if the } i\text{th white ball is selected} \\ 0 & \text{otherwise} \end{cases}$$

Note that the X_i's are not independent; but we only need the expectation.

Solution 1: Let $X = X_1 + X_2 + ... + X_m$, where

$$X_i = \begin{cases} 1 & \text{if the } i\text{th white ball is selected} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(X_i) =& P\{X_i = 1\} = P\{i\text{th white ball is selected}\} \\ =& \frac{\binom{1}{1}\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N} \end{aligned}$$

Therefore,

$$E(X) = \frac{mn}{N}.$$

Solution 2: Let $X = Y_1 + Y_2 + ... + Y_n$, where

$$Y_i = \begin{cases} 1 & \text{if the } i\text{th ball selected is white} \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(Y_i) =& P\{Y_i = 1\} = P\{i\text{th ball elected is white}\} \\ =& \frac{m}{N} \end{aligned}$$

Therefore,

$$E(X) = \frac{mn}{N}.$$

Expected Number of Matches

A group of N people throw their hats into the center of a room. The hats are mixed up and each person randomly selects one.

Q: Find the expected number of people that selects their own hat.

Solution: Let X denote the number of matches and let

$$X = X_1 + X_2 + ... + X_N,$$

where

$$X_i = \begin{cases} 1 & \text{if the } i \text{ th person selects his own hat} \\ 0 & \text{otherwise} \end{cases}$$

$$E(X_i) = P\{X_i = 1\} = \frac{1}{N}$$

Thus

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

Expectations May Be Undefined

A More Rigorous Definition: If X is a discrete random variable with mass function $p(x)$, then

$$E(X) = \sum_x xp(x), \qquad \text{provided } \sum_x |x|p(x) < \infty$$

If X is a continuous random variable with density $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx, \qquad \text{provided } \int_{-\infty}^{\infty} |x|f(x) < \infty$$

Example (Cauchy Distribution):

$$f(x) = \frac{1}{\pi}\frac{1}{x^2+1}, \qquad -\infty < x < \infty$$

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi}\frac{1}{x^2+1}dx = 0 \ \ ??? \qquad \text{(Answer is NO)}$$

because

$$\int_{-\infty}^{\infty} \frac{|x|}{\pi}\frac{1}{x^2+1}dx = \int_{0}^{\infty} \frac{2x}{\pi}\frac{1}{x^2+1}dx$$

$$= \int_{0}^{\infty} \frac{1}{\pi}\frac{1}{x^2+1}d[x^2+1]$$

$$= \frac{1}{\pi}\log(x^2+1)\Big|_0^{\infty}$$

$$= \frac{1}{\pi}\log\infty = \infty$$

Section 7.3: Moments of The Number of Events

$$I_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

$X = \sum_{i=1}^n I_i$: number of these events A_i that occur. We know

$$E(X) = \sum_{i=1}^n E(I_i) = \sum_{i=1}^n P(A_i)$$

Suppose we are interested in the number of pairs of events that occur, eg. A_iA_j

$X = \sum_{i=1}^{n} I_i$: number of these events A_i that occur.

$\frac{X(X-1)}{2} = \binom{X}{2}$: number of pairs of event that occur.

$\Longleftrightarrow$

$\sum_{i<j} I_i I_j$: number of pairs of event that occur, because

$$I_i I_j = \begin{cases} 1 & \text{if both } A_i \text{ and } A_j \text{ occur} \\ 0 & \text{otherwise} \end{cases}$$

Therefore,

$$E\left(\frac{X(X-1)}{2}\right) = \sum_{i<j} P\left(A_i A_j\right)$$

$$E\left(X^2\right) - E\left(X\right) = 2\sum_{i<j} P\left(A_i A_j\right)$$

Second Moments of a Binomial Random Variables

n independent trials, with each trial being a success with probability p.

A_i = the event that trial i is a success.

If $i \neq j$, then $P(A_iA_j) = p^2$.

$$E\left(X^2\right) - E\left(X\right) = 2\sum_{i<j} P\left(A_iA_j\right) = 2\binom{n}{2}p^2 = n(n-1)p^2$$

$$E(X^2) = n^2p^2 - np^2 + np$$

$$Var(X) = \left[n^2p^2 - np^2 + np\right] - \left[np\right]^2 = np(1-p)$$

Second Moments of a Hypergeometric Random Variables

n balls are randomly selected without replacement from an urn containing N balls, of which m are white.

A_i = event that the ith ball selected is white.

X = number of white balls selected. We have known

$$E(X) = \sum_{i=1}^{n} P(A_i) = n\frac{m}{N}.$$

$$P(A_iA_j) = P(A_i)P(A_j|A_i) = \frac{m}{N}\frac{m-1}{N-1}, \qquad i \neq j$$

$$E\left(X^2\right) - E\left(X\right) = 2\sum_{i<j} P\left(A_iA_j\right) = 2\binom{n}{2}\frac{m}{N}\frac{m-1}{N-1}$$

$$E(X^2) = n(n-1)\frac{m}{N}\frac{m-1}{N-1} + \frac{nm}{N}$$

$$Var(X) = \frac{nm(N-n)(N-m)}{N^2(N-1)} = n\left(\frac{m}{N}\right)\left(1-\frac{m}{N}\right)\frac{N-n}{N-1}$$

One can view $\frac{m}{N} = p$. Note that $\frac{N-n}{N-1} \approx 1$ if N is much larger than n.

Second Moments of the Matching Problem

Let A_i = event that person i selects his/her own hat in the match problem, $i = 1$ to N.

$$P(A_iA_j) = P(A_i)P(A_j|A_i) = \frac{1}{N}\frac{1}{N-1}, \qquad i \neq j$$

X = number of people who select their own hat. We have shown $E(X) = 1$.

$$E(X^2) - E(X) = 2\sum_{i<j} P(A_iA_j) = N(N-1)\frac{1}{N(N-1)} = 1$$

$$Var(X) = E(X^2) - E^2(X) = 1$$

Section 7.4: Covariance and Correlations

Proposition 4.1: If X and Y are independent, then

$$E(XY) = E(X)E(Y).$$

For any function h and g,

$$E[h(X)\,g(Y)] = E[h(X)]\,E[g(Y)].$$

Proof:

$$
\begin{aligned}
E[h(X)\,g(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\
&= \left[\int_{-\infty}^{\infty} g(x)f_X(x)dx\right]\left[\int_{-\infty}^{\infty} h(y)f_Y(y)dy\right] \\
&= E[h(X)]\,E[g(Y)]
\end{aligned}
$$

Covariance

Definition: The covariance between X and Y, is defined by

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

Equivalently

$$Cov(X,Y) = E(XY) - E(X)E(Y)$$

Proof:

$$\begin{aligned} Cov(X,Y) &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY - E[X]Y - E[Y]X + E[X]E[Y]] \\ &= E[XY] - E[E[X]Y] - E[E[Y]X] + E[E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Note that $E(X)$ is a constant, so is $E(X)E(Y)$

X and Y are independent $\Longrightarrow Cov(X,Y) = E(XY) - E(X)E(Y) = 0$.

However, that X and Y are uncorrelated $Cov(X,Y) = 0$,
does not imply that X and Y are independent

Properties of Covariance: Proposition 4.2

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,X) = Var(X)$
- $Cov(aX,Y) = aCov(X,Y)$

Directly follow from the definition $Cov(X,Y) = E(XY) - E(X)E(Y)$.

- $$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

Proof:

$$
\begin{aligned}
Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) =& E\left(\sum_{i=1}^n X_i \sum_{j=1}^m Y_j\right) - E\left(\sum_{i=1}^n X_i\right) E\left(\sum_{j=1}^m Y_j\right)\\
=& E\left(\sum_{i=1}^n \sum_{j=1}^m X_i Y_j\right) - \sum_{i=1}^n E(X_i) \sum_{j=1}^m E(Y_j)\\
=& \sum_{i=1}^n \sum_{j=1}^m E(X_i Y_j) - \sum_{i=1}^n \sum_{j=1}^m E(X_i) E(Y_j)\\
=& \sum_{i=1}^n \sum_{j=1}^m \left[E(X_i Y_j) - E(X_i) E(Y_j)\right]\\
=& \sum_{i=1}^n \sum_{j=1}^m Cov(X_i, Y_j)
\end{aligned}
$$

$$Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$$

If $n = m$ and $X_i = Y_i$, then

$$\begin{aligned}
Var\left(\sum_{i=1}^{n} X_i\right) =& Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)\\
=&\sum_{i=1}^{n} Cov(X_i, X_i) + \sum_{i\neq j} Cov(X_i, X_j)\\
=&\sum_{i=1}^{n} Var(X_i) + \sum_{i\neq j} Cov(X_i, X_j)\\
=&\sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)
\end{aligned}$$

If X_i's are pairwise independent, then

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Example 4b: $X \sim Binomial(n,p)$ can be written as

$$X = X_1 + X_2 + ... + X_n,$$

where $X_i \sim Bernoulli(p)$. X_i's are independent. Therefore

$$\begin{aligned} Var(X) =& Var\left(\sum_{i=1}^{n} X_i\right) \\ =& \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p(1-p) = np(1-p). \end{aligned}$$

Example 4a: Let X_1, X_2, ..., X_n be independent and identically distributed random variables having expected value μ and variance σ^2. Let $\bar{X}$ be the sample mean and S^2 be the sample variance, where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Q: Find $Var(\bar{X})$ and $E(S^2)$.

Solution:

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

We've already shown $E(\bar{X}) = \mu$. As the sample size n increases, the variance of $\bar{X}$, the unbiased estimator of μ, decreases.

The Less Clever Approach (different from the book)

$$\begin{aligned}
E(S^2) =& E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)\\
=&\frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i^2+\bar{X}^2-2\bar{X}X_i\right)\right)\\
=&\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2+n\bar{X}^2-2\bar{X}\sum_{i=1}^{n}X_i\right)\\
=&\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)\\
=&\frac{1}{n-1}\left(\sum_{i=1}^{n}E(X_i^2)-nE\left(\bar{X}^2\right)\right)\\
=&\frac{n}{n-1}\left(\mu^2+\sigma^2-E\left(\bar{X}^2\right)\right)
\end{aligned}$$

$$
\begin{aligned}
E(\bar{X}^2) =& \frac{1}{n^2} E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] = \frac{1}{n^2} E\left[\sum_{i}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\right] \\
=& \frac{1}{n^2}\left[\sum_{i}^{n} E(X_i^2) + 2\sum_{i<j} E(X_i X_j)\right] \\
=& \frac{1}{n^2}\left[\sum_{i}^{n} (\mu^2+\sigma^2) + 2\sum_{i<j} \mu^2\right] \\
=& \frac{1}{n^2}\left[n\left(\mu^2+\sigma^2\right) + n(n-1)\mu^2\right] = \frac{\sigma^2}{n} + \mu^2
\end{aligned}
$$

Or, directly,

$$
E(\bar{X}^2) = Var\left(\bar{X}\right) + E^2\left(\bar{X}\right) = \frac{\sigma^2}{n} + \mu^2
$$

Therefore,

$$
\begin{aligned}
E(S^2) =& \frac{n}{n-1}\left(\mu^2+\sigma^2-E\left(\bar{X}^2\right)\right)\\
=& \frac{n}{n-1}\left(\mu^2+\sigma^2-\frac{\sigma^2}{n}-\mu^2\right)\\
=& \sigma^2
\end{aligned}
$$

Sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ is an unbiased estimator of σ^2.

Why $\frac{1}{n-1}$ instead of $\frac{1}{n}$?

$\sum_{i=1}^n (X_i - \bar{X})^2$ is a sum of n terms with only $n-1$ degrees of freedom, because $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

The Standard Clever Technique for evaluating $E(S^2)$ (as in the textbook)

$$\begin{aligned}
(n-1)S^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{n}(\bar{X} - \mu)^2 - 2\sum_{i=1}^{n}(X_i - \mu)(\bar{X} - \mu) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2
\end{aligned}$$

$$\begin{aligned}
E[(n-1)S^2] =&(n-1)E(S^2) = E\left[\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\
=&nVar(X_i) - nVar(\bar{X}) = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2
\end{aligned}$$

The Correlation of Two Random Variables

Definition: The correlation of two random variables X and Y, denoted by $\rho(X,Y)$, is defined, as long as $Var(X) > 0$ and $Var(Y) > 0$, by

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

It can be shown that

$$-1 \leq \rho(X,Y) \leq 1$$

$$-1 \leq \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \leq 1$$

Proof: The basic trick: Variance $Var(X) \geq 0$ always.

$$\begin{aligned}
&Var\left(\frac{X}{\sqrt{Var(X)}} + \frac{Y}{\sqrt{Var(Y)}}\right)\\
=&\frac{Var(X)}{Var(X)} + \frac{Var(Y)}{Var(Y)} + \frac{2Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}\\
=&2\left(1+\rho(X,Y)\right) \geq 0
\end{aligned}$$

Therefore, $\rho(X,Y) \geq -1$.

$$
\begin{aligned}
&Var\left(\frac{X}{\sqrt{Var(X)}} - \frac{Y}{\sqrt{Var(Y)}}\right)\\
=&\frac{Var(X)}{Var(X)} + \frac{Var(Y)}{Var(Y)} - \frac{2Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}\\
=&2\left(1-\rho(X,Y)\right) \geq 0
\end{aligned}
$$

Therefore, $\rho(X,Y) \leq 1$.

- $\rho(X,Y) = +1 \Longrightarrow Y = a + bX$, and $b > 0$.
- $\rho(X,Y) = -1 \Longrightarrow Y = a - bX$, and $b > 0$.
- $\rho(X,Y) = 0 \Longrightarrow X$ and Y are uncorrelated
- $\rho(X,Y)$ is close to $+1 \Longrightarrow X$ and Y are highly positively correlated.
- $\rho(X,Y)$ is close to $-1 \Longrightarrow X$ and Y are highly negatively correlated.

Section 7.5: Conditional Expectation

If X and Y are jointly discrete random variables

$$E[X|Y=y] = \sum_x xP\{X=x|Y=y\}$$
$$= \sum_x xp_{X|Y}(x|y)$$

If X and Y are jointly continuous random variables

$$E[X|Y=y] = \int_{-\infty}^{\infty} xf_{X|Y}(x|y)dx$$

Example 5b: Suppose the joint density of X and Y is given by

$$f(x,y) = \frac{1}{y}e^{-x/y}e^{-y}, \qquad 0 < x, y < \infty$$

Q: Compute

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

What the answer should be like? (a function a y)

Solution:

$$
\begin{aligned}
f_{X|Y}(x|y) =& \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx} \\
=& \frac{\frac{1}{y}e^{-x/y}e^{-y}}{\int_0^{\infty} \frac{1}{y}e^{-x/y}e^{-y}dx} \\
=& \frac{\frac{1}{y}e^{-x/y}e^{-y}}{e^{-x/y}e^{-y}\Big|_{\infty}^{0}} \\
=& \frac{\frac{1}{y}e^{-x/y}e^{-y}}{e^{-y}} \\
=& \frac{1}{y}e^{-x/y}, \qquad y > 0
\end{aligned}
$$

$$
E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = \int_0^{\infty} \frac{x}{y}e^{-x/y}dx = y
$$

Computing Expectations by Conditioning

Proposition 5.1: The tower property

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

In some cases, computing $E(X)$ is difficult;

but computing $E(X|Y)$ may be much more convenient.

Proof: Assume X and Y are continuous. (Book assumes discrete variables)

$$
\begin{aligned}
E[E[X|Y]] &= \int_{-\infty}^{\infty} E(X|Y) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx\right] f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y) dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \left[f_{X|Y}(x|y) f_Y(y)\right] dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dx dy \\
&= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x,y) dy\right] dx \\
&= \int_{-\infty}^{\infty} x f_X(x) dx = E(X)
\end{aligned}
$$

Example 5d: Suppose N the number of people entering a store on a given day is a random variable with mean 50. Suppose further that the amounts of money spent by these customers are independent random variables having a common mean \$8. Assume the amount of money spent by a customer is also independent of the total number of customers to enter the store.

Q: What is the expected amount of money spent in the store on a given day?

Solution:

N: the number of people entering the store. $E(N) = 50$.

X_i: the amount spent by the ith customer. $E(X_i) = 8$.

X: the total amount of money spent in the store on a give day. $E(X) = ??$

We can not use the linearity directly:

$$E(X) = E\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} E(X_i) = \sum_{i=1}^{N} 8 \quad ???$$

because N, the number of terms in the summation, is also random.

$$\begin{aligned} E(X) =& E\left(\sum_{i=1}^{N} X_i\right) \\ =& E\left[E\left(\sum_{i=1}^{N} X_i | N\right)\right] \\ =& E\left[\sum_{i=1}^{N} E(X_i)\right] \\ =& E(N8) = 8E(N) \\ =& 8 \times 50 = \$400 \end{aligned}$$

Example 2e: A game is begun by rolling an ordinary pair of dice.

- If the sum of the dice is 2, 3, or 12, the player loses.
- If the sum is 7 or 11, the player wins.
- If the sum is any other numbers i,the player continues to roll the dice until the sum is either 7 or i.
  - If it is 7, the player loses;
  - If it is i, the player wins;

Q : Compute $E(R)$, where R= number of rolls in a game.

Conditional on S, the initial sum,

$$E(R) = \sum_{i=2}^{12} E(R|S=i)P(S=i)$$

Let $P(S=i) = P_i$.

$$P_2 = \frac{1}{36}, \quad P_3 = \frac{2}{36}, \quad P_4 = \frac{3}{36},$$
$$P_5 = \frac{4}{36}, \quad P_6 = \frac{5}{36}, \quad P_7 = \frac{6}{36},$$
$$P_8 = \frac{5}{36}, \quad P_9 = \frac{4}{36}, \quad P_{10} = \frac{3}{36},$$
$$P_{11} = \frac{2}{36}, \quad P_{12} = \frac{1}{36}$$

Equivalently,

$$P_i = P_{14-i} = \frac{i-1}{36}, \qquad i = 2, 3, ..., 7$$

Obviously,

$$E(R|S=i) = \begin{cases} 1, & \text{if } i = 2, 3, 7, 11, 12 \\ 1 + E(Z), & \text{otherwise} \end{cases}$$

Z = the number of rolls until the sum is either i or 7.

Note that $i \notin \{2, 3, 7, 11, 12\}$ at this point.

Z is a geometric random variable with success probability = $P_i + P_7$.

Thus $E(Z) = \frac{1}{P_i + P_7}$.

Therefore,

$$E(R) = 1 + \sum_{i=4}^{6} \frac{P_i}{P_i + P_7} + \sum_{i=8}^{10} \frac{P_i}{P_i + P_7} = 3.376$$

Computing Probabilities by Conditioning

If Y is discrete, then

$$P(E) = \sum_y P(E|Y=y)P(Y=y)$$

If Y is continuous, then

$$P(E) = \int_{-\infty}^{\infty} P(E|Y=y)f_Y(y)dy$$

The Best Prize Problem: Deal? No Deal!

We are presented with n distinct prizes in sequence. After being presented with a prize we must decide whether to accept it or to reject it and consider the next prize. The objective is to maximize the probability of obtaining the best prize. Assuming that all $n!$ orderings are equally likely.

Q: How well can we do? What would be a good strategy?

Many related applications: online algorithms, hiring decisions, dating, etc.

A plausible strategy: Fix a number k, $0 \leq k \leq n$. Reject the first k prizes and then accepts the first one that is better than all of those first k.

Let X = the location of the best prize, i.e., $1 \leq X \leq n$.

Let $P_k(\text{best})$ = the best prize is selected using this strategy.

$$P_k(\text{best}) = \sum_{i=1}^{n} P_k(\text{best}|X=i)P(X=i)$$
$$= \frac{1}{n}\sum_{i=1}^{n} P_k(\text{best}|X=i)$$

Obviously

$$P_k(\text{best}|X=i) = 0, \qquad \text{if } i \leq k$$

Assume $i > k$, then

$$\begin{aligned}P_k(\text{best}|X=i) =& P_k(\text{best of the first } i-1 \text{ is among the first } k|X=i)\\ =&\frac{k}{i-1}\end{aligned}$$

Therefore,

$$\begin{aligned}P_k(\text{best}) =& \frac{1}{n}\sum_{i=1}^{n} P_k(\text{best}|X=i)\\ =&\frac{k}{n}\sum_{i=k+1}^{n}\frac{1}{i-1}\end{aligned}$$

Next question: which k to use? We choose the k that maximizes P_k(best).

Exhaustive search is the best thing to do, which only involves a linear scan.

However, it is often useful to have an "analytical" (albeit approximate) expression.

A useful approximate formula

$$\sum_{i=1}^{n} \frac{1}{i} = 1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n} \approx \log n + \gamma$$

where Euler's constant

$\gamma = 0.57721566490153286060651209008240243104215933593992...$

The approximation is asymptotically (as $n \to \infty$) exact.

$$
\begin{aligned}
P_k(\text{best}) =& \frac{k}{n}\sum_{i=k+1}^{n}\frac{1}{i-1}\\
=& \frac{k}{n}\left(\frac{1}{k}+\frac{1}{k+1}+...+\frac{1}{n-1}\right)\\
=& \frac{k}{n}\left(\sum_{i=1}^{n-1}\frac{1}{i}-\sum_{i=1}^{k-1}\frac{1}{i}\right)\\
\approx& \frac{k}{n}\left(\log(n-1)-\log(k-1)\right)\\
=& \frac{k}{n}\log\frac{n-1}{k-1}
\end{aligned}
$$

$$P_k(\text{best}) \approx \frac{k}{n}\log\frac{n-1}{k-1} \approx \frac{k}{n}\log\frac{n}{k}$$

$$[P_k(\text{best})]' \approx \frac{1}{n}\log\frac{n}{k} - \frac{k}{n}\frac{1}{k} = 0 \Longrightarrow \log\frac{n}{k} = 1 \Longrightarrow k = \frac{n}{e}$$

Also, when $k = \frac{n}{e}$, $P_k(\text{best}) \approx \frac{n}{n}\frac{1}{e} \approx 0.36788$, as $n \to \infty$.

Conditional Variance

Proposition 5.2:

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

Proof: (from the opposite direction, compared to the book)

$$\begin{aligned}
& E[Var(X|Y)] + Var(E[X|Y]) \\
=& E\left(E(X^2|Y) - E^2(X|Y)\right) + \left(E(E^2(X|Y)) - E^2(E[X|Y])\right) \\
=& \left(E(E(X^2|Y)) - E^2(E[X|Y])\right) + \left(-E(E^2(X|Y)) + E(E^2(X|Y))\right) \\
=& E(X^2) - E^2(X) \\
=& Var(X)
\end{aligned}$$

Example 5o: Let X_1, X_2, ..., X_N be a sequence of independent and identically distributed random variables and let N be a non-negative integer-valued random variable independent of X_i.

Q: Compute $Var\left(\sum_{i=1}^N X_i\right)$.

Solution:

$$\begin{aligned}
Var\left(\sum_{i=1}^N X_i\right) =& E\left[Var\left(\sum_{i=1}^N X_i|N\right)\right] + Var\left(E\left[\sum_{i=1}^N X_i|N\right]\right)\\
=& E\left(Var(X)N\right) + Var(NE(X))\\
=& Var(X)E(N) + E^2(X)Var(N)
\end{aligned}$$