MATH 323 Lecture Notes

Prof. Masoud Asgharian-Dastenaei

Fall 2013

Contents

1 Lec. 1 - Wed. Sept. 4, 2013 **3**
1.1 Review of Set Notation 3
1.2 Laws 4
1.3 Some Definitions 5

2 Lec. 2 - Mon. Sept. 9, 2013 **5**
2.0.1 Preliminary Discussion About Relative Frequency 5
2.1 Axioms of Probability 6
2.2 Tools for Counting 7
2.2.1 The mn Rule 7
2.3 Exercise 2.7, p. 42 from the textbook 8

3 Lec. 3 - Wed. Sept. 11, 2013 **9**

4 Lec. 4 - Mon. Sept. 16, 2013 **9**
4.0.1 Elevator Question 9
4.1 Conditional Probability and Independence 9

5 Lec. 5 - Wed. Sept. 18, 2013 **11**
5.1 Tools for Probability 11
5.1.1 Multiplicative Law of Probability 11
5.1.2 Additive Law of Probability 12
5.1.3 Exercise 2.97, p. 60 from the textbook 13

6 Lec. 6 - Mon. Sept. 23, 2013 **14**
6.1 Bayes' Rule (or Theorem) 15
6.1.1 Exercise 2.124, p. 73 from the textbook 16
6.2 Chapter 3: Discrete Random Variables (RVs) 16

7 Lec. 7 - Wed. Sept. 25, 2013 **18**
7.0.1 Exercise 3.4 from the textbook 18
7.1 Expected Value of a RV or of a Function of a RV 19

8 Lec. 8 - Mon. Sept. 30, 2013 **21**

9 Lec. 9 - Wed. Oct. 2, 2013 **27**
9.1 Geometric Distribution . . . 27
9.2 Negative Binomial Distribution . . . 30
9.3 Hypergeometric Distribution . . . 31

10 Lec. 10 - Mon. Oct. 7, 2013 **31**
10.1 Hypergeometric Distribution . . . 31
10.1.1 p. 130 # 3.113 Jury of 6 person etc. . . . 32
10.2 Poisson Distribution . . . 33

11 Lec. 11 - Wed. Oct. 9, 2013 **33**
11.1 Poisson Distribution . . . 33
11.1.1 p. 136 # 3.127 . . . 36
11.1.2 p. 137 #3.140 . . . 36
11.2 Moments and Moment Generating Functions . . . 37

12 Lec. 12 - Wed. Oct. 16, 2013 **38**
12.1 Chebychev's Theorem . . . 38
12.2 Chapter 4: Continuous Random Variables and Their Products . . . 38
12.2.1 Probability Distribution Function . . . 38

13 Lec. 13 - Mon. Oct. 21 **40**
13.1 2. Expected Values for Continuous Random Variables . . . 40
13.2 3. Uniform Distribution . . . 42

14 Lec. 14 - Wed. Oct. 23 **42**

15 Lec. 15 - Mon. Oct. 28, 2013 **43**
15.1 4. Normal Distribution . . . 43
15.2 5. The Gamma Probability Distribution . . . 45

16 Lec. 16 - Wed. Oct. 30, 2013 **46**
16.1 Special cases . . . 48

17 Lec. 17 - Mon. Nov. 4, 2013 **48**
17.1 Section 6: The Beta Probability Distribution . . . 48
17.2 Section 7: Tchebysheff's Theorem . . . 49
17.3 Chapter 5: Multivariate Probability Distributions . . . 50
17.3.1 Section 1: Joint (Bivariate) Distributions . . . 50

18 Lec. 18 - Wed. Nov. 6, 2013 **51**
18.0.2 Joint Continuity and Separate Continuity . . . 53
18.1 Section 2: Marginal and Conditional Probability Distribution . . . 53

19 Lec. 19 - Mon. Nov. 11, 2013 **54**
19.1 Section 3. Independent Random Variables . . . 55
19.2 Section 4. The Expected Value of a Function of Random Variables . . . 56

20 Lec. 20 - Wed. Nov. 13, 2013 **56**
20.1 Section 5. The covariance of two random variables 56

21 Lec. 21 - Mon. Nov. 18, 2013 **58**
21.1 6. The Multinomial Probability Distribution 60

22 Lec. 22 - Wed. Nov. 20, 2013 **61**
22.1 7. Conditional Expectation 61
22.2 Chapter 6: Functions of Random Variables 62

23 Lec. 23 - Mon. Nov. 25, 2013 **63**
23.1 Summary of the Distribution Function Method 66

24 Lec. 24 - Wed. Nov. 27 **66**
24.1 2. The Method of Transformation 66
24.2 3. The Moment of Generating Functions 67

25 Lec. 25 - Mon. Dec. 2, 2013 **70**
25.1 Chapter 7: Sampling Distributions and the Central Limit Theorem (CLT) 70
25.1.1 1. Sampling Distribution 70

26 Lec. 25 - Tues. Dec. 3, 2013 **72**
26.1 Central Limit Theorem (CLT) 72
26.2 3. Application to Binomial Distribution 73

27 Glossary **75**

§1 Lec. 1 - Wed. Sept. 4, 2013

We collect samples from a population and use inductive logic to infer conclusions about the population. These samples must be *random* to well-represent the population.

Probability kicks in about the credibility of these conclusions.

§1.1 Review of Set Notation

Definition 1.1 A set is a collection of items, denoted by a capital letter

Examples: $\mathscr{A} = \{TV, duck, computer\}$ $\quad \mathscr{B} = \{1, 3, 5\}$

Notation. $\mathscr{A} \subseteq \mathscr{B} \quad \Leftrightarrow \quad$ *every element in $\mathscr{A}$ is also in $\mathscr{B}$*

$\mathscr{A} \cup \mathscr{B}$ denotes the set of all elements in $\mathscr{A}$, $\mathscr{B}$, or both.
$\mathscr{A} \cap \mathscr{B}$ denotes the set of all elements in both A and $\mathscr{B}$.

Eg. If $\mathscr{A} = \{1, 2, 3\}$ and $\mathscr{B} = \{x \in \mathbb{R} | x \leq 3\}$, then $\mathscr{A} \subseteq \mathscr{B}$ and $\mathscr{A} \cap \mathscr{B} = \{3\}$.

The null set is a set consisting of no elements and is denoted by $\varnothing$ or $\{\}$.

Suppose $\mathcal{S}$ is the set of all elements under consideration and $A \subset \mathcal{S}$.
Then $\mathcal{A}^c$ (or $\bar{\mathcal{A}}$) is the set of all elements in $\mathcal{S}$ which are not in $\mathcal{A}$.

To show that two sets are equal, you must show that they are subsets of each other.
This tells you that they have the exact same elements.

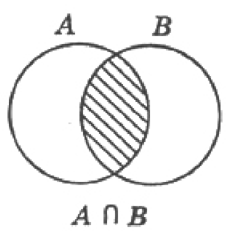
Definition 1.2 Set $\mathcal{A}$ and Set $\mathcal{B}$ are called disjoint or mutually exclusive if their intersection is empty i.e. $\mathcal{A} \cap \mathcal{B} = \varnothing$.

§1.2 Laws

1. $\mathcal{A} \cup \mathcal{B} = \mathcal{B} \cup \mathcal{A}$ — Commutative law
2. $\mathcal{A} \cap (\mathcal{B} \cup \mathcal{C}) = (\mathcal{A} \cap \mathcal{B}) \cup (\mathcal{A} \cap \mathcal{C})$ — Distributive law of intersection over union
3. $\mathcal{A} \cup (\mathcal{B} \cap \mathcal{C}) = (\mathcal{A} \cup \mathcal{B}) \cap (\mathcal{A} \cup \mathcal{C})$ — Distributive law of union over intersection

Theorem 1.1 *De Morgan's Law*

$$\overline{\mathcal{A} \cap \mathcal{B}} = \bar{\mathcal{A}} \cup \bar{\mathcal{B}} \qquad \overline{\mathcal{A} \cup \mathcal{B}} = \bar{\mathcal{A}} \cap \bar{\mathcal{B}}$$



To prove De Morgan's law, we observe the following LHS to RHS reasoning:

$$\begin{aligned} x \in \overline{(\mathcal{A} \cap \mathcal{B})} \quad &\Rightarrow \quad x \notin (\mathcal{A} \cap \mathcal{B}) \quad \Rightarrow \quad x \notin \mathcal{A}, x \notin \mathcal{B} \quad \Rightarrow \quad x \in \bar{\mathcal{A}} \textbf{ or } x \in \bar{\mathcal{B}} \\ &\Rightarrow \quad x \in \left(\bar{\mathcal{A}} \cup \bar{\mathcal{B}}\right) \\ \therefore \overline{(\mathcal{A} \cap \mathcal{B})} &\subseteq \bar{\mathcal{A}} \cup \bar{\mathcal{B}} \end{aligned}$$

Then we establish RHS to LHS reasoning to get equality.

§1.3 Some Definitions

Definition 1.3 An experiment is a process by which an observation is made

Eg. measure of IQ, lifetime of lightbulbs

Experiments can result in one or more outcomes called *events*.

Eg. Flipping a coin yields 2 outcomes, H or T.

We say an experiment is *random* if the outcome is uncertain.

The set of all possible outcomes of an experiment is called the sample space.

Eg. $\{1, 2, 3, 4, 5, 6\} \rightarrow$ for rolling a die

To differentiate between a single outcome and a set of possible outcomes, we use the terms *simple* and *compound*.

Eg. In rolling a die, $\{3\}$ is a simple event, $\{1, 3, 5\}$ is a compound event

Definition 1.4 A set $\mathcal{S}$ is countable if there exists a map from $\mathcal{S}$ to the natural numbers $\mathbb{N}$

Eg. $\mathbb{N} = \{1, 2, 3, ...\}$ is countable

$\mathbb{Q} = \{rational\ \ numbers\}$ is countable

$\mathbb{R}$ is **not** countable because there is no 1-to-1 correspondence from $\mathbb{N}$ to $\mathbb{R}$

If the sample space is finite and countable, then it is discrete.

§2 Lec. 2 - Mon. Sept. 9, 2013

Probability is the study of quantifying uncertainty.

§2.0.1 Preliminary Discussion About Relative Frequency

Suppose you flip a coin. The probability of it landing H or T is 50 %.

How do we know? Toss a coin 1000 times; the more times we toss, the closer the ratio of number of H to total tosses approaches that 50%. How close this *relative frequency* of H gets to 50% depends on and is indexed to the number of times you flip the coin.

Relative frequency is always non-negative.

Definition 2.1 We say $\lim_{n\to\infty} a_n = a$ if for any $\varepsilon > 0$, $\exists\ N_\varepsilon$ such that for $n > N_\varepsilon$, we have $|a_n - a| < \varepsilon$.

Even if a sequence converges, it may fluctuate. But in the case of counting the number of heads that show up on a coin with more and more flips, the chances of fluctuating away from the target gets smaller and smaller.

For the probability operation, the input is a set into a function $P()$ and the output is a number i.e. we assign non-negative probabilities to sets.

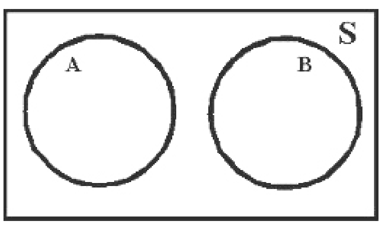
§2.1 Axioms of Probability

1. $P(A) \geq 0$ for any event A.

2. Let $\mathcal{S}$ denote the sample space. Then $P(\mathcal{S}) = 1$.

3. If $A_1, A_2, A_3, ... \subseteq \mathcal{S}$ is a sequence of pairwise mutually exclusive events (i.e. $A_i \cap A_j = \varnothing$ whenever $i \neq j$), then:

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \qquad \text{"Countable additivity"}$$



If we choose a point in the sample space $\mathcal{S}$, the chance of falling in A is $\frac{P(A)}{P(\mathcal{S})}$, and same for B. The chance of falling in A or B is simply $\frac{P(A)+P(B)}{P(\mathcal{S})}$ since they are mutually exclusive.

A probabilitistic model $(\mathcal{S}, \mathcal{P})$ consists of the sample space $\mathcal{S}$ associated with the experiment and the probability of each event in $\mathcal{S}$.

Exercise 2.9 and 2.16, p. 32-33 from the textbook

$$
\begin{aligned}
&\mathcal{S} = \left\{A^+, A^-, B^+, B^-, AB^+, AB^-, O^+, O^-\right\} \\
&\text{Part(a)}: \qquad P(O^+) = \frac{1}{3} \\
&\text{Part(b)}: \qquad P(O^+ \cup O^+) \underset{exclusive}{\overset{mutually}{=}} P(O^+) + P(O^-) = \frac{1}{3} + \frac{1}{15} \\
&\text{Part(c)}: \qquad P(A^+ \cup A^+) \underset{exclusive}{\overset{mutually}{=}} P(A^+) + P(A^-) = \frac{1}{3} + \frac{1}{16} \\
&\text{Part(d)}: \qquad P(O^+ \cup O^-) - P(A^+ \cup A^-) \\
&\qquad\qquad\quad\ \ \text{Let } C = \left\{O^+ \cup O^-\right\}, A = \left\{A^+ \cup A^-\right\}
\end{aligned}
$$

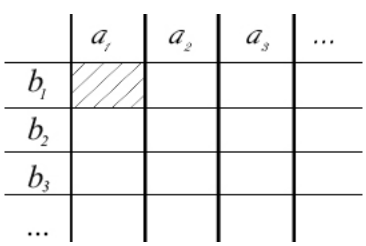
$$
P(\overline{A \cup O}) = 1 - P(A \cup O) = 1 - \{P(A) + P(0)\}
$$

§2.2 Tools for Counting

§2.2.1 The mn Rule

With m elements $a_1, ..., a_m$ and n elements $b_1, ..., b_n$, we can count mn pairs involving one element from each group.

Each cell below represents a pair. The shaded cell is the pair (a_1, b_1).



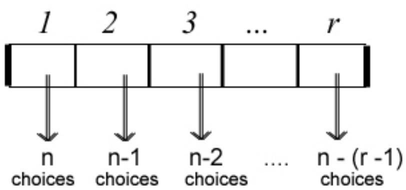
Definition 2.2 An ordered arrangement of r distinct objects is called a permutation.

Notation. *We write P_r^n to denote the number of ways of ordering n distinct objects taking r objects at a time.*

Factorials: $k! = 1 \cdot 2 \cdot 3 \cdots k$
$0! = 1$ *by convention*
$1! = 1$
$2! = 1 \cdot 2$

Theorem 2.1 $P_r^n = n(n-1)(n-2)\cdots(n-r+1) = \dfrac{n!}{(n-r)!}$

Proof. We have r spaces to fill with objects chosen from the n distinct objects.
When order matters, we call this an r-tuple from n objects.
(When order doesn't matter, we simply call it a set of r objects.)



We have n choices for the 1^{st} object. After the 1^{st} object has been selected, we have $n-1$ choices for the 2^{nd} object. After two objects have been selected, there remains $n-2$ choices for the 3^{rd} object, etc.

Hence, we repeatedly apply the mn rule to obtain the number of ways of choosing a r-tuple from n objects:

$$\begin{aligned} n(n-1)(n-2)\cdots(n-(r-1)) &= \frac{n(n-1)(n-2)\cdots(n-r+1)\left[(n-r)(n-r-1)\cdots 2\cdot 1\right]}{(n-r)(n-r-1)\cdots 2\cdot 1} \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

§2.3 Exercise 2.7, p. 42 from the textbook

We begin by calculating the cardinality (or "size") of the sample space.

$\mathcal{S} = \{$set of all $(\boxdot_1, \boxdot_2, \boxdot_3, ..., \boxdot_r)$ s.t. each $\boxdot$ represents a date between Jan 1 and Dec 31$\}$

Assume there are 365 possible choices for each $\boxdot$

$r = 20, n = 365$

$Card\{\mathcal{S}\} = 365^{20}$

If all the outcomes in $\mathcal{S}$ are equiprobable, then for outcome $\{O_i\} \in \mathcal{S}$,

$$1 = P(\mathcal{S}) = P\left(\bigcup_{i=1}^{365^{20}} \{O_i\}\right) = \sum_{i=1}^{365^{20}} P(\{O_i\}) \quad \overset{equiprobability}{\Rightarrow} \quad P(\{O_i\}) = \frac{1}{365^{20}}$$

$\mathcal{A} = \{$set of all $(\boxdot_1, ..., \boxdot_{20})$ s.t. no two dates are the same$\}$

$$P(\mathscr{A}) = P\left(\bigcup_{i=1}^{?} a_i\right) = \sum_{i=1}^{?} P(a_i) = \frac{?}{365^{20}} \qquad a_i \in \mathscr{A}$$

$$? = Card(\mathscr{A})$$

Using Theorem 2.1, we know that: $Card(\mathscr{A}) = P_{20}^{365} \quad \therefore \quad P(\mathscr{A}) = \dfrac{P_{20}^{365}}{365^{20}} = \dfrac{365!}{20!365^{20}}$

§3 Lec. 3 - Wed. Sept. 11, 2013

§4 Lec. 4 - Mon. Sept. 16, 2013

We solved some problems from *A First Course in Probability* by Sheldon Ross. If you do not have the problem and detailed solution, you may contact the prof.

§4.0.1 Elevator Question

(a) The objects are the floors, so n=6, and the procedure is repeated r=8 times.

$$\binom{n+r-1}{r} = \binom{6+8-1}{8} = \binom{13}{8} = \frac{13!}{5!\ 8!}$$
$$= \frac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8!}{8!\ 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13 \cdot 11 \cdot 9 = 1287$$

(b) Similar argument applied to men and women separately:

$$\text{Men: } \binom{n+r_m-1}{r_m} = \binom{6+5-1}{5} = \binom{10}{5} = 252$$

$$\Rightarrow 252 \cdot 56 = 14112$$

$$\text{Women: } \binom{n+r_w-1}{r_w} = \binom{6+3-1}{3} = \binom{8}{3} = 56$$

§4.1 Conditional Probability and Independence

We are interested in the probability of events *given* extra information.

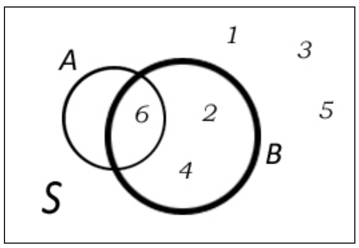
Eg. Roll a die; given that the result is even, what is the probability of getting a 6?
The result also changes when it's equiprobable.

Eg. What are the chances of you going to a party? $\Rightarrow$ Mandy is going $\Rightarrow$ 100%
$\Rightarrow$ Jim is going $\Rightarrow$ 0%
$\Rightarrow$ stranger is going $\Rightarrow$ 50%

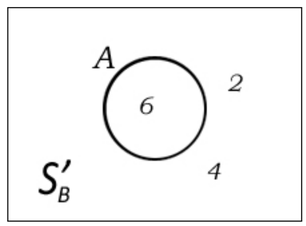
This brings up the issue of independence.

$P(A|B) = P_B(A) =$"probability of A given that B occurs"

Eg. $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ A=$\{6\}$ $B = \{2, 4, 6\}$ $A \cap B = \{6\}$



$\rightarrow$



When we say *given* that the outcome is even, it means we are in a new sample space $\mathcal{S}'_{\mathcal{B}} = \{2, 4, 6\}$.

If B has occured, then A can only occur if $A \cap B \neq \varnothing$.
The greater the P$(A \cap B)$ the more chance of P(A|B):

P(A|B) $\propto$ P$(A \cap B)$ $\Rightarrow$ P(A|B)=γ_B P$(A \cap B)$

where γ_B is the proportionality constant that depends only on B.

If A=B, then P(A|B)=1.
1=P(B|B)= γ_B P$(B \cap B)$=γ_B P(B) $\Rightarrow$ $\gamma_B = \dfrac{1}{P(B)}$.

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

where P(B)>0 since we cannot condition on an event that does not happen.

Definition 4.1 The conditional probability of an event A given that an event B has occurred with P(B)>0 is equal to:

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

Suppose event A is you attending a party and event B is Jim attending the party.

If P(A|B)=P(A), then A and B are independent and Jim going to the party doesn't affect you. Then we have:

$$P\left(A\right) = P\left(A|B\right) = \frac{A \cap B}{P\left(B\right)} \qquad \Rightarrow \qquad P\left(A \cap B\right) = P(A)P(B)$$

Definition 4.2 Events A and B are independent if P(A|B)=P(A), or equivalently,

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

It may not be intuitively obvious, but if A is independent of B, then B is independent of A.

$$\begin{aligned} &P(A|B) = P(A) \quad \Rightarrow \quad \underbrace{P(A \cap B) = P(A)P(B)} \\ &P(B|A) \stackrel{?}{=} P(B) \qquad P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A)P(B)}{P(B)} \end{aligned}$$

§5 Lec. 5 - Wed. Sept. 18, 2013

§5.1 Tools for Probability

§5.1.1 Multiplicative Law of Probability

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

If A and B are independent, then $P(A \cap B) = P(A)P(B)$.
Probability is a study that tries to quantify uncertainty.

Exercise 2.76 p. 56

Plumber A services 40 % of all customers. 10% of customers are not satisfied. Among these complaints, 50% were from plumber A doing the job. What is the probability of a customer being dissatisfied given that plumber A did the job? What is the probability of a customer being satisfied given that A did the job?

Let A={plumber A does the job}, $\quad C = \{\text{customer complains}\}$.

$P(A) = 0.4, P(C) = 0.1, P(A|C) = 0.5.$

We want to find $P(C|A)$ and $P(\bar{C}|A)$.

$$\begin{aligned} P(C|A) &= \frac{P(C\cap A)}{P(A)} \qquad \text{but } P(C\cap A) = P(A|C)P(C) = P(C|A)P(A) \quad \Rightarrow \\ &= \frac{P(A|C)P(C)}{P(A)} = \frac{0.5\cdot 0.1}{0.4} = 0.125 \end{aligned}$$

$$P(\bar{C}|A) = 1 - P(C|A) = 1 - 0.125 = 0.875.$$

Observe that $P(A|C) = 0.5$, but $P(C|A) = 0.125$, i.e. that 87.5% of the customers served by plumber A are satisfied.

We are seeing an example of *Simpson's Paradox*, where we observe the same trend in two groups of data, but that trend disappears when the two groups come together.

In general, $P(A) = 1 - P(\bar{A})$ and $P(A|B) = 1 - P(\bar{A}|B)$. We will prove this later."

§5.1.2 Additive Law of Probability

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

This is essentially the 3rd axiom but for the general case where A and B are not necessarily disjoint. If $P(A\cap B) = 0$, then $P(A\cup B) = P(A) + P(B)$.

Recall the 3rd axiom: $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\inf} P(A_i)$ but *only* if $A_i \cap A_j = \varnothing$ whenever $i \neq j$

Intuitively, or visually, this would be obvious, but we must prove it.

Theorem 5.1 *Suppose $C \subseteq D$. Then*

$$\begin{aligned} P(D\backslash C) &= P\left(\textit{all elements in } D \textit{ that aren't in } C\right) \\ &= P(D\cap C^c) = P(D) - P(C) \end{aligned}$$

Proof.

$$D = D\cap \mathcal{S} = D\cap (C\cup C^c) = \underbrace{(D\cap C)\cup (D\cap C^c)}_{\text{This is the union of two disjoint sets}}$$

Now we apply the 3rd axiom:

$$P(D) = P(D\cap C) + P(D\cap C^c)$$

If $C\subseteq D$, then

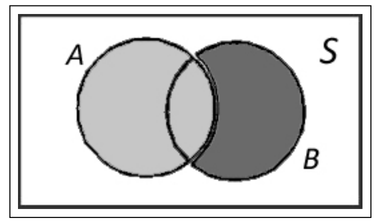
$$P(C) + P(D\backslash C) = P(D)$$

$$\Rightarrow P(D\backslash C) = P(D) - P(C)$$

Corollary 5.2.

$$\begin{aligned} P(A^c) &= 1 - P(A) \textit{for any event } A \\ A^c &= \mathcal{S} \backslash A \\ P(A^c) &= P(\mathcal{S} \backslash A) = P(\mathcal{S}) - P(A) = 1 - P(A) \end{aligned}$$

Theorem 5.3 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



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

Using the 3rd axiom, $P(A \cup B^c) = P(A) + P(B \backslash A)$

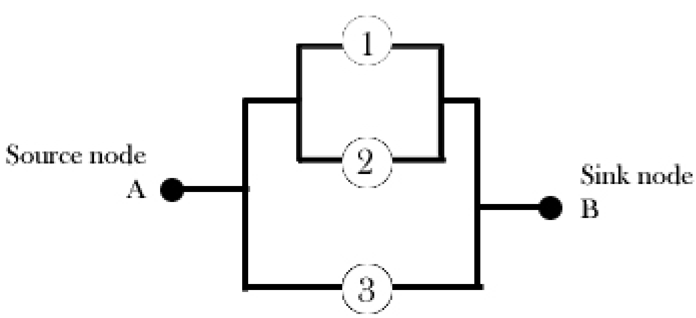
On the RHS, $P(B \backslash A) P(B) - P(A \cup B)$

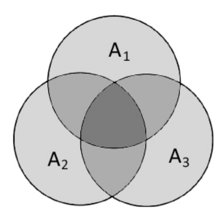
Substituting gives us: $P(A \cup B) = P(A) + P(B) - P(A \cup B)$

§5.1.3 Exercise 2.97, p. 60 from the textbook

$A_i = \{$Relays functions properly$\}$ where i=1,2,3

$D = \{$current flows when the relays are activated$\}$





$$P(D) = P\left(\bigcup_{i=1}^{\infty} A_I\right) = \sum_{i=1}^{3} P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3)$$

(The darkest region was subtracted completely by the 2nd term, so we add it back in)

$$\begin{aligned} P(D) &= 3(0.9) - 3\left[(0.9)(0.9)\right] + (0.9)(0.9)(0.9) \\ &= 0.999 \end{aligned}$$

An alternative way of doing this:

$$P(D) = 1 - P(D^c)$$
$$P(D^c) = P\left\{\bigcup_{i=1}^{3} A_i^c\right\} = P(A_1^c \cap A_2^c \cap A_3^c)) \qquad \text{by De Morgan's Law}$$

Since A_i, $i = 1,2,3$ are independent, so are A_i^c, $i = 1,2,3$. Check that $D^c = A_1^c \cap A_2^c \cap A_3^c$. Therefore,

$$P(D^c) = P(A_1^c \cap A_2^c \cap A_3^c) = P(A_1^c) + P(A_2^c) + P(A_3^c) = (1 - 0.9)^3 = 0.001$$
$$\Rightarrow \quad P(D) = 1 - 0.001 = 0.999$$

$$P(A_1|D) = \frac{A_1 \cap D}{P(D)} = \frac{P(D|A_1)P(A_1)}{P(D)} = \frac{1 \cdot 0.4}{0.999} = 0.9009.$$

Definition 5.1 A collection of sets $\{B_1, B_2, ..., B_k\}$ is called a partition of $\mathcal{S}$ if:

(a) $\mathcal{S} = \bigcup_{i=1}^{k} B_i$ (b) $B_i \cap B_j = \varnothing$ if $i \neq j$.

Theorem 5.4 *Law of Total Probability*

Assume that $\{B_1, ..., B_k\}$ is a partition of $\mathcal{S}$ such that $P(B_i) > 0$, $i = 1, 2, ..., k$. Then for every event A,

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

Proof. Using (a): $A = A \cap \mathcal{S} = A \cap \left(\bigcup_{i=1}^{k} B_i\right) = \bigcup_{i=1}^{k} (A \cap B_i)$

Using (b) and axiom 3: $P(A) = \sum_{i=1}^{k} P(A \cap B_i)$

Using the multiplicative law: $P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$

§6 Lec. 6 - Mon. Sept. 23, 2013

§6.1 Bayes' Rule (or Theorem)

Assume that $\{B_1, ..., B_k\}$ is a partition of $\mathcal{S}$ such that $P(B_i) > 0$ for all $i = 1, 2, ..., k$.

Then $P(B_j|A) = \dfrac{P(A|B_j)P(B_j)}{\sum\limits_{i=1}^{k} P(A|B_i)P(B_i)}$

Proof. $$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{P(A)} = \frac{P(A|B_j)P(B_j)}{\sum\limits_{i=1}^{k} P(A|B_i)P(B_i)}$$

Eg. Famous Prisoner's Dilemma

It is crucial to get the sample space right. When the prisoner didn't ask, there were 3 possibilities. After he asked, there were 4.

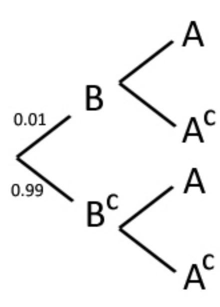
Eg. Blood donation problem

Probabilities are given for when we know if an antibody is present.

The question is: given that the test shows positive, what is the chance of a person having an antibody? $\Rightarrow$ Bayes' Theorem tells us.

$B = \{\text{contaminated blood}\}$ $B^c = \{\text{not contaminated blood}\}$

$A = \{\text{test positive}\}$ $A^c = \{\text{test negative}\}$



$P(A|B) = 0.997$

$P(A^c|B) = 0.003$

$P(A|B^c) = 0.015$

$P(B^c|A^c) = 0.988$

We want to find $P(A)$ and $P(B|A)$.

$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = (0.997)(0.01) + (0.015)(0.99) = 0.2482$

Using this, we get: $P(B|A) = \dfrac{P(A|B)P(B)}{P(A)} = \dfrac{(0.997)(0.01)}{0.02482} = 0.399$

Why is the probability of having antibodies so low given that a the test returns positive?

This is because we are testing for the *uncommon* situation of contaminated blood. Therefore, most positives would actually be false positives.

§6.1.1 Exercise 2.124, p. 73 from the textbook

$$E = \{favouring\} \qquad P(D) = 0.6 \qquad P(R) = 0.4$$
$$P(E|R) = 0. \qquad P(E|D) = 0.7$$
$$\begin{aligned} P(D|E) &\overset{Bayes'}{=} \frac{P(E|D)P(D)}{P(E)} \\ &= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|R)P(R)} = \frac{(0.7)(0.6)}{(0.7)(0.6) + (0.3)(0.4)} = 0.77 \end{aligned}$$

§6.2 Chapter 3: Discrete Random Variables (RVs)

Definition 6.1 A random variable $X : \mathcal{S} \to \mathbb{R}$ is a real valued function whose domain is a sample space $\mathcal{S}$.

Eg. Coin toss $\Rightarrow$ $\mathcal{S} = \{H, T\}$ $\quad X(H) = 1 \quad X(T) = 0$

Roll die $\Rightarrow$ $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$ $\quad X(i) = i \quad i = 1, 2, 3, 4, 5, 6$

More formally, $X : (\mathcal{S}, \mathcal{F}, \mathcal{P}) \to (\mathbb{R}, \mathcal{F}_{\mathbb{R}})$

where $\mathcal{F}$ is a class of subsets of $\mathcal{S}$,

$\mathcal{P}$ is a probability measure,

$\mathcal{F}_{\mathbb{R}}$ is a class of subsets of the real numbers $\mathbb{R}$

Knowing $\mathbb{R}, \mathcal{F}_{\mathbb{R}}, \mathcal{F}$ find $\mathcal{P}$.

Definition 6.2 A RV Y is said to be discrete if it can assume only countably many different values.

Eg. Sizes of families is discrete.

Given $\mathcal{S}$ and $\mathcal{P}$, we try to find the distribution of Y (i.e. how the values of Y are distributed with respect to its probability measure).

If $Y : \mathcal{S} \to \mathbb{R}$, what is $P(Y = y)$ for some $y \in \mathbb{R}$?

Eg. Let X be the random value we get by rolling a die. Then $X(i) = i$ for $i = 1, 2, .., 6$.

$$
\begin{aligned}
&(1) \quad \text{If we roll a fair die,} && P(\{i\}) = \frac{1}{6} \text{ for } i = 1,2,3,4,5,6. \\
&(2) \quad \text{If we roll a loaded die,} && P(\{2i\}) = \frac{1}{4} \text{ for } i = 1,2,3 \\
& && P(\{2i-1\}) = \frac{1}{12} \text{ for } i = 1,2,3.
\end{aligned}
$$

Letting $X : \mathcal{S} = \{1, 2, 3, 4, 5, 6\} \to \mathbb{R}$:

$$
(1) \quad P(X = i) = \frac{1}{6} \text{ for } i = 1, 2, ..., 6.
$$

$$
(2) \quad P(X = i) = \begin{cases} \dfrac{1}{4} & \text{if } i \text{ is even} \\[2ex] \dfrac{1}{12} & \text{if } i \text{ is odd} \end{cases}
$$

We can consider a random variable Y as a way of splitting up the sample space.

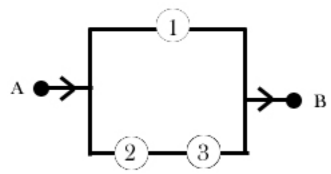
$$
\begin{aligned}
& Y : \mathcal{S} = \{1, 2, ..., 6\} \to \mathbb{R} \\
\text{Let} \quad & Y(2i) = 1, \qquad Y(2i-1) = 0, \qquad \text{for } i = 1,2,3. \\
& A_1 = \{\text{even}\} = \{2,4,6\}, \quad A_0 = \{\text{odd}\} = \{1,3,5\}
\end{aligned}
$$

$$
\text{Then} \quad P(Y = y) = \sum_{s \in A_y} P(s) \qquad \text{where } A_y = \{s \in \mathcal{S} : Y(s) = y\}
$$

$$
\begin{aligned}
(1) \quad & P(Y=1) = P(A_1) = P(\{2,4,6\}) = \sum_{s \in A_1} P(s) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \\
& P(Y=0) = P(A_0) = P(\{1,3,5\}) = \sum_{s \in A_0} P(s) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \\
(2) \quad & P(Y=1) = P(A_1) = \sum_{s \in A_1} P(s) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \\
& P(Y=0) = P(A_0) = \sum_{s \in A_0} P(s) = \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = \frac{1}{4}
\end{aligned}
$$

§7 Lec. 7 - Wed. Sept. 25, 2013

§7.0.1 Exercise 3.4 from the textbook



Let Y be the number of complete paths from A to B.

Y can take 3 values: 0, 1, and 2.

$$\mathcal{S} = \left\{ \underbrace{\overbrace{\{1,2,3\}}^{S_1}}_{\text{"all work"}}, \underbrace{\overbrace{\left\{1,2,\bar{3}\right\}}^{S_2}}_{\text{"3 doesn't work"}}, \left\{1,\bar{2},3\right\}, \overset{etc...}{\left\{1,\bar{2},\bar{3}\right\}}, \left\{\bar{1},2,3\right\}, \left\{\bar{1},\bar{2},3\right\}, \left\{\bar{1},2,\bar{3}\right\}, \overbrace{\left\{\bar{1},\bar{2},\bar{3}\right\}}^{S_8} \right\}$$

$$\begin{aligned} Y: \quad & \mathcal{S} \to \mathbb{R} \\ & Y(S_1) = 2 \\ & Y(S_2) = Y(S_3) = Y(S_4) = Y(S_5) = 1 \\ & Y(S_6) = Y(S_7) = Y(S_8) = 0 \end{aligned}$$

$$\begin{aligned} \text{Let} \quad A_0 &= \left\{s \in \mathcal{S} : Y(s) = 0\right\} = \{S_6, S_7, S_8\} \\ A_1 &= \left\{s \in \mathcal{S} : Y(s) = 1\right\} = \{S_2, S_3, ..., S_5\} \\ A_2 &= \left\{s \in \mathcal{S} : Y(s) = 2\right\} = \{S_1\} \\ & v \end{aligned}$$

$$\begin{aligned}
P(Y=0) &= \sum_{s\in A_0} P(s) \\
&= P(S_6)+P(S_7)+P(S_8) \\
&= P(\left\{\bar{1},2,\bar{3}\right\},\left\{\bar{1},\bar{2},3\right\},\left\{\bar{1},\bar{2},\bar{3}\right\}) \\
&= (0.2)(0.8)(0.2)+(0.2)(0.2)(0.8)+(0.2)(0.2)(0.2) \quad \text{by independence} \\
&= 0.072 \\
P(Y=2) &= P(S_1) = (0.8)^3 = 0.512 \\
\Rightarrow \quad P(Y=1) &= \sum_{s\in A_1} P(s) = \sum_{i=2}^{5} P(S_i) \quad \text{(more terms than we like to sum)} \\
&= 1-P(Y=0)-P(Y=1) = 1-(0.072+0.512) = 0.416
\end{aligned}$$

Thus we denote the *probability distribution* of Y as $P_Y(y)$:

$$P_Y(y) = \begin{cases} 0.072 & \text{if y=0} \\ 0.416 & \text{if y=1} \\ 0.512 & \text{if y=2} \end{cases}$$

A probability distribution $p(y)$ must always satisfy two properties:

$$\begin{aligned}
&(1) \quad 0 \le p(y) \le 1 \quad \text{for all } y \\
&(2) \quad \sum_{\text{all } y} p(y) = 1.
\end{aligned}$$

§7.1 $\boxed{\textbf{Expected Value of a RV or of a Function of a RV}}$

If we have n observations $x_1,...,x_n$, then we say the *average* value $\bar{x} = \frac{1}{n}\sum_{i=1}^{\infty} x_i$

Eg. The average of 2, 3, 3, 16, 5, 7, 21, 5, 7, 5 is:

$$\begin{aligned}
\bar{x} &= \frac{1}{10}(2+3+3+16+5+7+21+5+7+5) = \frac{74}{10} \\
\bar{x} &= \frac{1}{10}(2+2\cdot 3+3\cdot 5+2\cdot 7+16+21) \\
&\quad \text{(we only need to keep track of distinct values)} \\
&= \underbrace{\frac{1}{10}}(2)+\underbrace{\frac{2}{10}}(3)+\underbrace{\frac{3}{10}}(5)+\underbrace{\frac{2}{10}}(7)+\underbrace{\frac{1}{10}}(16)+\frac{1}{10}(21)
\end{aligned}$$

these are "weights", or *relative frequencies*

The idea of relative frequency gives us a way of computing an average when there are different probabilities of getting each value.

Definition 7.1 Let Y be a discrete RV with probability function (or distribution) $p(y)$. Then the expected value of Y is:

$$E[Y] = \sum_y yp(y) = \mu_Y \quad \text{sometimes } \mu \text{ only}$$

If Y is a continuous RV, the $\sum$ is replaced with $\int$.

Theorem 7.1 *Let Y be a discrete RV with probability function $p_Y(y)$ and $g(Y)$ be a real-valued function of Y. Then the expected value of $g(Y)$ (which is also an RV) is:*

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

This is a useful theorem because if $Z = g(Y)$, then

$$E[Z] = E[g(Y)] = \sum_z zp_Z(z) = \sum_z g(y)p_Z(y)$$

If Z is a nice function of Y, we can still calculate $E[Z]$ even if we don't have $p_Z(z)$.

Proof. In the textbook. Read if interested!

Properties of the expected value operator E[ ] on some RV Y:

(1) Let G be a constant. Then E(G)=G.
Think: $g(Y) = G$

$$\Rightarrow E[G] = E[g(Y)] = \sum_y g(y)p_Y(y) = \sum_y Gp_Y(y) = G\underbrace{\sum_y p_Y(y)}_{=1} = G$$

(2) Let C be a constant and $g(\cdot)$ a real valued function.
Then $E[Cg(Y)] = CE[g(Y)]$.

Proof: $$E[Cg(Y)] = \sum_y Cg(y)p_Y(y) = C\sum_y g(y)p_Y(y) = CE[g(Y)]$$

(3) Let g_i, $i = 1, 2, ..., k$ be k real valued functions

Then $E[\sum_{i=1}^k g_i(Y)] = \sum_{i=1}^k E[g_i(Y)]$

Proof: Notice if we let $\varphi(Y) = \sum_{i=1}^k g_i(Y)$ that $\varphi(Y)$ is then a function of Y

$$\Rightarrow E[\sum_{i=1}^{k} g_i(Y)] = E[\varphi(Y)] = \sum_{y} \varphi(y)p_Y(y) = E[\sum_{i=1}^{k} g_i(Y)p_Y(y)]$$
$$= \sum_{i=1}^{k} E[g_i(Y)p_Y(y)] = \sum_{i=1}^{k} E[g_i(Y)]$$

What if Y can take on both positive and negative values?

$$E[|Y|] = \sum_{y} |y|\, p_Y(y) < \infty \quad \Rightarrow \quad E[Y] = \sum_{y} yp_Y(y) < \infty$$

(but **not** the other direction)

Definition 7.2 The variance of a RV Y is defined to be $Var[Y] = E[(Y - \mu_Y)^2]$ where $\mu_Y = E[Y]$.

The square root of $Var[Y]$ is the standard deviation.

Notice: $E[Y - \mu_Y] = E[Y] - E[\mu_Y] = \mu_Y - \mu_Y = 0$.

There is also $E[|Y - \mu_Y|]$ (called the first absolute central moment of Y) but the graph of $|y|$ is not differentiable at 0 while y^2 is. The quadratic is more widely used for its nice properties and applicability.

Theorem 7.2 *Let Y be a RV with probability function $p_Y(y)$. Then,*

$$\sigma_Y^2 = Var[Y] = E[(Y - \mu_Y)^2] = E[Y^2] - \mu_Y^2$$

§8 Lec. 8 - Mon. Sept. 30, 2013

Theorem 8.1 *Let Y be a RV with probability function $p_Y(y)$.*
Then $\sigma_Y^2 = Var[Y] = E[(Y - \mu_Y)^2] = E[Y^2] - \mu_Y^2$, where $\mu_Y = E[Y]$.

Proof.

$$\begin{aligned} E[(Y - \mu_Y)^2] &= E[Y^2 - 2\mu_Y Y + \mu_Y^2] \\ &= E[Y^2] + E[-2\mu_Y^2] + E[M_Y^2], \quad \text{by the distributivity of expected value} \\ &= E[Y^2] - 2\mu_Y E[Y] + \mu_Y^2, \qquad \text{since} \mu_Y = E[Y] \text{ , which is a constant.} \\ &= E[Y^2] - 2\mu_Y \mu_Y + \mu_Y^2 \\ &= E[Y]^2 - \mu_Y^2 \end{aligned}$$

The Binomial Distribution

Characteristics of the Binomial Experiment:

1. The experiment consists of a fixed number n of identical trials.

(1) Each trial results in one of the two possible outcomes. We say an outcome is either a success S or a failure F

(2) The probability of success on a single trial is equal to the same value p and remains the same from trial to trial. Thus the probability of a failure is equal to $q = 1 - p$.

(3) Trials are independent.

(4) The random variables of interest is Y, the number of successes observed in n trials.

Eg. Tossing a fair coin 3 times.

This is equivalent to a sampling with replacement AND with order.

$S = \{\{H,H,H\},\{H,H,T\},\{H,T,H\},\{H,T,T\},\{T,H,H\},\{T,H,T\},\{T,T,H\},\{T,T,T\}\}$.
Let Y denote the number of heads and $A_i = \left\{S_i^1, S_i^2, S_i^3, ...\right\},\ \ i = 0,1,2,3$ the set of tosses with i heads. Then we have:

$$\begin{aligned}
A_0 &= \{\{T,T,T\}\}\\
A_1 &= \{\{H,T,T\},\{T,H,T\},\{T,T,H\}\}\\
A_2 &= \{\{H,H,T\},\{H,T,H\},\{T,H,H\}\}\\
A_3 &= \{\{H,H,H\}\}
\end{aligned}$$

and by the independence of trials:

$$\begin{aligned}
&S_0^1 = P(T)P(T)P(T)\\
&S_1^1 = P(H)P(T)P(T), S_1^2 = P(T)P(H)P(T), S_1^3 = P(T)P(T)P(H)\\
&S_2^1 = P(H)P(H)P(T), S_2^2 = P(H)P(T)P(H), S_2^3 = P(T)P(H)P(H)\\
&S_3^1 = P(H)P(H)P(H)
\end{aligned}$$

Since $P(T) = 1 - P(H) = 1 - \frac{1}{2} = \frac{1}{2}$,

$$P(Y=0) = \sum_{S_0^i \in A_0} P(S_0^i) = \frac{1}{8}$$
$$P(Y=1) = \sum_{S_1^i \in A_1} P(S_1^i) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$
$$P(Y=2) = \sum_{S_2^i \in A_2} P(S_2^i) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}$$
$$P(Y=3) = \sum_{S_3^i \in A_3} P(S_3^i) = \frac{1}{8}$$

Eg. Suppose we repeat the same experiment as in the previous example but with a loaded coin with $P(H) = \rho$.
Then we have $P(T) = 1 - P(Y) = 1 - \rho$ and

$$\begin{aligned}
P(Y=0) \quad &= \sum_{S_0^i \in A_0} P(S_0^i) \\
&= P(T)P(T)P(T) = (1-\rho)^3 \\
P(Y=1) \quad &= \sum_{S_1^i \in A_1} P(S_1^i) \\
&= P(H)P(T)P(T) + P(T)P(H)P(T) + P(T)P(T)P(H) = 3\rho(1-\rho)^2 \\
P(Y=2) \quad &= \sum_{S_2^i \in A_2} P(S_2^i) \\
&= P(H)P(H)P(T) + P(H)P(T)P(H) + P(T)P(H)P(H) = 3\rho^2(1-\rho) \\
P(Y=3) \quad &= \sum_{S_3^i \in A_3} P(S_3^i) \\
&= P(H)P(H)P(H) = \rho^3
\end{aligned}$$

From the previous examples, we now generalise to an experiment where a loaded coin with $P(H) = \rho$ is tossed n times.
The probability of having k heads, i.e. the PMF of the binomial distribution is

$$P(Y=k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2 \ldots$$

To check that this is actually a PMF, we need to verify that

(1) $P(Y=k) \geq 0,\ \forall\, k \in [0, n]$

(2) $\sum_{k=0}^{n} \Pr(Y=k) = 1$

Proof.

(1) Since $0 \leq \rho \leq 1$, the properties of exponents and binomials guarantee this condition.

(2) Recall the binomial expansion formula: $(x+y)^n = \sum_{k=0}^{n} x^k + y^{n-k}$.
Substituting below,

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

Bernoulli Distribution

A special case of the binomial distribution is when $n = 1$. Given the probability of success ρ, the PMF is

$$P(Y = k) = \rho^k (1-\rho)^{1-k}, k = 0, 1$$

This is called the Bernoulli variable.

Let $Z_i, i = 1, 2, ...n$ be Bernoulli variables. Then according to the binomial distribution,

$$Y = \sum_{i=1}^{n} Z_i$$

.

If a variable Y has a binomial distribution with given n and ρ, we write $Y\ Bin(n, \rho)$.

Theorem 8.2 *If $Y \sim Bin(n, \rho)$ then $\mu_Y = E[Y] = n\rho$ and $\sigma_Y^2 =$ Var[Y]$=n\rho(1-\rho)$.*

Proof.

$$
\begin{aligned}
E[Y] &= \sum_{allYs} YP(Y) \\
&= \sum_{k=0}^{n} kP(k) \\
&= \sum_{k=0}^{n} k \binom{n}{k} \rho^k (1-\rho)^{n-k} \\
&= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} \rho^k (1-\rho)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} \rho^k (1-\rho)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)!(n-k)!} \rho^k (1-\rho)^{n-k} \\
&= \sum_{k=1}^{n} \binom{n-1}{k-1} \rho^k (1-\rho)^{n-k} \\
&= n\rho \sum_{k=1}^{m} \binom{n-1}{k-1} \rho^{k-1} (1-\rho)^{n-k} && \text{making the change of variable} r = k-1 \\
&= n\rho \sum_{r=0}^{n-1} \binom{n-1}{r} \rho^{r} (1-\rho)^{n-(r+1)} \\
&= n\rho \sum_{n-1}^{r} \rho^{r} (1-\rho)^{(n-1)-r} && \text{recalling the binomial expansion for} (x+y)^{n-1} \\
&= n\rho[\rho + (1-\rho)]^{n-1} = n\rho * 1^{n-1} = n\rho
\end{aligned}
$$

$\sigma_Y^2 = Var[Y] = E[Y^2] - \mu_Y^2$, where $\mu_Y = E[Y] = n\rho$

Noting that,

$$
\begin{aligned}
E[Y^2] &= \sum_{allYs} Y\rho(Y) \\
&= \sum_{k=0}^{n} k^2 P(k) \\
&= \sum_{k=0}^{n} k^2 \binom{n}{k} \rho^k (1-\rho)^{n-k} \\
&= \sum_{k=1}^{n} k^2 \binom{n}{k} \rho^k (1-\rho)^{n-k} \qquad \text{the terms for } k = 0 \text{are null} \\
&= E[Y(Y-1)+Y] \\
&= E[Y(Y-1)] + E[Y] \text{ and}
\end{aligned}
$$

$$
\begin{aligned}
E[Y(Y-1)] &= \sum_{k=0}^{n} k(k-1)\rho k \\
&= \sum_{k=2}^{n} k(k-1) \binom{n}{k} \rho^k (1-\rho)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1) \frac{n!}{k(k-1)(k-1)!(n-k)} \rho^k (1-\rho)^{n-k} \\
&= \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} \rho^k (1-\rho)^{n-k} \\
&= n(n-1) \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} \rho^k (1-\rho)^{n-k} \\
&= n(n-1)\rho^2 \sum_{k=2}^{n} \binom{n-2}{k-2} \rho^{k-2} (1-\rho)^{n-k} \\
&= n(n-1)\rho^2 \sum_{r=0}^{n-2} \binom{n-2}{r} \rho^{r} (1-\rho)^{(n-2)-r} \qquad \text{setting } r = k-2 \\
&= n(n-1)\rho^2 [\rho + (1-\rho)]^{n-2} \qquad \text{but } \rho + (1-\rho) = 1 \\
&= n(n-1)\rho^2
\end{aligned}
$$

Hence, $E[Y^2] = E[Y(Y-1)] + E[Y] = n(n-1)\rho^2 + n\rho = n^2\rho^2 - n\rho^2 + n\rho$ and

$Var[Y] = E[Y^2] - \mu_Y^2 = [n^2\rho^2 - n\rho^2 + n\rho] - (n\rho)^2 = n^2\rho^2 - n\rho^2 + n\rho - n^2\rho^2 = n\rho(1-\rho)$.

Exercise 3.39 p. 111

Let Y be the number of components that last longer than 1000 hours, $n = 4$, and $Y \sim Bin(n = 4, \rho = 0.8)$.

(a) $\Pr(Y=2) = \binom{4}{2}(0.8)^2(1-0.8)^{4-2}$

(b)

$$\begin{aligned}\Pr(Y \geq 2) &= \Pr(Y=2) + \Pr(Y=3) + \Pr(Y=4) \\ &= \binom{4}{2}(0.8)^2(0.2)^2 + \binom{4}{3}(0.8)^3(0.2)^1 + \binom{4}{4}(0.8)^4\end{aligned}$$

§9 Lec. 9 - Wed. Oct. 2, 2013

§9.1 Geometric Distribution

Q1 What is the chance of winning the 649 lotto jackpot this Saturday?

Q2 What is the chance of winning the game on your 2nd, 3rd, ... try?

Q3 On average, how long does it take to win the jackpot of 649?

All these questions can be considered using the Geometric Probability Distribution.

1. An infinite sequence of identical and independent tries
2. Each try can only result in one of two outcomes: S (win the 649 jackpot) or F (everything else)
3. The probability of success is equal to p and remains the same from try to try
4. The geometric random variables Y is the # of tries on which the first success occurs.

$\mathcal{S}$ consists of infinitely many points.

s_1**:** S: success at the first try

s_2**:** F,S: failure on the first try and success on the second

s_3**:** F,F,S

s_k**:** $\underbrace{\text{F,F,\ldots,S}}_{k-1 \text{ times}}$: Failure at the first $k-1$ tries, but success at the k^{th} trial

$\mathcal{S} = \{s_1, s_2, s_3, \ldots, s_k, \ldots\}$

Y, the geometric random variable, is a map $Y : \mathcal{S} \to \mathbb{R}$. $Y(s_k) = k$; that is, Y maps s_1 to the first trial, s_3 to the third trial, etc.

$$p(Y=k) = p(s_k) = p(\underbrace{\text{F,F,\ldots,S}}_{k-1 \text{ times}})$$

The only way $Y = k$ is if s_k happens.

$$p(\underbrace{\text{F,F,...,S}}_{k-1 \text{ times}}) = \underbrace{(1-p)(1-p)\dots(1-p)}_{k-1} p = (1-p)^{k-1}p$$

So we have $P(Y = k) = (1-p)^{k-1}p$ for $k = 1, 2, \dots$.
Is this a probability mass function? Check:

1. $p(k) = p(Y = k) \geq 0$ ✓
2. $\sum_{k=1}^{\infty} p(k) = 1$:

$$\sum_{k=1}^{\infty}(1-p)^{k-1}p = p\sum_{k=1}^{\infty}(1-p)^{k-1}$$

This is a geometric series:

$$p\sum_{k=1}^{\infty}(1-p)^{k-1} = p\left[\frac{1}{1-(1-p)}\right] = p \cdot \frac{1}{p} = 1 \checkmark$$

Aside How did we evaluate the geometric series?
If we have $\sum_{k=0}^{n} x^k - f_n(x)$ (1) and $xf_n(x) = \sum_{k=0}^{n} x^{k+1}$ (2), then (1)-(2):

$$f_n(x) - xf_n(x) = 1 - x^{n+1}$$

(where all but the first term of (1) and the last term of (2) cancelled)
So,

$$(1-x)f_n(x) = 1 - x^{n+1} \implies f_n(x) = \frac{1-x^{n+1}}{1-x}$$

If $|x| < 1$, then $\lim_{n\to\infty} x^n = 0$, so

$$\lim_{n\to\infty} f_n(x) = \frac{1}{1-x}\lim_{n\to\infty}[1 - x^{n+1}] = \frac{1}{1-x}\left[1 - \underbrace{\lim_{n\to\infty} x^{n+1}}_{=0}\right] = \frac{1}{1-x}$$

So, $\sum_{k=0}^{\infty} x^k = \lim_{n\to\infty}\sum_{k=0}^{n} x^k = \frac{1}{1-x}$ for $|x| < 1$. So choose $x = 1 - p$.

For these kind of questions, calculate μ and θ^2 first.
If $Y \sim \text{Geo}(p)$ then $\text{E}[Y] = \frac{1}{p}$ since the lower the chances of success, the more you need to try.
ex. if chance of getting S is 0.01 then try 100 times.
$\text{Var}[Y] = \frac{1-p}{p^2}$

$$
\begin{aligned}
\mathrm{E}[Y] &= \sum_y y p(y) = \sum_{k=1}^{\infty} k p(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1} p \qquad \leftarrow \text{Not a geo series} \\
&= p \sum_{k=1}^{\infty} k q^{k-1} \qquad \leftarrow \text{if we replace } 1-p \text{ with } q \\
&= p \sum_{k=1}^{\infty} \tfrac{\mathrm{d}}{\mathrm{d}q} q^k \qquad \leftarrow \text{now we need to interchange } \tfrac{\mathrm{d}}{\mathrm{d}q} \text{ and } \Sigma \\
&= p \tfrac{\mathrm{d}}{\mathrm{d}q} \left[\sum_{k=1}^{\infty} q^k \right] \qquad \leftarrow \text{by calc. theorem} \\
&= p \tfrac{\mathrm{d}}{\mathrm{d}q} \left[\frac{1}{1-q} - 1 \right] \qquad \leftarrow \text{already explained} \\
&= p \frac{1}{(1-q)^2} = p \frac{1}{p^2} = \frac{1}{p}
\end{aligned}
$$

Now back to the questions.

Sampling without replacements, 49 numbers, 6 chosen.

Q1: $p = \frac{1}{\binom{49}{6}}$

Q2: $Y \sim \text{Geo}\left(p = \frac{1}{\binom{49}{6}}\right)$

Q3: $\mathrm{E}[Y] = \frac{1}{p} = \frac{1}{\frac{1}{\binom{49}{6}}} = \binom{49}{6} = 13,983,816$

Assume that each year is 52.5 weeks and you have two chances to play per week. So $52.5 * 2 = 105$ times to play per year.

Then $\binom{49}{6}/105 = 133,180$ years on average to win.

Does choosing the same number every trial affect the chances? No, because each trial is independent.

$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mu_Y^2$$

$$
\begin{aligned}
\mathrm{E}[Y^2] &= \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p \qquad \leftarrow \text{hard to calc., use same trick as binomial}\\
&= \mathrm{E}[Y(Y-1)+Y]\\
&= \mathrm{E}(Y(Y-1)] + \mathrm{E}[Y]\\
E[Y(Y-1)] &= \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-1}p \qquad \leftarrow \text{take derivative}\\
&= p(1-p)\sum_{k=2}^{\infty} \tfrac{\mathrm{d}^2}{\mathrm{d}q^2} q^k \qquad \leftarrow q = 1-p\\
&= p(1-p)\tfrac{\mathrm{d}^2}{\mathrm{d}q^2}\left[\sum_{k=2}^{\infty} q^k\right]\\
&= p(1-p)\tfrac{\mathrm{d}^2}{\mathrm{d}q^2}\left[\frac{1}{1-q} - q - 1\right]\\
&= p(1-p)\tfrac{\mathrm{d}^1}{\mathrm{d}q^1}\left[\frac{1}{(1-q)^2} - 1\right]\\
&= p(1-p)\frac{2}{(1-q)^3} = p(1-p)\frac{2}{p^3} = \frac{2}{p^2} - \frac{2}{p}
\end{aligned}
$$

hence

$$E[Y^2] = \frac{2}{p^2} - \frac{2}{p} + \frac{1}{p} = \frac{2}{p^2} - \frac{1}{p}$$

so

$$\mathrm{Var}[Y] = \frac{2}{p^2} - \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{1}{p^2} - \frac{1}{p} = \frac{1-p}{p^2}$$

Know these manipulations as they might end up on exams.

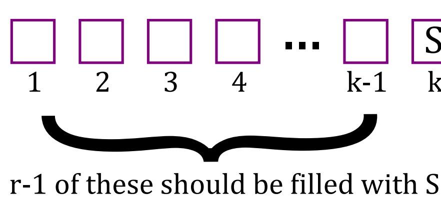
§9.2 Negative Binomial Distribution

The negative binomial distribution: a generalization of the geometric distribution.

I repeat the trial until I have r successes. If $r = 1$, then it's geometric distribution.

Y = number of the trial at which the r^{th} success occurs.

$P(Y = k)$



So we have to choose $r-1$ of the $k-1$ boxes to be filled with S and the remaining $(k-1)-(r-1)$ should be filled with F. Lastly, we need the last box to be filled with S.

Then

$$P(Y = k) = \binom{k-1}{r-1}(1-p)^{(k-1)-(r-1)}p^{r-1}p = \binom{k-1}{r-1}(1-p)^{k-r}p^r$$

$$P(x = k) = \underbrace{\binom{n}{k}}_{k \text{ is changing}} (1-p)^{n-k}p^k \qquad \leftarrow \text{binomial} \qquad \leftarrow \text{need to be non-negative and sum up to 1}$$

Called negative binomial because r is changing.
The negative binomial distribution if a sum of geo just like binomial is a sum of Bernoulli.
Sums to 1:

$$\sum_{k=r}^{\infty} P(Y = k) = \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r q^{k-r} = p^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r}$$

Since $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$, if we differentiate this r times, we come up with q^{k-r}. Then we see it sums to 1.

§9.3 Hypergeometric Distribution

Useful for knowing things like " how many wolves are on Mont Tremplant?"

Capture/recapture technique: we capture a # of animals from a large population, then tag and release them. After a period of time, recapture a certain number of them and see how many are tagged. This allows us to estimate the size of the population.

§10 Lec. 10 - Mon. Oct. 7, 2013

§10.1 Hypergeometric Distribution

Capture → recapture right away so there is no population change.
After capture, we tag r fish
In the Recapture stage, RV Y is the # of tagged fish that we recapture.
Suppose the total number of fish in the world is $N =$ population.
If I take a sample of size n in the second stage, what is $p(Y = y)$?

$$p(Y = y) = \frac{\overbrace{\binom{r}{y}}^{\text{tagged}} \overbrace{\binom{N-r}{N-y}}^{\text{untagged}}}{\binom{N}{n}} = \frac{\text{possible ways to choose } y \text{ tagged}}{\text{possible samples}}$$

$$y = 0, 1, 2, \ldots, \min(r, n)$$

Check if this is a probability mass function

1. $p(Y = y) \geq 0$? Numerator of the ratio is non-negative and denominator is always positive

2. $1 \stackrel{?}{=} \sum y = 0^r p(Y = y) = \sum_{y=0}^{r} \frac{\binom{r}{y}\binom{n-r}{n-y}}{\binom{N}{n}}$ equivalent to this condition: $\sum_{y=0}^{r} \binom{r}{y}\binom{N-r}{n-y} = \binom{N}{n}$.

$$\begin{aligned}(1+x)^N =&(1+x)^r(1+x)^{N-r}\\ \binom{N}{w} \qquad & \binom{r}{y} \quad \binom{N-r}{z}\end{aligned}$$

The number of tagged y in a sample must obviously be at most the size of the sample n (and at most equal to the total number of tagged animals r) so we specify $y \le \min(r, n)$ and $n \le r$ so $y \le n$. So now we can do $\sum_{y=0}^{n} P(Y = y)$.

Remember: the whole point is to estimate N. r, n are known, y is observed, what is left is N from the identity above.

$$P_N(Y = y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = f(N)$$

How do we find N? Suppose $r = 10, n = 8, y = 4$.
$\frac{r}{N} = \frac{y}{n}$ so $\mathrm{E}[Y] = N$. It's an estimation of the population size based on the sample size.

$$\hat{N} = \frac{nr}{y} = \frac{8 \times 10}{4} = 20 \qquad \leftarrow \text{estimate}$$

$\Pr(Y = y)$ is the link from observable parameter $r \,\&\, n, y$ to unobservable N.
We choose N to maximize $\Pr(Y = y)$ because what we observe is what we are supposed to see.
So choose N so that seeing a tagged number of y (in our sample)is most *likely*.
Take the first derivative of $f(N)$; N^* is a local maxima:

$$\frac{f(N^*)}{f(N^*+1)} \ge 1 \text{ and } \frac{f(N^*)}{f(N^*-1)} \ge 1$$

so if $f(N^*) \ge f(N)\ \forall N$ then we can think of it like a threshold has been reached.
This is called the maximum likelihood method.
Where do we get this from?

$$\frac{f(x+\Delta x) - f(x)}{\Delta x} = f(x+1) - f(x) \qquad \text{if } \Delta x = 1$$

When I throw out the net I may not know how many fish we can catch so n is a random variable in real life. So when we assume above that $n \le r$, it is for simplification.

1. If n happens to be greater than r, then

§10.1.1 p. 130 # 3.113 Jury of 6 person etc.

Hypergeometeric is obtained in the recaptured stage $Y \sim HG(N, n, r)$.

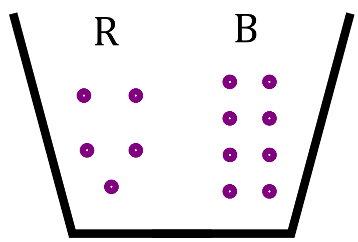
$$N = 20 \qquad r = 8 \qquad n = 6 \qquad y = 1$$

$$p(Y = 1) = \frac{\binom{8}{1}\binom{20-8}{6-1}}{\binom{20}{6}} = \frac{8\binom{20}{5}}{\binom{20}{6}} = \frac{8 \times 12 \times 11}{20 \times 19 \times 17} = 0.163$$

If $Y \sim HG(N, n, r)$, then $\mathrm{E}[Y] = \frac{nr}{N}$ and $\mathrm{Var}[Y] = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$.
$\mathrm{E}[Y] = \frac{nr}{N} = \frac{6 \times 8}{20} = 2.4$.
Then calculate variance and construct an interval and see if 1 falls in it.

Definition: $p = \frac{r}{N}$
$E[Y] = nr/N = np$
$\text{Var}[Y] = np(1-p)$
Note: $\lim_{N\to\infty} \frac{N-n}{N-1} = 1$.

$$\lim_{N\to\infty} \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}} = \binom{n}{y} p^y (1-p)^{n-y} \qquad \leftarrow \text{binomial where } p = \frac{r}{N}$$



$$n \qquad \underbrace{\frac{R}{R+B}}_{=N}$$

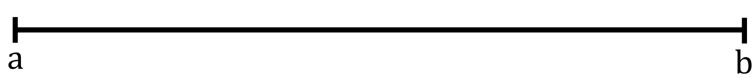
this is sampling with replacement

$$P(Y=r) = \binom{n}{r} \left(\frac{R}{N}\right)^r \left(1 - \frac{R}{N}\right)^{n-r} \qquad \leftarrow \text{binomial with replacement}$$

$$p(Y=r) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}} \qquad \leftarrow \text{sampling without replacement}$$

If N is very large that the variance of the hypergeometric is very similar to binomial. Their $E[Y]$ is already the same, so HG$\sim$Binomial when N is large.

§10.2 Poisson Distribution

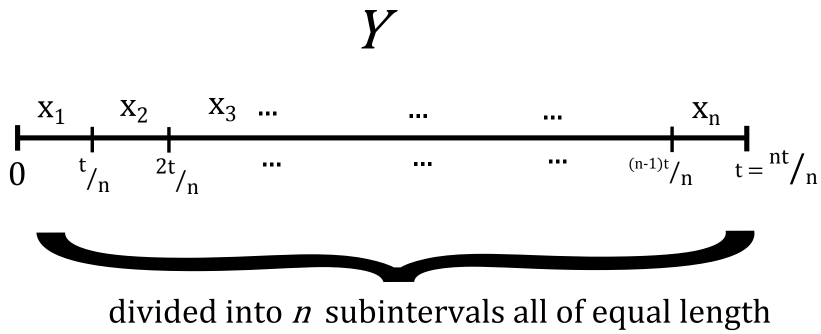


Interested in the # of events in the interval, for example, the interval could be 1 week.
$p(\text{no accident}) \approx 1 - p$ $\qquad \leftarrow$ goes down as intervals shorten
$p(\text{one accident}) \approx p$
$p(\text{more than one accident}) \approx 0$
So this is almost like Bernoulli.

§11 Lec. 11 - Wed. Oct. 9, 2013

§11.1 Poisson Distribution

The Poisson distribution is used for rare events.



1. We assume that the number of events in each interval is independent
2. We assume the number of events in a subinterval depends on the length only and not on the start and end points

The smaller the subinterval, the less the chance of having an event. I can chose n large enough that I have at most 1 event per subinterval. If I make such an assumption, then

$p(\text{no accident}) \approx 1 - p_n \qquad \leftarrow$ goes down as intervals shorten

$p(\text{one accident}) \approx p_n$

$p(\text{more than one accident}) \approx 0$

Note that p_n depends on the size of the subinterval; the bigger n, the smaller p_n.

Make sure to assume that $\underbrace{np_n}_{\text{gives the average}} \approx \lambda \quad \leftarrow$ constant

Y is the number of events in $[0, t]$

$x_i =$ the number of events in $\left[\frac{(i-1)t}{n}, \frac{it}{n}\right]$ for $i = 1, 2, \ldots, n$

$Y = \sum_{i=1}^{n} x_i$. We assumed that the number of events in each interval are independent, so

$$x_i = \begin{cases} 0 & 1 - p_n \\ 1 & p_n \end{cases}$$

which does not change with i.

x_i's are independent; we have "stationary increments"

$Y \sim \text{Bin}(n, p_n)$ since x_i independent

$$p(Y = y) = \binom{n}{y} p_n^y (1 - p_n)^{n-y}$$

$$\begin{aligned} \lim_{n\to\infty} \binom{n}{y} p_n^y (1 - p_n)^{n-y} &= \lim_{n\to\infty} \frac{n(n-1)\ldots(n-y+1)}{y!} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y} \\ &= \frac{\lambda^y}{y!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \left[\left(1 - \frac{\lambda}{n}\right)^{-y} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{y-1}{n}\right)\right] \end{aligned}$$

1. $\lim_{n\to\infty}\left(1-\frac{k}{n}\right)^{-y}=1$ $\leftarrow$ y is a fixed number
2. $\lim_{n\to\infty}\left(1-\frac{k}{n}\right)=1,\ k=1,2,\ldots,y-1$
3. $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=e^{-\lambda}$

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

Aside Suppose $x_n\to x_0$ as $n\to\infty$ and g is continuous at x_0. Then $\lim_{n\to\infty} g(x_n)=g\left(\lim_{n\to\infty} x_n\right)=g(x_0)$.

Say that $A=\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}$.

Then

$$\log A=\lim_{n\to\infty} n\log\left(1-\frac{\lambda}{n}\right)=\lim_{n\to\infty}\frac{\log\left(1-\frac{\lambda}{n}\right)}{1/n}$$

so then we can use L'Hopital's rule.

$$\lim_{n\to\infty}\binom{n}{y}p_n^y(1-p_n)^{n-y}=\frac{\lambda^y}{y!}e^{-\lambda}$$

So

$$P(Y=y)=\frac{e^{-\lambda}\lambda^y}{y!}\qquad y=0,1,2,\ldots$$

where $e=2.71\ldots$ Euler's number.

1. $P(Y=y)\geq 0$ since $\lambda>0$.
2. $\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^y}{y!}\stackrel{?}{=}1$

$$\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^y}{y!}=e^{-\lambda}\underbrace{\sum_{y=0}^{\infty}\frac{\lambda^y}{y!}}_{e^{\lambda}}=1$$

where we've used the power series expansion of e^x.

If $Y\sim\text{Po}(\lambda)$ $\leftarrow$ Poisson

Then

$$\text{E}[Y]=\sum_{y=0}^{\infty}yP(Y=y)=\sum_{y=0}^{\infty}y\frac{e^{-\lambda}\lambda^y}{y!}$$

as n increases, the average remains the same.

$$\text{E}[Y]=\sum_{y=1}^{\infty}y\frac{e^{-\lambda}\lambda^y}{y!}=\lambda e^{-\lambda}\sum_{y=1}^{\infty}\frac{\lambda^{y-1}}{(y-1)!}$$

Set $z = y - 1$. Then

$$\mathrm{E}[Y] = \lambda e^{-\lambda} \underbrace{\sum_{z=0}^{\infty} \frac{\lambda^z}{z!}}_{=e^{\lambda}} = \lambda$$

So,

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

$$E[Y^2] = E[Y(Y-1) + Y]$$

Then

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

This time, set $z = y - 2$.

$$\mathrm{E}[Y(Y-1)] = \lambda^2 e^{-\lambda} \underbrace{\sum_{z=0}^{\infty} \frac{\lambda^z}{z!}}_{=e^{\lambda}} = \lambda^2$$

So, $\mathrm{E}[Y^2] = \lambda^2 + \lambda$.

$$\mathrm{Var}[Y^2] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

§11.1.1 p. 136 # 3.127

$$Y \sim \mathrm{Po}(\lambda = 4)$$

If $Y = \#$ of error/page.

$$P(Y \le 4) = \sum_{y=1}^{4} P(Y = y) = \frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} + \frac{e^{-\lambda}\lambda^2}{2!} + \frac{e^{-\lambda}\lambda^3}{3!} + \frac{e^{-\lambda}\lambda^4}{4!} = e^{-4} \times 34.33$$

There is a trick to calculate this without a calculator.

§11.1.2 p. 137 #3.140

$C = \text{cost} = 100\left(\frac{1}{2}\right)^y$

$Y \sim \mathrm{Po}(\lambda = 2)$.

Y is RV, so is C.

$$\mathrm{E}[C] = \mathrm{E}\left[100\left(\frac{1}{2}\right)^Y\right] = 100\,\mathrm{E}\left[\left(\frac{1}{2}\right)^Y\right]$$

$$E[g(Y)] = \sum_y g(y)P(Y = y)$$

So

$$\mathrm{E}\left[\left(\frac{1}{2}\right)^Y\right] = \sum_{y=0}^{\infty}\left(\frac{1}{2}\right)^Y \frac{e^{-\lambda}\lambda^y}{y!} = e^{-\lambda}\sum_{y=0}^{\infty}\frac{\left(\frac{1}{2}\right)^y}{y!} = e^{-\lambda}e^{\frac{\lambda}{2}} = e^{-\lambda/2} = e^{-2/2} = e^{-1}$$

So,

$$E[C] = \frac{100}{e}$$

§11.2 Moments and Moment Generating Functions

Moment:

- About the origin $\mu'_k = \mathrm{E}[Y^k]$
- About the mean $\mu_k = \mathrm{E}[(Y-\mu)^k]$ where $\mu = \mathrm{E}[Y]$

Definition 11.1 The moment generating function (mgf) $m(t)$ of a random variable Y is defined by $m_Y(t) = \mathrm{E}[e^{tY}]$, where the exponential is interpreted as a power series. We say that the mgf of Y exists if there exists a positive constant b such that $m(t)$ is finite for $|t| \leq b$.

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{E}[e^{tY}] = \mathrm{E}\left[\tfrac{\mathrm{d}}{\mathrm{d}t}e^{tY}\right] = \mathrm{E}[Ye^{tY}]$$

Then,

$$\left.\tfrac{\mathrm{d}m(t)}{\mathrm{d}t}\right|_{t=0} = \left.\mathrm{E}[Ye^{tY}]\right|_{t=0} = \mathrm{E}[Y]$$

Theorem 11.1 *If $m(t)$ exist, then for any positive integer k, $\left.\frac{\mathrm{d}^k}{\mathrm{d}t^k}m(t)\right|_{t=0} = m^{(k)}(0) = \mu'_k$.*

Suppose

1. $Y \sim \mathrm{Bin}(n,p)$.

$$\begin{aligned}
m_Y(t) = \mathrm{E}[e^{tY}] &= \sum_{y=0}^{n} e^{tY}\binom{n}{y}p^y(1-p)^{n-y} \\
&= \sum_{y=0}^{n}\binom{n}{y}\underbrace{(pe^t)^y}_{a}\underbrace{(1-p)^{n-y}}_{b} \\
&= (a+b)^n = \left[pe^t + (1-p)\right]^n \qquad \leftarrow \text{holds for any } t \\
\left.\tfrac{\mathrm{d}^k}{\mathrm{d}t^k}m(t)\right|_{t=0} &= \left. n[pe^t][pe^t+(1-p)]^{n-1}\right|_{t=0} = n[pe^0][pe^0+(1-p)] = n[p][p+1-p] = np
\end{aligned}$$

2. If $Y \sim \mathrm{Geo}(p)$,

$$\begin{aligned}
m_Y(t) = E[e^{tY}] &= \sum_{y=0}^{\infty} e^{tY}(1-p)^{y-1}p = \frac{p}{1-p}\sum_{y=0}^{\infty}e^{tY}(1-p)^y = \frac{p}{1-p}\sum_{y=0}^{\infty}[e^t(1-p)]^y \\
&= \frac{p}{1-p}\frac{1}{1-e^t(1-p)}
\end{aligned}$$

If $e^t(1-p) < 1 \Leftrightarrow e^t < (1-p)^{-1} \Leftrightarrow t < -\log(1-p)$. So,

$$m_Y(t) = \begin{cases} \frac{p}{1-p}\frac{1}{1-e^t(1-p)} & \text{if } t < -\log(1-p) \\ \text{does not exist otherwise} \end{cases}$$

3. If $Y \sim \text{Po}(\lambda)$

$$m_Y(t) = \text{E}[e^{tY}] = \sum_{y=0}^{\infty} e^{ty} \frac{e^{-\lambda}\lambda^y}{y!} = e^{-\lambda} \underbrace{\sum_{y=0}^{\infty} \frac{(\lambda e^t)^y}{y!}}_{e^{\lambda e^t}} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

for all t.

§12 Lec. 12 - Wed. Oct. 16, 2013

§12.1 Chebychev's Theorem

Theorem 12.1 *(Chebychev's (or Tchebysheff's) Theorem)* *Let Y be a random variable with mean μ and finite variance σ^2. Then, for any constant $K > 0$,*

$$P(|Y - \mu| < K\sigma) \geq 1 - \frac{1}{K^2}$$

or, equivalently,

$$P(|Y - \mu| \geq K\sigma) \leq \frac{1}{K^2}$$

§12.2 Chapter 4: Continuous Random Variables and Their Products

§12.2.1 Probability Distribution Function

Q. Why do we need continuous random variable?

Definition 12.1 Let Y denote any random variable. The distribution function of Y, denoted by $F(Y)$ is given by

$$F(Y) = P(Y \leq y)$$

for $-\infty < y < \infty$. ($F(Y)$ is sometimes called the cumulative distribution function).

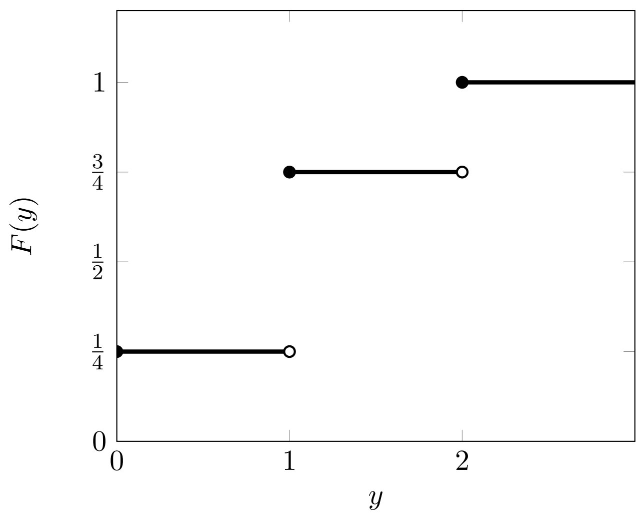
Example 12.2 Suppose $Y \sim \text{Bin}(2, 1/2)$. (i.e. $n = 2, p = \frac{1}{2}$). Then

$$p(y) = \binom{2}{y} \left(\frac{1}{2}\right)^y \left(\frac{1}{2}\right)^{2-y}$$

for $y = 0, 1, 2, \ldots$. and thus, $p(0) = \frac{1}{4}$, $p(1) = \frac{1}{2}$, $p(2) = \frac{1}{4}$. The distribution function of Y is therefore

$$F(Y) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{1}{4} & \text{if } 0 \leq y < 1 \\ \frac{3}{4} & \text{if } 1 \leq y < 2 \\ 1 & \text{if } y \geq 2 \end{cases}$$

The graph of $F(Y)$ is



Note that $F(Y)$ for $Y \sim \text{Bin}(2, 1/2)$ is not continuous; in fact, it it is a step function.

Remark. Distribution functions for discrete random variables are always step functions because the cumulative distribution function increases only at a countable number of points.

Theorem 12.3 ***(Properties of a Dist. Function)*** *If $F(Y)$ is a distribution function, then*

1. $F(-\infty) = \lim_{y=-\infty} F(y) = 0$
2. $F(\infty) = \lim_{y\to\infty} F(Y) = 1$
3. *$F(y)$ is a nondecreasing function of y (if y_1 and y_2 are any values such that $y_1 < y_2$, then $F(y_1) \le F(y_2)$.*
4. *$F(y)$ is right continuous, i.e. $\lim_{z\to y^+} F(z) = F(y)$ for all y.*

Using the above remark we have the following definition.

Definition 12.2 Let Y denote a random variable with distribution function $F(Y)$. Y is said to be continuous if the distribution function $F(Y)$ is continuous, for $-\infty < y < \infty$.

Q. What is the counterpart of $p(y)$ when we deal with continuous random variables?

Q. Suppose Y is a discrete random variable taking values $0, 1, 2, \ldots$. What is the relationship between $p(y) = P(Y = y)$ and $F(y) = P(Y \le y)$?

Answer:

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

Thus, $p(y)$ is sort of a "derivative".

This observation bring the notion of density function, defined below, to our attention.

Definition 12.3 Let $F(Y)$ be the distribution function for a continuous random variable Y. Then $f(y)$, given by

$$f(y) = \frac{\mathrm{d}F(y)}{\mathrm{d}y} = F'(y)$$

whenever the derivative exists, is called the probability density function for the random variable Y.

Theorem 12.4 ***(Properties of a density function)*** *If $f(y)$ is a density function, then*

1. $f(y) \geq 0$ *for any value of* y
2. $\int_{-\infty}^{\infty} f(y)\,\mathrm{d}y = 1$

Remark. Compare these properties to those of $p(y)$.

§13 Lec. 13 - Mon. Oct. 21

Theorem 13.1 *If the random variable Y has a density function $f(y)$ and $a \leq b$, then the probability that Y falls in the interval $[a, b]$ is*

$$P(a \leq Y \leq b) = \int_a^b f(y)\,\mathrm{d}y$$

Do ex. 4.4 on page 159.

§13.1 2. Expected Values for Continuous Random Variables

Everything is the same as that of a discrete random variable. We only replace "$\sum$" with "$\int$".

Definition 13.1 The expected value of a continuous random variable Y is

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} y f(y)\,\mathrm{d}y < \infty$$

provided the integral exists.

Remark. Technically, $\mathrm{E}(Y)$ is said to exist if

$$\mathrm{E}(|Y|) = \int_{-\infty}^{\infty} |y| f(y)\,\mathrm{d}y < \infty$$

Theorem 13.2 *Let $g(y)$ be a function of Y. Then*

$$\mathrm{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) f(y)\,\mathrm{d}y$$

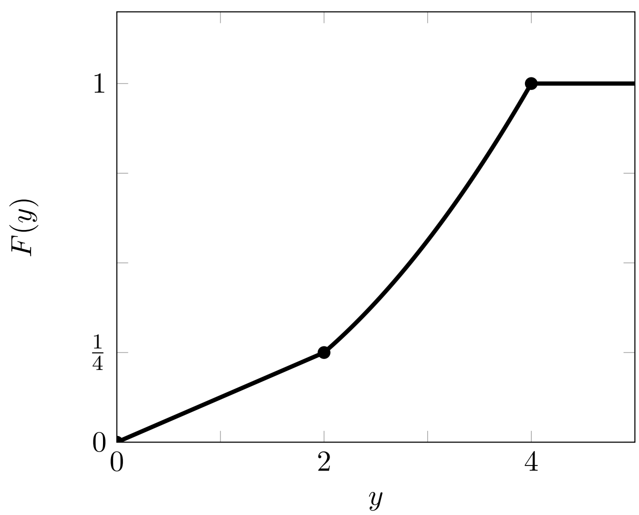
provided that the integral exists.

Theorem 13.3 *Let C be a constant and $g(Y), g_1(Y), \ldots, g_k(Y)$ be functions of a continuous random variable Y. Then*

1. $\mathrm{E}(C) = C$
2. $\mathrm{E}[Cg(Y)] = C\,\mathrm{E}[g(Y)]$
3. $\mathrm{E}\left[\sum_{i=1}^{k} g_i(Y)\right] = \sum_{i=1}^{k} \mathrm{E}[g_i(Y)]$

Do ex. 4.19 on page 164.

$$F(Y) = \begin{cases} 0 & y < 0 \\ \dfrac{y}{8} & 0 \le y < 2 \\ \dfrac{y^2}{16} & 2 \le y < 4 \\ 1 & y \ge 4 \end{cases} \longrightarrow \quad f(y) = \begin{cases} 0 & y < 0 \\ \dfrac{1}{8} & 0 \le y < 2 \\ \dfrac{y}{8} & 2 \le y < 4 \\ 0 & y \ge 4 \end{cases}$$



Note that $F(y)$ is a continuous function, while $f(y)$ is not.

$$\begin{aligned}
\mathrm{E}[Y] &= \int_{-\infty}^{\infty} y f(y)\,\mathrm{d}y \\
&= \underbrace{\int_{-\infty}^{0} y f(y)\,\mathrm{d}y}_{=0} + \int_{0}^{2} y f(y)\,\mathrm{d}y + \int_{2}^{4} y f(y)\,\mathrm{d}y + \underbrace{\int_{4}^{\infty} y f(y)\,\mathrm{d}y}_{=0} \\
&= \int_0^2 y \frac{1}{8}\,\mathrm{d}y + \int_2^4 y \frac{y}{8}\,\mathrm{d}y \\
&= \frac{1}{8}\left[\frac{1}{2}y^2\right]_0^2 + \frac{1}{8}\left[\frac{1}{3}y^3\right]_2^4 \\
&= \frac{1}{4} + \frac{56}{4} = \frac{62}{24} \approx 2.51
\end{aligned}$$

$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - [\mathrm{E}[Y]]^2 = \ldots \text{(Exercise)}$$

§13.2 3. Uniform Distribution

Definition 13.2 If $\theta_1 < \theta_2$, a random variable Y is said to have a continuous uniform probability distribution on the interval (θ_1, θ_2) if and only if the density function of Y is

$$f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1} & \text{if } \theta_1 < y < \theta_2 \\ \\ 0 & \text{if otherwise} \end{cases}$$

Notation: If Y is distributed according to the uniform distribution on an interval (θ_1, θ_2), we usually use the following notation:

$$Y \sim \mathrm{Unif}(\theta_1, \theta_2) \text{ or just } Y \sim \mathrm{U}(\theta_1, \theta_2)$$

Remark. The constants that determine the specific form of a density function are called parameters of the density function. θ_1, θ_2 are parameters of the uniform distribution.

§14 Lec. 14 - Wed. Oct. 23

Importance of the uniform distribution:

1. Simulation studies are a valuable technique for validating models in statistics. For example, to obtain a sample from a random variable Y with cdf F, often we can take observations from Unif$(0, 1)$ and transform them into observations from Y based on what F is. (This can be done because the values that F takes are precisely in the interval (0,1)!) For this reason, most computer systems contain a random number generator, which provides a starting point for generating samples.

2. Relationship between Poisson distribution and uniform distribution. Suppose the number of calls coming into a switchboard, that occur in the interval $(0, t)$, has a Poisson distribution. If it is known (conditional probability) that exactly one such event has occured in the interval $(0, t)$, then the actual true of occurrences is distributed uniformly over this interval.

Now suppose $Y \sim \mathrm{U}(\theta_1, \theta_2)$, then

$$\mathrm{E}[Y] = \frac{\theta_1 + \theta_2}{2}$$

For

$$\mu_Y = \mathrm{E}[Y] = \int_{\theta_1}^{\theta_2} y f(y)\, \mathrm{d}y = \int_{\theta_1}^{\theta_2} \frac{y}{\theta_2 - \theta_1}\, \mathrm{d}y = \frac{1}{\theta_2 - \theta_1}\left[\frac{1}{2}y^2\right]_{\theta_1}^{\theta_2} = \frac{1}{2}\frac{1}{\theta_2^2 - \theta_1^2}\left[\theta_2^2 - \theta_1^2\right] = \frac{\theta_1 + \theta_2}{2}$$

$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mu_Y^2$$

$$
\begin{aligned}
\mathrm{E}[Y^2] &= \int_{\theta_1}^{\theta_2} y^2 f(y)\,\mathrm{d}y = \int_{\theta_1}^{\theta_2} y^2 \frac{1}{\theta_2-\theta_1}\,\mathrm{d}y = \frac{1}{\theta_2-\theta_1}\,\frac{1}{3}y^3\Big|_{\theta_1}^{\theta_2} = \frac{\theta_2^2+\theta_1\theta_2+\theta_1^2}{3} \\
\mathrm{Var}[Y] &= \frac{\theta_2^2+\theta_1\theta_2+\theta_1^2}{3} - \left(\frac{\theta_1+\theta_2}{2}\right)^2 \\
&= \frac{4\left[\theta_2^2+\theta_1\theta_2+\theta_1^2\right] - 3\left[\theta_1^2+2\theta_1\theta_2+\theta_2^2\right]}{12} \\
&= \frac{(\theta_2-\theta_1)^2}{12}
\end{aligned}
$$

Do ex. 4.45 on page 170.
$D =$ diameter of a spherical particle.
$D \sim \mathrm{U}(0.01, 0.05)$, $\theta_1 = 0.01$, $\theta_2 = 0.05$, (recall $r = \frac{D}{2}$)

$$
\begin{aligned}
\mathrm{E}(\text{Volume}) &= \mathrm{E}\left[\frac{4}{3}\pi\,({}^{D}\!/_{2})^3\right] = \frac{4}{3}\pi\,\mathrm{E}\left(\frac{D^3}{8}\right) \\
&= \frac{4}{3}\pi\,\mathrm{E}\left(\frac{D^3}{8}\right) = \frac{\pi}{6}\,\mathrm{E}[D^3] = \frac{\pi}{6}\int_{0.01}^{0.05}\frac{x^3}{0.04}\,\mathrm{d}x \\
&= \frac{\pi}{6}\,\frac{1}{4}x^4\Big|_{0.01}^{0.05} \\
&= \frac{\pi}{24}\left[(0.05)^4-(0.01)^4\right] = \frac{\pi}{24}\frac{624}{10^8} = \frac{65\pi}{10^7}
\end{aligned}
$$

$$
\mathrm{Var}(\text{Volume}) = \mathrm{E}[(\text{Volume})^2] - \mu^2_{\text{volume}}
$$

$$
\mathrm{E}[(\text{volume})^2] = \mathrm{E}\left[\left(\frac{4}{3}\pi\left(\frac{D}{2}\right)^3\right)^2\right] = \left(\frac{4\pi}{3}\right)^2\frac{1}{2^6}\,\mathrm{E}[D^6]
$$

§15 Lec. 15 - Mon. Oct. 28, 2013

§15.1 $\boxed{\text{4. Normal Distribution}}$

Normal distribution is the most widely used continuous probability distribution. Its importance stems mostly from the so-called Central Limit Theorem (CLT) which will be discussed in Chapter 7.

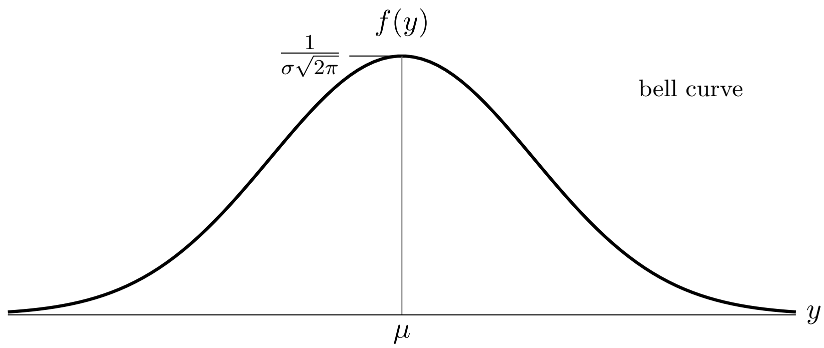
Definition 15.1 A random variable Y is said to have a Normal probability distribution if and only if, for $\sigma > 0$ and $-\infty < \mu < \infty$, the density function of Y is

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

As $\int_{-\infty}^{\infty} f(y)\,\mathrm{d}y = 1$, we have that $\int_{-\infty}^{\infty} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\,\mathrm{d}y = \sigma\sqrt{2\pi}$.

Note. If Y is Normally distriubted, we write $Y \sim \mathrm{N}(\mu, \sigma)$.

The unknown constants μ and σ are parameters of the Normal distribution.



Note. If $Y \sim \mathrm{N}(\mu, \sigma)$, then $\Pr(a \le Y \le b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \,\mathrm{d}y$. There is no closed form for this integral; the evaluation of this integral thus requires numerical integration techniques.

Important property of Normal distribution: define $z = \frac{Y-\mu}{\sigma}$. Then $Y \sim \mathrm{N}(\mu, \sigma)$ if and only if $Z \sim \mathrm{N}(0, 1)$.

Mgf of $Y \sim \mathrm{N}(\mu, \sigma)$

$m_Y(t) = \mathrm{E}[e^{tY}] = ?$. We first notice that $Y = \sigma z + \mu$ so we have that

$$m_Y(t) = \mathrm{E}[e^{tY}] = \mathrm{E}[e^{t(\sigma z + \mu)}] = \mathrm{E}[\underbrace{e^{t\mu}}_{\text{non-random}} \cdot e^{t\sigma z}] = e^{t\mu}\,\mathrm{E}[e^{t\sigma z}]$$

define $\mathfrak{s} = t\sigma$

$$m_Y(t) = e^{t\mu}\,\mathrm{E}[e^{\mathfrak{s}z}] = e^{t\mu} m_z(\mathfrak{s})$$

It then suffices to find $m_z(\mathfrak{s})$:

$$\begin{aligned} m_z(\mathfrak{s}) = \mathrm{E}[e^{\mathfrak{s}z}] &= \int_{-\infty}^{\infty} e^{\mathfrak{s}\xi} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} \,\mathrm{d}\xi \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\xi - \mathfrak{s})^2 + \frac{\mathfrak{s}^2}{2}} \,\mathrm{d}\xi = e^{\mathfrak{s}/2} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z-\mathfrak{s})^2}{2}} \,\mathrm{d}z}_{=1} \\ &= e^{\mathfrak{s}/2} \end{aligned}$$

Aside Proof that $I = \int_0^\infty e^{-x^2/2} \,\mathrm{d}x = \sqrt{\frac{\pi}{2}}$:

$$I^2 = \left(\int_0^\infty e^{-x^2/2} \,\mathrm{d}x\right)^2 = \int_0^\infty \int_0^\infty e^{-\frac{x^2+y^2}{2}} \,\mathrm{d}x \,\mathrm{d}y$$

Set $x = r\cos\theta$ and $y = r\sin\theta$. Then,

$$I^2 = \int_0^\infty \int_0^{\pi/2} e^{-r^2/2} J(r, \theta) \,\mathrm{d}\theta \,\mathrm{d}r$$

where

$$
\begin{aligned}
J(r,\theta) &= \left|\det\begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix}\right| = \left|\det\begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}\right| \\
&= \left|r\cos^2\theta - (r\sin^2\theta)\right| = |r| = r \\
I^2 &= \int_0^\infty \int_0^{\pi/2} e^{-r^2/2} r \,\mathrm{d}\theta \,\mathrm{d}r = \int_0^\infty re^{-r^2/2}\left(\int_0^{\pi/2}\mathrm{d}\theta\right)\mathrm{d}r \\
&= \frac{\pi}{2}\int_0^\infty re^{-r^2/2}\,\mathrm{d}r
\end{aligned}
$$

Let $v = \frac{r^2}{2} \implies \mathrm{d}v = r\,\mathrm{d}r$ and hence

$$
I^2 = \frac{\pi}{2}\int_0^\infty e^{-v}\,\mathrm{d}v = \frac{\pi}{2}\left(-e^{-v}\big|_0^\infty\right) = \frac{\pi}{2}
$$

Thus, $I = \sqrt{\frac{\pi}{2}}$.

Then, we have

$$
m_Y(t) = e^{t\mu}\cdot e^{s^2/2} = e^{t\mu + \frac{s^2}{2}} = e^{t\mu + \frac{t^2\sigma^2}{2}}
$$

Now,

$$
\begin{aligned}
\mathrm{E}[Y] &= \frac{\mathrm{d}m_Y(t)}{\mathrm{d}t}\bigg|_{t=0} = (\mu + t\sigma^2)e^{t\mu+\frac{t^2\sigma^2}{2}}\bigg|_{t=0} = \mu \\
\text{and } \mathrm{E}[Y^2] &= \frac{\mathrm{d}^2 m_Y(t)}{\mathrm{d}t^2}\bigg|_{t=0} = \left[\sigma^2 e^{t\mu+\frac{t^2\sigma^2}{2}} + (\mu t + \sigma^2)^2 e^{t\mu+\frac{t^2\sigma^2}{2}}\right]_{t=0} \\
&= \sigma^2 + \mu^2
\end{aligned}
$$

So,

$$
\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mu_Y^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2
$$

Do 4.88 on page 184 (7^{th} ed.)

§15.2 5. The Gamma Probability Distribution

Definition 15.2 A random variable Y is said to have a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ if and only if the density function Y is

$$
f(y) = \begin{cases} \frac{1}{\beta^\alpha\Gamma(\alpha)} y^{\alpha-1}e^{-\frac{y}{\beta}} & \text{if } 0 \le y < \infty \\ 0 & \text{if elsewhere} \end{cases}
$$

where

$$
\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\,\mathrm{d}y
$$

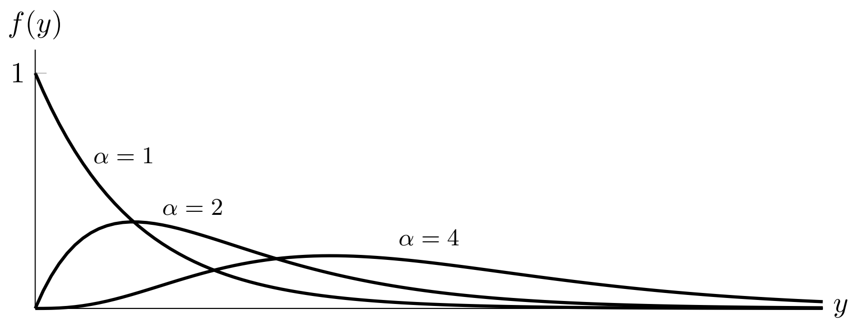


Figure 1. Gamma density functions with $\beta = 1$.

Using integration by parts, we can obtain $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ for any $\alpha > 1$. If $\alpha = n$, an integer number, then by induction, $\Gamma(n) = (n-1)!$.

Note.

(a) $f(y) \geq 0 \;\forall y \geq 0$

(b) $\int_0^\infty f(y)\,\mathrm{d}y = 1$ implies that $\int_0^\infty y^{\alpha-1}e^{-y/\beta}\,\mathrm{d}y = \beta^\alpha\Gamma(\alpha)$.

Notice that the shape of the gamma density function differs for different values of α. That is why α is called the shape parameter. The parameter β is called the scale parameter, as multiplying a gamma-distributed random variable by a positive constant (and thereby changing the scale on which the measurement is made) produces a random variable that also has a gamma distribution with the same value of α (shape parameter) but with a different value of β.

§16 Lec. 16 - Wed. Oct. 30, 2013

The main application of the Gamma distribution is in modeling life time distribution and in general waiting time to an event (such as death or failure in case that lifetime is the random variable of interest).

Suppose $Y \sim \text{Gamma}(\alpha, \beta)$ (For brevity, we sometimes write $Y \sim \text{G}(\alpha, \beta)$. Then,

$$m_Y(t) = \mathrm{E}[e^{tY}] = \int_0^\infty e^{tY}\frac{1}{\beta^\alpha\Gamma(\alpha)}y^{\alpha-1}e^{-\frac{y}{\beta}}\,\mathrm{d}y = \int_0^\infty \frac{1}{\beta^\alpha\Gamma(\alpha)}y^{\alpha-1}e^{-y\left(\frac{1}{\beta}-t\right)}\,\mathrm{d}y$$

Setting $\lambda^{-1} = \frac{1}{\beta} - t$,

$$= \frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty y^{\alpha-1}e^{-\frac{y}{\lambda}}\,\mathrm{d}y$$

We want to evaluate $\int_0^\infty y^{\alpha-1}e^{-\frac{y}{\lambda}}\,\mathrm{d}y$. Define $z = \frac{y}{\lambda} \implies \mathrm{d}z = \frac{1}{\lambda}\,\mathrm{d}y$. Then

$$\int_0^\infty y^{\alpha-1}e^{-y/\lambda}\,\mathrm{d}x = \int_0^\infty (\lambda z)^{\alpha-1}e^{-z}\lambda\,\mathrm{d}z = \lambda^\alpha\int_0^\infty z^{\alpha-1}e^{-z}\,\mathrm{d}z$$
$$\int_0^\infty y^{\alpha-1}e^{-y/\lambda}\,\mathrm{d}x$$

Using $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\,\mathrm{d}y$, we have

$$= \lambda^\alpha \Gamma(\alpha)$$

Note. This also implies that $\frac{1}{\beta^\alpha\Gamma(\alpha)}\int_0^\infty y^{\alpha-1}e^{-\frac{y}{\beta}}\,\mathrm{d}y = 1$, i.e. $\int_0^\infty f(y)\,\mathrm{d}y = 1$.

So,

$$m_Y(t) = \frac{1}{\beta^\alpha\Gamma(\alpha)}\lambda^\alpha\Gamma(\alpha) = \frac{\lambda^\alpha}{\beta^\alpha}$$

if $\lambda > 0$, or, equivalently,

$$\frac{1}{\beta} - t > 0 \Leftrightarrow t < \frac{1}{\beta}$$

Thus,

$$m_Y(t) = \left(\frac{\lambda}{\beta}\right)^\alpha \text{ if } t < \frac{1}{\beta}$$

and therefore,

$$m_Y(t) = \begin{cases} \frac{1}{(1-\beta t)^\alpha} & \text{if } t < \frac{1}{\beta} \\ \\ \text{does not exist} & \text{if otherwise} \end{cases}$$

$$\mu_Y = \mathrm{E}[Y] = \left.\tfrac{\mathrm{d}m_Y(t)}{\mathrm{d}t}\right|_{t=0} = \left.-\alpha(-\beta)(1-\beta t)^{-\alpha-1}\right|_{t=0} = \alpha\beta \tag{1}$$

$$\mathrm{E}[Y^2] = \left.\tfrac{\mathrm{d}^2 m_Y(t)}{\mathrm{d}t^2}\right|_{t=0} = \alpha\beta\left[-(\alpha+1)(-\beta)(1-\beta t)^{-\alpha-2}\right]_{t=0} = \alpha(\alpha+1)\beta^2$$

$$\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mu_Y^2 = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \cancel{\alpha^2\beta^2} + \alpha\beta^2 - \cancel{\alpha^2\beta^2} = \alpha\beta^2 \tag{2}$$

Aside Note that $\Gamma(n+1) = n\Gamma(n)$. This is not too hard to prove:

$$\Gamma(n+1) = \int_0^\infty y^{(n+1)-1}e^{-y}\,\mathrm{d}y = \int_0^\infty y^n e^{-y}\,\mathrm{d}y$$

Let $v = y^n$ and $\mathrm{d}w = e^{-y}\,\mathrm{d}y$. Then $\mathrm{d}v = ny^{n-1}\,\mathrm{d}y$ and $w = -e^{-y}$. Now,

$$\begin{aligned}\Gamma(n+1) &= \int_0^\infty y^n e^{-y}\,\mathrm{d}y = \int_0^\infty v\,\mathrm{d}w = vw|_0^\infty - \int_0^\infty w\,\mathrm{d}v \\ &= \underbrace{-y^n e^{-y}\big|_0^\infty}_{=0} - \int_0^\infty ny^{n-1}\left(-e^{-y}\right)\mathrm{d}y \\ &= n\int_0^\infty y^{n-1}e^{-y}\,\mathrm{d}y = n\Gamma(n)\end{aligned}$$

Thus, we have $\boxed{\Gamma(n+1) = n\Gamma(n)}$. Using this relationship repeatedly, we have $\Gamma(n+1) = n!$. This also gives us a way to define the factorial for non-integer values. Note that in the above derivation, we don't have to have an integer value. In fact, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ for any $\alpha > 0$. Next, we compute $\Gamma\left(\frac{1}{2}\right)$:

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty y^{-\frac{1}{2}}e^{-y}\,\mathrm{d}y$$

Let $x = \sqrt{2y} \implies y = \frac{x^2}{2} \implies \mathrm{d}y = x\,\mathrm{d}x$. Thus,

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{\sqrt{2}}{\cancel{x}} e^{-\frac{x^2}{2}} \cdot \cancel{x}\,\mathrm{d}x = \sqrt{2}\int_0^\infty e^{-x^2/2}\,\mathrm{d}x$$
$$= \sqrt{2}\sqrt{\frac{\pi}{2}} = \sqrt{\pi}$$

---

§16.1 Special cases

1.

Definition 16.1 Let v be a positive integer. A random variable Y is said to have a chi-square distribution with v degrees of freedom if and only if Y is a gamma-distributed random variable with parameters $\alpha = v/2$ and $\beta = 2$.

If Y is distributed according to chi-square, we usually denote this by $Y \sim \chi_v^2$. Using (1) and (2) above, we have

$$\mathrm{E}[Y] = \alpha\beta = \frac{v}{2} \cdot 2 = v \qquad \mathrm{Var}[Y] = \alpha\beta^2 = \frac{v}{2} \cdot 2^2 = 2v$$

2.

Definition 16.2 A random variable Y is said to have an exponential distribution with parameter $\beta > 0$ if and only if the density function of Y is

$$f(y) = \begin{cases} \frac{1}{\beta} e^{-y/\beta} & \text{if } 0 \le y < \infty \\ 0 & \text{if otherwise} \end{cases}$$

If Y is exponentially distributed with parameter β, we use the following notation: $Y \sim \mathrm{Exp}(\beta)$. If $Y \sim \mathrm{Exp}(\beta)$, then $\mathrm{E}[Y] = \beta$ and $\mathrm{Var}[Y] = \beta^2$.

Note. $\chi_v^2 = \mathrm{G}(v/2, 2)$ and $\mathrm{Exp}(\beta) = \mathrm{G}(1, \beta)$.

Do exercise 4.94 on page 191 (7^{th} ed.).

§17 Lec. 17 - Mon. Nov. 4, 2013

§17.1 Section 6: The Beta Probability Distribution

Definition 17.1 A random variable Y is said to have a beta probability distribution with parameters $\alpha > 0$ and $\beta > 0$ if and only if the density function of Y is

$$f(y) = \begin{cases} \frac{y^{\alpha-1}(1-y)^{\beta-1}}{\beta\mathrm{eta}(\alpha,\beta)} & \text{if } 0 \le y \le 1 \\ 0 & \text{if otherwise} \end{cases}$$

where $\beta\mathrm{eta}(\alpha, \beta) = \int_0^1 y^\alpha (1-y)^{\beta-1}\,\mathrm{d}y = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

Our notation for Beta distribution is $Y \sim \text{Beta}(\alpha, \beta)$.

The Beta probability distribution emerges naturally when we study the ratio of two positive random variables and we know the numerator is less than or equal to the denominator.

Suppose that $Y \sim \text{Beta}(\alpha, \beta)$. Then

$$\begin{aligned}
\mu_Y = \mathrm{E}(Y) &= \int_0^1 y \frac{y^{\alpha-1}(1-y)^{\beta-1}}{\beta\text{eta}(\alpha,\beta)}\,\mathrm{d}y = \frac{1}{\beta\text{eta}(\alpha,\beta)}\int_0^1 y^{\alpha}(1-y)^{\beta-1}\,\mathrm{d}y \\
&= \frac{\beta\text{eta}(\alpha+1,\beta)}{\beta\text{eta}(\alpha,\beta)} = \frac{\frac{\Gamma(\alpha+1)\cancel{\Gamma(\beta)}}{\Gamma(\alpha+\beta+1)}}{\frac{\Gamma(\alpha)\cancel{\Gamma(\beta)}}{\Gamma(\alpha+\beta)}} = \frac{\frac{\alpha\cancel{\Gamma(\alpha)}}{(\alpha+\beta)\cancel{\Gamma(\alpha+\beta)}}}{\frac{\cancel{\Gamma(\alpha)}}{\cancel{\Gamma(\alpha+\beta)}}} \\
&= \frac{\alpha}{\alpha+\beta} \\
\mathrm{E}(Y^2) &= \frac{\beta\text{eta}(\alpha+2,\beta)}{\beta\text{eta}(\alpha,\beta)} = \frac{\frac{\Gamma(\alpha+2)\cancel{\Gamma(\beta)}}{\Gamma(\alpha+\beta+2)}}{\frac{\Gamma(\alpha)\cancel{\Gamma(\beta)}}{\Gamma(\alpha+\beta)}} \\
&= \frac{\frac{\alpha(\alpha+1)\cancel{\Gamma(\alpha)}}{(\alpha+\beta)(\alpha+\beta+1)\cancel{\Gamma(\alpha+\beta)}}}{\frac{\cancel{\Gamma(\alpha)}}{\cancel{\Gamma(\alpha+\beta)}}} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} \\
\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mu_Y^2 &= \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\alpha+1}{\alpha+\beta+1}\right) - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \left(\frac{\alpha}{\alpha+\beta}\right)\left[\frac{\alpha+1}{\alpha+\beta+1} - \frac{\alpha}{\alpha+\beta}\right] \\
&= \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\beta}{(\alpha+\beta)(\alpha+\beta+1)}\right) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

Do ex. 4.131 on page 199 (7^{th} edition).

§17.2 Section 7: Tchebysheff's Theorem

Theorem 12.1 ***(Chebychev's (or Tchebysheff's)Theorem)*** *Let Y be a random variable with mean μ and finite variance σ^2. Then, for any constant $K > 0$,*

$$P(|Y-\mu| < K\sigma) \geq 1 - \frac{1}{K^2}$$

or, equivalently,

$$P(|Y-\mu| \geq K\sigma) \leq \frac{1}{K^2}$$

Tchebysheff's theorem is a special case of Markov's inequality:

Theorem 17.1 ***(Markov's inequality)*** *Let $g(Y)$ be a function of the random variable Y with $\mathrm{E}(|g(Y)|) < \infty$. Then*

$$P(|g(Y)| \geq \lambda) \leq \frac{\mathrm{E}(|g(Y)|)}{\lambda}$$

Proof.

$$\begin{aligned}
\mathrm{E}(|g(Y)|) &= \int_{-\infty}^{\infty} |g(Y)| f(y)\,\mathrm{d}y \\
&= \int_{y:|g(y)|\geq\lambda} |g(y)| f(y)\,\mathrm{d}y + \int_{y:|g(y)|<\lambda} |g(y)| f(y)\,\mathrm{d}y \\
&\geq \int_{y:|g(y)|\geq\lambda} |g(y)| f(y)\,\mathrm{d}y \\
&\geq \lambda \int_{y:|g(y)|\geq 1} f(y)\,\mathrm{d}y = \lambda P[|g(Y)| \geq \lambda]
\end{aligned}$$

Thus,

$$P[|g(Y)| \geq \lambda] \leq \frac{\mathrm{E}[|g(Y)|]}{\lambda}$$

□

Now, define $g(y) = (y-\mu)^2$ and $\lambda = K^2\sigma^2$, we therefore have the LHS as:

$$P[|g(Y)| \geq \lambda] = P[(Y-\mu_Y)^2 \geq K^2\sigma^2] = P(|Y-\mu_Y| \geq K\sigma]$$

and the RHS:

$$\frac{\mathrm{E}[|g(Y)|]}{\lambda} = \frac{\mathrm{E}[(Y-\mu_Y)^2]}{K^2\sigma^2} = \frac{\sigma^2}{K^2\sigma^2} = \frac{1}{K^2}$$

This implies that

$$P[|Y-\mu_Y| \geq k\sigma] \leq \frac{1}{K^2}$$

§17.3 Chapter 5: Multivariate Probability Distributions

§17.3.1 Section 1: Joint (Bivariate) Distributions

Q. Is your weight proportional to your height?

Q. What is a Normal proportion?

Q. How can we define a Normal proportion?

To answer questions of this type we need to study the intersection of a pair of events associated with height-weight measurements. For instance, suppose Y_1 is the height and Y_2 is the weight, we then need to study events of the following type:

$$\{Y_1 = y_1\} \cap \{Y_2 = y_2\}$$

which is usually denoted as $(Y_1 = y_1, Y_2 = y_2)$.

Definition 17.2 Let Y_1, Y_2 be discrete random variables. The joint (or bivariate) probability distribution for Y_1 and Y_2 is given by

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2) \qquad -\infty < y_i < \infty, \quad i = 1, 2$$

The function $p(y_1, y_2)$ will be referred to as the joint probability (mass) function.

Properties of $p(y_1, y_2)$:

a) $p(y_1, y_2) \geq 0$ for all y_1, y_2

b) $\sum_{y_1} \sum_{y_2} p(y_1, y_2) = 1$

We can also define the joint distribution function as follows:

Definition 17.3 For any random variables Y_1 and Y_2, the joint (bivariate) distribution function $F(y_1, y_2)$ is given by

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2) \qquad -\infty < y_i < \infty, \quad i = 1, 2$$

Do problem 5.4 on page 232 (6^{th} edition).

§18 Lec. 18 - Wed. Nov. 6, 2013

We also have the continuous counterpart of the above notions. We first need to define what we mean by "continuous" when we have more than one random variable.

Definition 18.1 Let Y_1, Y_2 be continuous random variables with joint distribution function $F(y_1, y_2)$. If there exists a non-negative function $f(y_1, y_2)$ such that

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2)\, \mathrm{d}t_2\, \mathrm{d}t_1, \qquad \text{for all } -\infty < y_i < \infty, \quad i = 1, 2$$

then Y_1 and Y_2 are said to have be jointly continuous random variables. The function $f(y_1, y_2)$ is called the joint probability density function.

Properties

a) $f(y_1, y_2) \geq 0$ for all y_1, y_2.

b) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(y_1, y_2)\, \mathrm{d}y_1\, \mathrm{d}y_2 = 1$

Bivariable cumulative distribution functions satisfy a set of properties similar to those specified for univariate cumulative distribution functions.

Theorem 18.1 *If Y_1 and Y_2 are random variables with joint distribution function $F(y_1, y_2)$ then*

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$
2. $F(+\infty, +\infty) = 1$
3. *If $y^\star \geq y$ and $y_2^\star \geq y_2$, then*

$$\underbrace{\overbrace{F(y_1^\star, y_2^\star)}^{A+B+C+D} - \overbrace{F(y_1^\star, y_2)}^{B+C} - \overbrace{F(y_1, y_2^\star)}^{B+A} + \overbrace{F(y_1, y_2)}^{B}}_{\underbrace{P(y_1 < Y_1 \leq y_1^\star, y_2 < Y_2 \leq y_2^\star)}_{D}} \geq 0$$

where A, B, C, D correspond to the regions of figure 2.

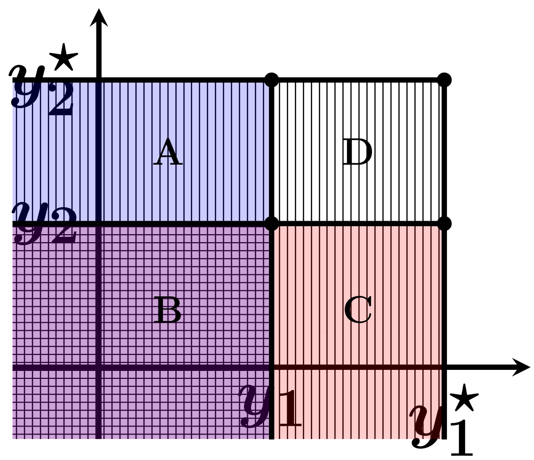


Figure 2. Regions in event space contributing to F evaluated at different arguments.

Do 5.11 on page 234:

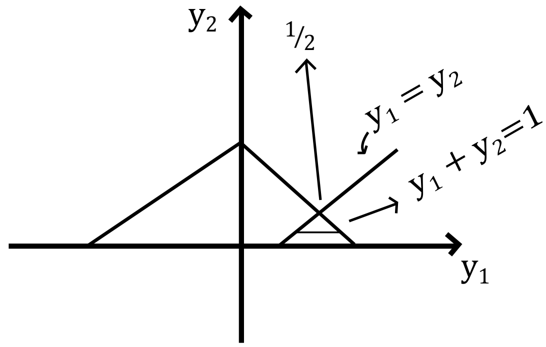
$$f(y_1, y_2) = \begin{cases} 1 & \text{if } -1 + y_2 \leq y_1 \leq 1 - y_2, \ 0 \leq y_2 \leq 2 \\ \\ 0 & \text{if otherwise} \end{cases}$$

a)

$$P(Y_1 \leq {}^3\!/\!{}_4, Y_2 \leq {}^3\!/\!{}_4) = \underbrace{\int_0^{1/4} \int_{-1+y_2}^{3/4} \mathrm{d}y_1 \, \mathrm{d}y_2}_{I} + \underbrace{\int_{1/4}^{3/4} \int_{-1+y_2}^{1-y_2} \mathrm{d}y_1 \, \mathrm{d}y_2}_{II}$$

$$\begin{aligned}
I &= \int_0^{1/4} \int_{-1+y_2}^{3/4} \mathrm{d}y_1 \, \mathrm{d}y_2 = \int_0^{1/4} \left[\int_{-1+y_2}^{3/4} \mathrm{d}y_1 \right] \mathrm{d}y_2 \\
&= \int_0^{1/4} [y_1]_{-1+y_2}^{3/4} \, \mathrm{d}y_2 = \int_0^{1/4} \left[3/4 - (-1 + y_2) \right] \mathrm{d}y_2 \\
&= \int_0^{1/4} \left[7/4 - y_2 \right] \mathrm{d}y_2 = \left[7/4 y_2 - 1/2 y_2^2 \right]_0^{1/4} \\
&= \frac{7}{4} \cdot \frac{1}{4} - \frac{1}{2} \cdot \left(\frac{1}{4} \right)^2 = \frac{7}{16} - \frac{1}{32} = \frac{13}{32} \\
II &= \int_{1/4}^{3/4} \int_{-1+y_2}^{1-y_2} \mathrm{d}y_1 \, \mathrm{d}y_2 = \int_{1/4}^{3/4} \left[\int_{-1+y_2}^{1-y_2} \mathrm{d}y_1 \right] \mathrm{d}y_2 = \int_{1/4}^{3/4} \left[(1 - y_2) - (-1 + y_2) \right] \mathrm{d}y_2 \\
&= \int_{1/4}^{3/4} [2 - 2y_2] \, \mathrm{d}y_2 = 2 \int_{1/4}^{3/4} [1 - y_2] \, \mathrm{d}y_2 \\
&= 2 \left[y_2 - 1/2 y_2^2 \right]_{1/4}^{3/4} = 2 \left[\frac{1}{2} - \frac{1}{2} \left((3/4)^2 - (1/4)^2 \right) \right] \\
&= 1 - \left[\frac{9}{16} - \frac{1}{16} \right] = \frac{1}{2}
\end{aligned}$$

Thus, $P(Y_1 \leq 3/4, Y_2 \leq 3/4) = \frac{13}{32} + \frac{1}{2} = \frac{29}{32}$.



b)

$$\begin{aligned}
P(Y_1 - Y_2 \geq 0) &= \int_0^{1/2} \int_{y_2}^{1-y_2} \mathrm{d}y_1 \, \mathrm{d}y_2 \\
&= \int_0^{1/2} \left[\int_{y_2}^{1-y_2} \mathrm{d}y_1 \right] \mathrm{d}y_2 = \int_0^{1/2} [(1-y_2) - y_2] \, \mathrm{d}y_2 \\
&= \int_0^{1/2} [1 - 2y_2] \, \mathrm{d}y_2 = [y_2 - y_2^2]_0^{1/2} \\
&= \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}
\end{aligned}$$

§18.0.2 Joint Continuity and Separate Continuity

Recall that $g(x, y)$ is called jointly continuous (or simply just continuous) at (x_0, y_0) if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$[(x - x_0)^2 + (y - y_0)^2] < \delta \implies |g(x, y) - g(x_0, y_0)| < \varepsilon$$

An example of a function that is separately continuous in both arguments but is not continuous is

$$g(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

It is obvious that $g(x, y)$ is a continuous function of x (y) for any fixed value of y (x). We now show that it is not continuous at $(0, 0)$. Let $\varepsilon = \frac{1}{4}$. For any $\delta > 0$, choose $x = y = \sqrt{\delta}/2$. Thus,

$$x^2 + y^2 = \frac{\delta}{4} + \frac{\delta}{4} = \frac{\delta}{2} < \delta$$

But,

$$\frac{xy}{x^2 + y^2} = \frac{\sqrt{\delta}/2 \cdot \sqrt{\delta}/2}{(\sqrt{\delta}/2)^2 + (\sqrt{\delta}/2)^2} = \frac{\delta/4}{\delta/4 + \delta/4} = \frac{\cancel{\delta}/4}{\cancel{\delta}/2} = \frac{1}{2} > \frac{1}{4}$$

§18.1 Section 2: Marginal and Conditional Probability Distribution

Definition 18.2 Let Y_1 and Y_2 be jointly discrete (respectively, continuous) random variables with probability function $p(y_1, y_2)$ (resp. probability density function $f(y_1, y_2)$). Then the marginal probability functions (resp. marginal density functions) of Y_1 and Y_2, respectively, are given by

$$p_1(y_1) = \sum_{y_2} p(y_1, y_2) \qquad p_2(y_2) = \sum_{y_1} p(y_1, y_2)$$

$$\left(\text{resp.} \quad f_1(y_1) = \int_{-\infty}^{+\infty} f(y_1, y_2)\,\mathrm{d}y_2 \qquad f_2(y_2) = \int_{-\infty}^{+\infty} f(y_1, y_2)\,\mathrm{d}y_1\right)$$

Conditional probability (resp. density) functions can be similarly defined:

Discrete case:

$$p(y_1 \mid y_2) = p(Y_1 = y_1 \mid Y_2 = y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)}$$

provided that $p_2(y_2) > 0$.

Continuous case:

$$f(y_1 \mid y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

provided that $f_2(y_2) > 0$.

Do 5.22 on page 242 and 5.30 on page 244.

§19 Lec. 19 - Mon. Nov. 11, 2013

 # 5.20 on page 230:

b)

$$P(Y_2 = 0 \mid Y_1 = 0) = \frac{P(Y_2 = 0, Y_1 = 0)}{P(Y_1 = 0)} = \frac{.38}{.76} = \frac{1}{2}$$

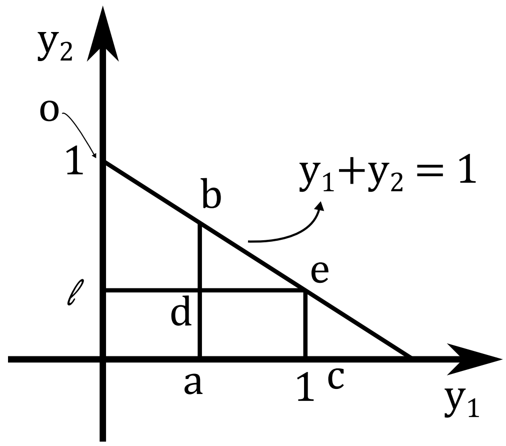
$$P(Y_2 = 1 \mid Y_1 = 0) = \frac{0.14}{.76} \approx 0.18$$

$$P(Y_2 = 2 \mid Y_1 = 0) = \frac{0.24}{0.76} \approx 0.32$$

c)

$$P(Y_1 = 0 \mid Y_2 = 2) = \frac{P(Y_2 = 2, Y_1 = 0)}{P(Y_2 = 2)} = \frac{0.24}{0.29}$$

 # 5.26, on page 231:



a)

$$P(Y_1 \geq 1/2, \mid Y_2 \leq 1/4) = \frac{P(Y_1 \geq 1/2, Y_2 \leq 1/4)}{P(Y_2 \leq 1/4)}$$

$$P(Y_1 \geq 1/2, Y_2 \leq 1/4) \iint_{(y_1,y_2): y_1 \geq 1/2, y_2 \leq 1/4} 2 \, \mathrm{d}y_1 \, \mathrm{d}y_2 = 2[\text{Area of } A]$$

$$\text{Area } A = \text{Area of } \overset{\triangle}{abc} - \text{Area of } \overset{\triangle}{bde}$$

$$= \frac{1}{2}\left[\frac{1}{2} \cdot \frac{1}{2}\right] - \frac{1}{2}\left[\frac{1}{4} \cdot \frac{1}{4}\right] = \frac{3}{32} \implies P(Y_1 \geq 1/2, Y_2 \leq 1/4) = \frac{6}{32}$$

$$P(Y_2 \leq 1/4) = 1 - \overset{\triangle}{ole} = 1 - \frac{1}{2}\left[\frac{3}{4} - \frac{3}{4}\right] = 1 - \frac{9}{32} = \frac{23}{32}$$

a) $\frac{6/32}{23/32} = \frac{6}{23}$

b)

$$P(Y_1 \geq 1/2 \mid Y_2 = 1/4) = \int_{1/2}^{3/4} f_{Y_1|Y_2=1/4}(y_1 \mid y_2 = 1/4) \, \mathrm{d}y_1 = 4/3[3/4 - 1/2] = 1/3$$

$$f_{Y_1|Y_2=1/4}(y_1 \mid y_2 = 1/4) = \frac{f(y_1, y_2)}{f_{Y_2}(1/4)} = \begin{cases} \frac{2}{3/2} & \text{if } 0 \leq y_1 \leq 3/4 \\ 0 & \text{if otherwise} \end{cases}$$

$$f_{Y_2}(1/4) = \int_0^{3/4} f(y_1, y_2) \, \mathrm{d}y_1 = \int_0^{3/4} 2 \, \mathrm{d}y_1 = 2 \cdot \frac{3}{4} = \frac{3}{2}$$

Note that $f_{Y_2}(y_2) = \int_0^{1-y_2} f(y_1, y_2) \, \mathrm{d}y_1 = 2(1-y_2)$ for $0 \leq y_2 \leq 1$, so $f_{Y_2}(1/4) = 2(1-1/4) = \frac{3}{2}$.

§19.1 Section 3. Independent Random Variables

Having defined conditional probability (resp. density) functions, we can define independence between two random variables.

Definition 19.1 Let Y_1 and Y_2 be discrete (resp. continuous) random variables with joint probability (resp. density) function $p(y_1, y_2)$ (resp. $f(y_1, y_2)$) and marginal probability (resp. density) functions $p_1(y_1)$ and $p_2(y_2)$ (resp. $f_1(y_1)$ and $f_2(y_2)$), respectively, then Y_1 and Y_2 are said to be independent if and only if

$$p(y_1, y_2) = p_1(y_1) \cdot p_2(y_2)$$

$$(\text{resp. } f(y_1, y_2) = f_1(y_1) \cdot f_2(y_2))$$

for all pairs of (y_1, y_2).

Do #5.48 on page 252.

§19.2 Section 4. The Expected Value of a Function of Random Variables

Definition 19.2 Let $g(Y_1, \ldots, Y_K)$ be a function of random variables $Y_1, Y_2, \ldots, Y_k$. Then the expected value of $g(Y_1, \ldots, Y_k)$ is

$$\mathrm{E}[g(Y_1, \ldots, Y_k)] = \begin{cases} \sum_{y_k} \cdots \sum_{y_1} g(y_1, \ldots, y_k) p(y_1, \ldots, y_k) & \text{if } Y_1, \ldots, Y_k \text{ are discrete} \\ \int_{y_k} \cdots \int_{y_1} g(y_1, \ldots, y_k) f(y_1, \ldots, y_k)\, \mathrm{d}y_1 \cdots \mathrm{d}y_k & \text{if } Y_1, \ldots, Y_k \text{ are continuous} \end{cases}$$

Theorem 19.1 *Let $Y_1 \perp Y_2$ (Y_1 and Y_2 are independent) and $g(Y_1)$ and $h(Y_2)$ functions of Y_1 and Y_2. Then,*

$$\mathrm{E}[g(Y_1)h(Y_2)] = \mathrm{E}[g(Y_1)]\, \mathrm{E}[h(Y_2)]$$

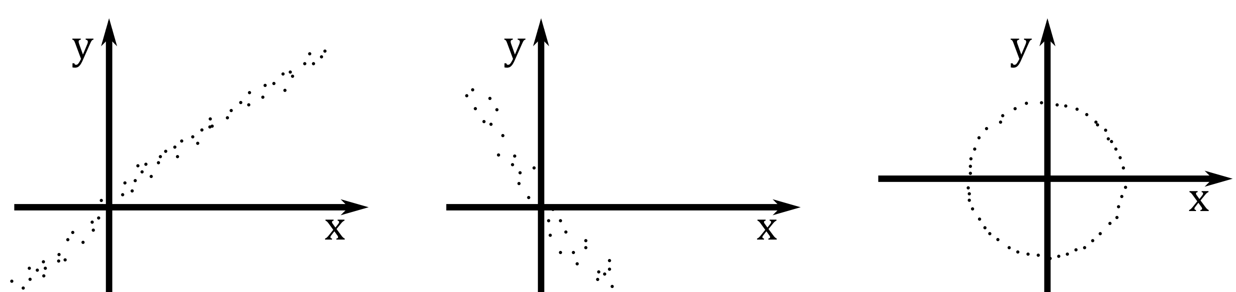
if the expectations exist.

§20 Lec. 20 - Wed. Nov. 13, 2013

§20.1 Section 5. The covariance of two random variables

Q. How can we measure a relationship between two random variables? Let's restrict ourselves to linear relationships (the simplest type of relationship).

Now I should draw some graphics like (a) and (b) on page 249:



(a) (b) (c)

To introduce a measure we always start with the extreme cases. For example, (a) and (b) show close to a perfect linear relationship while (c) is not linear at all. Thus, our measure should be maximal for (a) and (b) and minimal for (c). Thus, we might start with the following which is simple and also reflects our common sense:

$$\sum_{i=1}^{n} x_i y_i \text{ for } n \text{ sample pairs } (x_i, y_i),\ i = 1, 2, \ldots, n$$

To make this quantity location-invariant, we introduce

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

and to have a form like average we divide the latter form by n. We arrive at

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

Thus, we have the following definition:

Definition 20.1 Let Y_1 and Y_2 be random variables with means μ_1 and μ_2 respectively. Then the covariance of Y_1 and Y_2 is given by

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$

Note that

$$\mathrm{Cov}(aY_1, bY_2) = \mathrm{E}[(aY_2 - a\mu_1)(bY_2 - b\mu_2)] = ab\,\mathrm{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)] = ab\,\mathrm{Cov}(Y_1, Y_2)$$

Thus, Cov is not scale invariant. To make it scale invariant, we define the correlation coefficient:

$$\rho(Y_1, Y_2) = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2},\ \sigma_i = [V(Y_i)]^{1/2},\ i = 1, 2$$

$$\begin{aligned}
\mathrm{Cov}(Y_1, Y_2) &= \mathrm{E}[(Y_1 - \mu_1)(Y_2 - \mu_2)] = \mathrm{E}[Y_1Y_2 - \mu_2 Y_1 - \mu_1 Y_2 + \mu_1\mu_2] \\
&= \mathrm{E}[Y_1Y_2] - \mu_2\,\mathrm{E}(Y_1) - \mu_1\,\mathrm{E}(Y_2) + \mu_1\mu_2 \\
&= \mathrm{E}[Y_1, Y_2] - \mu_2\mu_1 - \cancel{\mu_1\mu_2} + \cancel{\mu_1\mu_2} \\
&= \mathrm{E}[Y_1Y_2] - \mu_1\mu_2
\end{aligned}$$

Theorem 20.1 *If Y_1 and Y_2 are independent random variables, then*

$$\mathrm{Cov}(Y_1, Y_2) = 0$$

Proof.

$$\mathrm{Cov}(Y_1, Y_2) = \mathrm{E}(Y_1Y_2) - \mu_1\mu_2$$

Using $Y_1 \perp Y_2$,

$$= \mathrm{E}[Y_1]\,\mathrm{E}[Y_2] - \mu_1\mu_2 = \mu_1\mu_2 - \mu_1\mu_2 = 0$$

$\square$

Theorem 20.2 *Let Y_i, $i = 1, 2, \ldots, n$ and X_j, $j = 1, \ldots, m$ be random variables with* $\mathrm{E}(Y_i) = \mu_i$ *and* $\mathrm{E}(X_j) = \xi_i$. *Define*

$$U_1 = \sum_{i=1}^{n} a_i Y_i \qquad \text{and} \qquad U_2 = \sum_{j=1}^{m} b_i X_j$$

for constants a_i, $i = 1, 2, \ldots, n$ and b_j, $j = 1, \ldots, m$. Then,

a) $\mathrm{E}(U_1) = \sum_{i=1}^{n} a_i \mu_i$, $\mathrm{E}(U_2) = \sum_{j=1}^{m} b_j \xi_j$.

b) $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum\sum_{i<j} a_i a_j \operatorname{Cov}(Y_i, Y_j)$

c) $\operatorname{Cov}(U_1, U_w) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \operatorname{Cov}(Y_i, X_j)$.

Example 20.3 (Example 1) Define

$$U_1 = U_1 = \frac{Y_1}{\sigma_1} + \frac{Y_2}{\sigma_2}, \qquad U_2 = \frac{Y_1}{\sigma_1} - \frac{Y_2}{\sigma_2}$$

Thus,

$$\begin{aligned} 0 \le \operatorname{Var}(U_1) &= \frac{\operatorname{Var}(Y_1)}{\sigma_1^2} + \frac{\operatorname{Var}(Y_2)}{\sigma_2^2} + 2\frac{\operatorname{Cov}(Y_1, Y_2)}{\sigma_1 \sigma_2} \\ &= 2 + 2\rho(Y_1, Y_2) \end{aligned}$$

which implies

$$\rho(Y_1, Y_2) \ge -1 \tag{1}$$

Likewise, we have

$$\begin{aligned} 0 \le \operatorname{Var}(U_2) = \frac{\operatorname{Var}(Y_1)}{\sigma_1^2} + \frac{\operatorname{Var}(Y_2)}{\sigma_2^2} - 2\frac{\operatorname{Cov}(Y_1, Y_2)}{\sigma_1, \sigma_2} \\ = 2 - 2\rho(Y_1, Y_2) \end{aligned}$$

and so,

$$\rho(Y_1, Y_2) \le 1 \tag{2}$$

Using (1) and (2),

$$-1 \le \rho(Y_1, Y_2) \le 1$$

Example 20.4 $X_1, \ldots, X_n$ are iid with mean μ and variance σ^2. Then

$$\mathrm{E}[\bar{X}] = \mu \quad \& \quad \operatorname{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

§21 Lec. 21 - Mon. Nov. 18, 2013

Example 21.1 Mean value and variance for hypergeometric distribution: An urn contains r red balls and $N - r$ black balls. A random sample of n balls is drawn without replacement, and Y, the number of red balls in the sample, is observed. It is clear that $Y \sim HG(N, r, n)$.

Suppose that the sampling is done sequentially and we observe outcomes for $X_1, \dots, X_n$, where

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ draw results in a red ball} \\ \\ 0 & \text{if otherwise} \end{cases}$$

It is clear that $P(X_1 = 1) = \frac{r}{N}$.

$$\begin{aligned} \to P(X_2 = 1) &= P(X_2 = 2 \mid X_1 = 1)P(X_1 = 1) + P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0) \\ &= \frac{r-1}{N-1} \cdot \frac{r}{N} + \frac{r}{N-1} \cdot \frac{N-r}{N} \\ &= \frac{r(r(-1) + r(N-r)}{N(N-1)} = \frac{\cancel{r^2} - r + rN - \cancel{r^2}}{N(N-1)} = \frac{r\cancel{N-1}}{N\cancel{N-1}} = \frac{r}{N} \end{aligned}$$

Thus, X_1 and X_2 have the same distribution. Likewise,

$$\to P(X_k = 1) = \frac{r}{N}, \quad k = 1, 2, \dots, n$$

In a similar fashion, one can show that

$$\to P(X_j = 1, X_k = 1) = \frac{r(r-1)}{N(N-1)}, \quad j \neq k$$

Now,

$$Y = \sum_{i=1}^{n} X_i$$

Thus,

$$\mathrm{E}(Y) = \sum_{i=1}^{n} \mathrm{E}(X_i) = \sum \frac{r}{N} = \frac{nr}{N}$$

On the other hand, X_i is a Bernoulli random variable wtih $p = \frac{r}{N}$. Then,

$$V(X_i) = \frac{r}{N}\left(1 - \frac{r}{N}\right)$$

and

$$\mathrm{Cov}(X_i, X_j) = \mathrm{E}(X_i X_j) - \mathrm{E}(X_i)\,\mathrm{E}(X_j) = \frac{r(r-1)}{N(N-1)} - \left(\frac{r}{N}\right)^2 = -\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{1}{N-1}\right)$$

Now,

$$\begin{aligned} \to V(Y) = V\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} V(X_i) + 2\sum\sum_{i<j} \mathrm{Cov}(X_i, X_j) \\ &= \sum_{i=1}^{n}\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right) + 2\sum\sum_{i<j}\left[-\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{1}{N-1}\right)\right] \\ &= n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right) - \cancel{2}\left(\frac{n(n-1)}{\cancel{2}}\right)\frac{r}{N}\left(1 - \frac{r}{N}\right)\frac{1}{N-1} \\ &= n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left[1 - \frac{n-1}{N-1}\right] = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) \end{aligned}$$

§21.1 6. The Multinomial Probability Distribution

Definition 21.1 The Multinomial Experiment

1. The experiment consists of n identical balls
2. The outcome of each trial falls into one of k classes or cells with probability p_i $\left(\sum_{i=1}^k p_i = 1\right)$.
3. The trials are independent
4. The random variables of interest are Y_i, $i = 1, 2, \ldots, k$ where $Y_i = \#$ of trials for which the outcome falls into cell i. (Note that $\sum_{i=1}^n Y_i = n$).

$$p(y_1, \ldots, p_k) = \binom{n}{y_1, \ldots, y_k} \prod_{i=1}^k p_i^{y_i}, \quad 0 \le y_i \le n,\ i = 1, 2, \ldots, k, \quad \sum_{i=1}^k y_i = 1$$

Define

$$X_{i\ell} = \begin{cases} 1 & \text{if the } \ell\text{-th trial results in the } i\text{-th cell} \\ \\ 0 & \text{if othewise} \end{cases}$$

where $i = 1, 2, \ldots, k$ and $\ell = 1, 2, \ldots, n$.

$P(X_{i\ell} = 1) = p_i$. $X_{i1}, \ldots, X_{1n}$ are independent.

$$Y_i = \sum_{\ell=1}^n X_{i\ell} \text{ so } \mathrm{E}(Y_i) = \sum_{\ell=1}^n \mathrm{E}(X_{i\ell}) = np_i$$

and $V(Y_i) = np_iq_i$. In fact,

$$Y_i \sim \mathrm{Bin}(n, p_i), \quad i = 1, 2, \ldots, k$$

$$\begin{aligned} \mathrm{Cov}(Y_i, Y_j) &= \mathrm{Cov}\left(\sum_{\ell=1}^n X_{i\ell}, \sum_{r=1}^n X_{jr}\right) \\ &= \sum_{\ell=1}^n \sum_{r=1}^n \mathrm{Cov}(X_{i\ell}, X_{jr}) \end{aligned}$$

$$\mathrm{Cov}(X_{i\ell}, X_{jr}) = \begin{cases} \mathrm{E}(X_{i\ell})\,\mathrm{E}(X_{jr}) - \mathrm{E}(X_{i\ell})\,\mathrm{E}(X_{jr}) = 0 & \text{if } \ell \ne r \\ \\ \mathrm{E}(X_{i\ell})\,\mathrm{E}(X_{j\ell}) - \mathrm{E}(X_{i\ell})\,\mathrm{E}(X_{jr}) = 0 - p_ip_j & \text{if } \ell = r \end{cases}$$

$$X_{i\ell}X_{j\ell} \equiv 0$$

Therefore, $\mathrm{Cov}(Y_i, Y_j) = \sum_{\ell=1}^n -p_ip_j$ and hence, $\mathrm{Cov}(Y_i, Y_j) = -np_ip_j$.

Do # 5.123 on page 283.

§22 Lec. 22 - Wed. Nov. 20, 2013

§22.1 7. Conditional Expectation

Definition 22.1 Let Y_1 and Y_2 be two random variables and $g(Y_1)$ be a function of Y_1. Then,

$$\begin{aligned} \mathrm{E}[g(Y_1) \mid Y_2 = y_2] &= \int_{-\infty}^{+\infty} g(y_1) f(y_1 \mid y_2)\,\mathrm{d}y_1 \qquad \text{if } Y_1 \text{ and } Y_2 \text{ are continuous} \\ &= \sum_{y_1} g(y_1) p(y_1 \mid y_2) \qquad \text{if } Y_1 \text{ and } Y_2 \text{ are discrete} \end{aligned}$$

Note that $\mathrm{E}[g(Y_1) \mid Y_2 = y_2]$ is a function of y_2. Thus, $\Phi(y_2) = \mathrm{E}[g(Y_1) \mid Y_2 = y_2]$ is a random variable (a function of Y_2).

Theorem 22.1

$$\mathrm{E}(Y_1) = \mathrm{E}[\mathrm{E}[Y_1 \mid Y_2]]$$

Proof.

$$\begin{aligned} \mathrm{E}[Y_1] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y_1 f(y_1, y_2)\,\mathrm{d}y_1\,\mathrm{d}y_2 \\ &= \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} y_1 f(y_1 \mid y_2) \right\} f(y_2)\,\mathrm{d}y_2 \\ &= \int_{-\infty}^{+\infty} \mathrm{E}[Y_1 \mid Y_2 = y_2] f(y_2)\,\mathrm{d}y_2 \\ &= \mathrm{E}[\mathrm{E}[Y_1 \mid Y_2]] \end{aligned}$$

□

Using the definition of $\mathrm{E}[g(Y_1) \mid Y_2 = y_2]$, we can define

Definition 22.2

$$\mathrm{Var}(Y_1 \mid Y_2 = y_2) = \mathrm{E}[Y_1 \mid Y_2 = y_2] - [\mathrm{E}(Y_1 \mid Y_2 = y_2)]^2$$

5.15 on page 287

$$\mathrm{Var}(Y_1) = \mathrm{E}[V(Y_1 \mid Y_2)] + \mathrm{Var}[\mathrm{E}(Y_1 \mid Y_2)]$$

Proof.

$$\begin{aligned} \mathrm{Var}(Y_1) &= \mathrm{E}[Y_1^2] - [\mathrm{E}(Y_1)]^2 \\ &= \mathrm{E}\left\{\mathrm{E}[Y_1^2 \mid Y_2]\right\} - \{\mathrm{E}[\mathrm{E}(Y_1 \mid Y_2)]\}^2 \end{aligned}$$

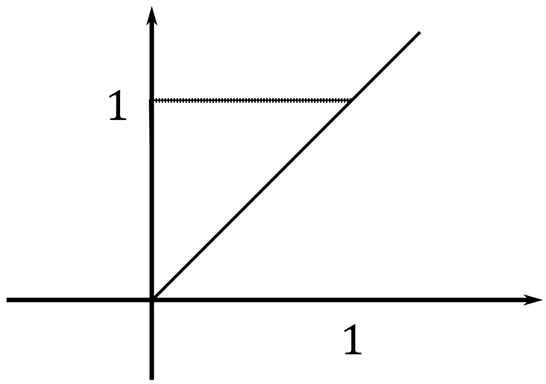
add and subtract $\mathrm{E}\{[\mathrm{E}(Y_1 \mid Y_2)]\}$

$$= \left(\mathrm{E}\left\{\mathrm{E}[Y_1^2 \mid Y_2]\right\} - \mathrm{E}\left\{[\mathrm{E}(Y_1 \mid Y)2]^2\right\}\right) + \left(\mathrm{E}\left\{\overbrace{[\mathrm{E}(Y_1 \mid Y_2)]^2}^{\varphi(Y_2)}\right\} - \{\mathrm{E}[\mathrm{E}(Y_1 \mid Y_2)]\}^2\right)$$

$$= \mathrm{E}[\mathrm{Var}(Y_1 \mid Y_2)] + \mathrm{Var}[\mathrm{E}(Y_1 \mid Y_2)]$$

□

Do #5.133 on page 289:



$$f(y_1, y_2) = \begin{cases} 6(1-y_2) & \text{if } 0 \le y_1 \le y_2 \le 1 \\ \\ 0 & \text{if otherwise} \end{cases}$$

$$\begin{aligned} f(y_1 \mid y_2) &= \frac{f(y_1, y_2)}{f(y_2)} \\ f(y_2) &= \int_{-\infty}^{+\infty} f(y_1, y_2)\,\mathrm{d}y_2 = \int_0^{y_2} 6(1-y_2)\,\mathrm{d}y_1 \\ &= 6y_2(1-y_2) \text{ if } 0 \le y_2 \le 1 \end{aligned}$$

Thus,

$$f(y_1 \mid y_2) = \begin{cases} \frac{1}{y_2} & \text{if } 0 \le y_1 \le y_2 \\ \\ 0 & \text{if otherwise} \end{cases}$$

a) $\mathrm{E}[Y_1 \mid Y_2 = y_2] = \int_0^{y_2} y_1 \cdot \frac{1}{y_2}\,\mathrm{d}y_2 = \frac{1}{y_2}\left[\frac{1}{2}y_1^2\right]_0^{y_2} = \frac{1}{y_2}\frac{1}{2}y_2^2 = \frac{y_2^2}{2}$

b) $\mathrm{E}[Y_1] = \mathrm{E}[\mathrm{E}(Y_1 \mid Y_2)] = \mathrm{E}\left[\frac{Y_2}{2}\right] = \frac{1}{2}\,\mathrm{E}[Y_2]$

$$\begin{aligned} \mathrm{E}[Y_2] &= \int_0^1 y_2 6y_2(1-y_2)\,\mathrm{d}y_2 = 6\int_0^1 y_2^2(1-y_2)\,\mathrm{d}y_2 \\ &= 6\beta\mathrm{eta}(3,2) = 6\frac{\Gamma(3)\Gamma(2)}{\Gamma(3+2)} = 6\frac{2!1!}{4!} = \frac{1}{2} \end{aligned}$$

Thus, $\mathrm{E}[Y_1] = \frac{1}{4}$.

§22.2 Chapter 6: Functions of Random Variables

We learnt that when calculating $\mathrm{E}[(g(Y_1))^r]$ where g is a function of Y we don't need to have the distribution of $g(y)$ as

$$\mathrm{E}[(g(Y))^r] = \int_y (g(y))^r f(y)\,\mathrm{d}y$$

where $f(y)$ is the density of Y.

§23 Lec. 23 - Mon. Nov. 25, 2013

Sometimes, we, however, need the probability/density function of Y. In this chapter, we will discuss three methods:

1. The method of distribution functions
2. The method of transformations
3. The method of moment-generating functions

We illustrate each method by the means of examples.

1. Method of Distribution Functions:
Do # 6.2 on page 307.

$$f(y) = \begin{cases} (3/2)y^2 & \text{if } -1 \leq y \leq 1 \\ 0 & \text{if otherwise} \end{cases}$$

a) Density of $U_1 = 3Y$.

$$\begin{aligned} F_{U_1}(u) &= P(U_1 \leq u) = P(3Y \leq u) = P(Y \leq u/3) = F_Y(u/3) \\ &= \int_{-1}^{u/3} (3/2)y^2 \, \mathrm{d}y \end{aligned}$$

Note that the possible values for U_1 is $[-3, 3]$.

$$= (3/2)\int_{-1}^{u/3} u^2 \, \mathrm{d}y = \frac{3}{2}\left[\frac{1}{3}y^3\right]_{-1}^{u/3} = \frac{1}{2}\left[\frac{u^3}{3^3} - (-1)^3\right] = \frac{1}{2}\left[\frac{u^3}{27} + 1\right]$$

Thus,

$$F_{U_1}(u) = \begin{cases} 0 & \text{if } u < -3 \\ \frac{1}{2}\left[\frac{u^3}{27} + 1\right] & \text{if } -3 \leq u \leq 3 \\ 1 & \text{if } u > 3 \end{cases}$$

Now differentiating F_{U_1}, we obtain

$$f_{U_1}(u) = \begin{cases} \frac{u^2}{18} & \text{if } -3 \leq u \leq 3 \\ 0 & \text{if otherwise} \end{cases}$$

b)

$$\begin{aligned}
F_{U_2}(u) &= P(U_2 \le u) = P(3-Y \le u) = P(3-u \le Y) = 1 - P(Y \le 3-u) = 1 - F_Y(3-u) \\
&= 1 - \int_{-1}^{3-u} (3/2) y^2 \, \mathrm{d}y = 1 - \frac{3}{2}\int_{-1}^{3-u} y^2 \, \mathrm{d}y = 1 - \frac{1}{2}\left[y^3\right]_{-1}^{3-u} \\
&= 1 - \frac{1}{2}\left[(3-u)^3 - (-1)^3\right] \\
&= \frac{1}{2}[1-(3-u)^3]
\end{aligned}$$

Note that $2 \le 3 - Y \le 4$

$$F_{U_2}(u) = \begin{cases} 0 & \text{if } u < 2 \\ \frac{1}{2}[1-(3-u)^3] & \text{if } 2 \le u \le 4 \\ 1 & \text{if } u > 4 \end{cases}$$

Now differentiating F_{U_2}, we obtain

$$f_{U_2}(u) = \begin{cases} \frac{3}{2}(3-u)^2 & \text{if } 2 \le u \le 4 \\ 0 & \text{if otherwise} \end{cases}$$

c) $U_3 = Y^2$

$$F_{U_3}(u) = P(U_3 \le u) = P(Y^2 \le u) = P(|Y| \le u^{1/2})$$

Note that U_1 and U_2 are monotone functions of Y. U_3 is not.

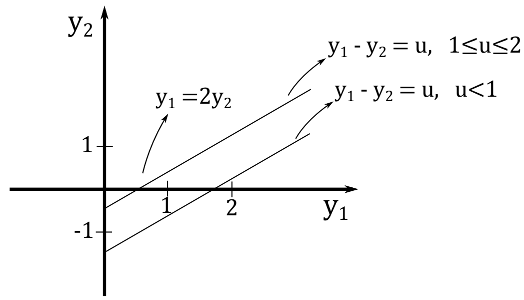
$$\begin{aligned}
&= P(-u^{1/2} \le Y \le u^{1/2}) = F_Y(u^{1/2}) - F_Y(-u^{1/2}) \\
&= \int_{-u^{1/2}}^{u^{1/2}} (3/2) y^2 \, \mathrm{d}y = \left.\frac{1}{2}y^3\right|_{-u^{1/2}}^{u^{1/2}} = u^{3/2}
\end{aligned}$$

$$F_{U_3}(u) = \begin{cases} 0 & \text{if } u < 0 \\ u^{3/2} & \text{if } 0 \le u \le 1 \\ 1 & \text{if } u > 1 \end{cases}$$

Now differentiating $F_{U_3}(u)$, we have

$$f_{U_3}(u) = \begin{cases} 3/2 u^{1/2} & \text{if } 0 \le u \le 1 \\ 0 & \text{if otherwise} \end{cases}$$

Do 6.6 on page 308:



$$f(y_1, y_2) = \begin{cases} 1 & \text{if } 0 \le y_1 \le 2, 0 \le y_2 \le 1, 2y_2 \le y_2 \\ \\ 0 & \text{if otherwise} \end{cases}$$

$$U = Y_1 - Y_2$$

$$F_U(u) = P(U \le u) = P(Y_1 - Y_2 \le u) = \begin{cases} \int_0^u \int_0^{y_1/2} 1 \,\mathrm{d}y_2 \,\mathrm{d}y_1 + \int_2^{2u} \int_{y_1-u}^{y_1/2} 1 \,\mathrm{d}y_2 \,\mathrm{d}y_1 & \text{if } 0 \le u < 1 \\ \\ 1 - \int_u^2 \int_0^{y_1-u} 1 \,\mathrm{d}y_2 \,\mathrm{d}y_1 & \text{if } 1 \le u \le 2 \end{cases}$$

$$= \begin{cases} \int_0^u y_1/2 \,\mathrm{d}y_1 + \int_u^{2u} [y_1/2 - (y_1 - u)] \,\mathrm{d}y_1 & \text{if } u < 1 \\ \\ 1 - \int_u^2 (y_1 - u) \,\mathrm{d}y_1 & \text{if } 1 \le u \le 2 \end{cases}$$

$$= \begin{cases} \int_0^u y_1/2 \,\mathrm{d}y_1 - \int_0^{2u} y_1/2 + \int_u^{2u} u \,\mathrm{d}y_1 & \text{if } 1 \le u \le 2 \\ \\ 1 - \int_0^{2-u} z \,\mathrm{d}z & \text{if } 1 \le u \le 2 \end{cases}$$

where $z = y_1 - u$

$$= \begin{cases} \frac{1}{4}u^2 - \frac{3}{4}u^2 + u^2 & \text{if } 1 \le u \le 2 \\ \\ 1 - \frac{1}{2}(2-u)^2 & \text{if } 1 \le u \le 2 \end{cases}$$

Thus,

$$F_U(u) = \begin{cases} 0 & \text{if } u < 0 \\ \\ \frac{1}{2}u^2 & \text{if } 0 \le u < 1 \\ \\ 1 - \frac{1}{2}(2-u)^2 & \text{if } 1 \le u \le 2 \\ \\ 1 & \text{if } u > 2 \end{cases}$$

Differentiating $F_U(u)$, we obtain

$$f_U(u) = \begin{cases} u & \text{if } 0 \le u < 1 \\ 2-u & \text{if } 1 \le u < 2 \\ 0 & \text{if otherwise} \end{cases}$$

§23.1 Summary of the Distribution Function Method

Let U be a function of the random variables $Y_1, \ldots, Y_n$. Then in order to find $f_U(u)$,

1. Find the region $U = u$ in the $(y_1, \ldots, y_n)$ space
2. Find the region $U \le u$
3. Find $F_U(u) = P(U \le u)$ by integrating $f(y_1, \ldots, y_n)$ over the region where $U \le u$
4. Find the density function $f_U(u)$ by differentiating $F_U(u)$.

§24 Lec. 24 - Wed. Nov. 27

§24.1 2. The Method of Transformation

Let Y have probability density function $f_Y(y)$. If $h(y)$ is either *increasing* or *decreasing* for all y such that $f_Y(y) > 0$, then $U = h(Y)$ has density function $f_U(u) = f_Y(h^{-1}(u)) \left|\frac{\mathrm{d}h^{-1}(u)}{\mathrm{d}u}\right|$.
Note that it is crucial that $h(y)$ is either increasing or decreasing for all y.

Do #6.28 on page 317:

$$f_Y(y) = \begin{cases} 1 & \text{if } 0 \le y \le 1 \\ 0 & \text{if otherwise} \end{cases}$$

$$U = -2\ln(Y) \text{ so } Y = e^{-\frac{U}{2}}$$

Note $0 < Y < 1 \implies \ln Y < 0 \implies U > 0$.

$$\frac{\mathrm{d}y}{\mathrm{d}u} = -\frac{1}{2}e^{-U/2}$$

$$f_U(u) = f_y\left(e^{-U/2}\right)\left|\frac{\mathrm{d}y}{\mathrm{d}u}\right| = 1 \cdot \left|-\frac{1}{2}e^{-U/2}\right| = \frac{1}{2}e^{-U/2} \text{ if } u > 0$$

Do #5.31 on page 317:

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{1}{8}y_1 e^{-\frac{y_1+y_2}{2}} & \text{if } y_1 > 0,\, y_2 > 0 \\ 0 & \text{if otherwise} \end{cases}$$

$$U = \frac{Y_2}{Y_1}$$

Step 1: We find joint Y_1 and U. To do this we first fix $Y_1 = y_1$. Thus, $U = \frac{Y_2}{y_1}$ and so $Y_2 = y_1 U$,which implies $h^{-1}(u) = y_1 u$. Therefore, $\left|\frac{\mathrm{d}h^{-1}(u)}{\mathrm{d}u}\right| = y_1$.

$$
\begin{aligned}
f_{Y_1,U}(y_1,u) &= f_{Y_1,Y_2}(y_1, y_1 u)\left|\frac{\mathrm{d}h^{-1}(u)}{\mathrm{d}u}\right| \\
&= \begin{cases} \frac{1}{8} y_1 e^{-\frac{y_1+y_1u}{2}}\cdot y_1 & \text{if } y_1>0,\ y_2>0 \\ 0 & \text{if otherwise} \end{cases} = \begin{cases} \frac{1}{8} y_1^2 e^{-\frac{y_1(1+u)}{2}} & \text{if } y_1>0,\ y_2>0 \\ 0 & \text{if otherwise} \end{cases} \\
f_U(u) &= \int_0^\infty \frac{1}{8} y_1^2 e^{-\frac{y_1(1+u)}{2}}\,\mathrm{d}y_1 \\
v = y_1^2 &\qquad \mathrm{d}w = e^{-\frac{y_1(1+u)}{2}}\,\mathrm{d}y_1 \\
\mathrm{d}v = 2y_1\,\mathrm{d}y_1 &\qquad w = -\frac{2}{1+u} e^{-\frac{y_1(1+u)}{2}} \\
f_U(u) &= \frac{1}{8}\left[-\frac{2y_1^2}{1+u} e^{-\frac{y_1(1+u)}{2}}\Bigg|_0^\infty + \int_0^\infty \frac{4y_1}{1+u} e^{-\frac{y_1(1+u)}{2}}\,\mathrm{d}y_1 \right] \\
&= \frac{4}{1+u}\int_0^\infty y_1 e^{-\frac{y_1(1+u)}{2}}\,\mathrm{d}y_1 \\
v = y_1 &\qquad \mathrm{d}w = e^{-\frac{y_1(1+u)}{2}}\,\mathrm{d}y_1 \\
\mathrm{d}v = \mathrm{d}y_1 &\qquad w = -\frac{2}{1+u} e^{-\frac{y_1(1+u)}{2}} \\
f_U(u) &= \frac{1}{\cancelto{2}{8}}\cdot \frac{\cancel{4}}{1+u}\left[-\frac{2y_1}{1+u} e^{-\frac{y_1(1+u)}{2}}\Bigg|_0^\infty + \int_0^\infty \frac{2}{1+u} e^{-\frac{y_1(1+u)}{2}}\,\mathrm{d}y_1 \right] \\
&= \frac{1}{(1+u)^2}\int_0^\infty e^{-\frac{y_1(1+u)}{2}}\,\mathrm{d}y_1 = \frac{1}{(1+u)^2}\left[-\frac{2}{1+u} e^{-\frac{y_1(1+u)}{2}}\right]_0^\infty = \frac{2}{(1+u)^3} \\
f_U(u) &= \begin{cases} \frac{2}{(1+u)^3} & \text{if } u>0 \\ 0 & \text{if otherwise} \end{cases}
\end{aligned}
$$

§24.2 3. The Moment of Generating Functions

Theorem 24.1 ***(Theorem 1)*** *Let $m_X(t)$ and $m_Y(t)$ denote the moment-generating functions of random variables X and Y respectively. If both moment-generating functions exist and $m_X(t) = m_Y(t)$ for all values of t, then X and Y have the same probability distribution*

Theorem 24.2 ***(Theorem 2)*** *Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with moment-generating functions $m_{Y_1}(t), \ldots, m_{Y_n}(t)$ respectively. If $U = \sum_{i=1}^n Y_i$, then $m_U(t) = \prod_{i=1}^n m_{Y_i}(t)$.*

$$
\left(i.e.\ m_{\sum_{i=1}^n Y_i}(t) = \prod_{i=1}^n m_{Y_i}(t) \right)
$$

Proof.

$$
\begin{aligned}
m_U(t) = \mathrm{E}(e^{tU}) &= \mathrm{E}\left(e^{t\sum_{i=1}^n Y_i}\right) \\
&= \mathrm{E}\left(e^{\sum_{i=1}^n tY_i}\right) = \mathrm{E}\left(\prod_{i=1}^n e^{tY_i}\right) = \prod_{i=1}^n \mathrm{E}\left(e^{tY_i}\right) = \prod_{i=1}^n m_{y_i}(t)
\end{aligned}
$$

□

Theorem 24.3 ***(Theorem 3)*** *Let $Y_1, \dots, Y_n$ be independent normally distributed random variables with $\mathrm{E}(Y_i) = \mu_i$ and $\mathrm{Var}(Y_i) = \sigma_i^2$, for $i = 1, 2, \dots, n$, and let $a_1, a_2, \dots, a_n$ be constants. If $U = \sum_{i=1}^n a_i Y_i$, then*

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

Proof. We first recall that if $Y_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, then $m_{Y_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}$. Now, using Theorem 2 we have

$$
\mu_U(t) = \prod_{i=1}^n m_{a_iY_i}(t)
$$

We also note that

$$
m_{a_iY_i}(t) = \mathrm{E}\left(e^{ta_iY_i}\right) = m_{Y_i}(a_i t)
$$

Thus,

$$
\begin{aligned}
m_U(t) = \prod_{i=1}^n m_{Y_i}(a_i t) &= \prod_{i=1}^n \exp\left\{\mu_i(a_i t) + \frac{\sigma_i^2(a_i t)^2}{2}\right\} \\
&= \prod_{i=1}^n \exp\left\{a_i\mu_i t + \frac{a_i^2\sigma_i^2 t^2}{2}\right\} \\
&= \exp\left\{t\left(\sum_{i=1}^n a_i\mu_i\right) + \frac{t^2}{2}\left(\sum_{i=1}^n a_i^2\sigma_i^2\right)\right\}
\end{aligned}
$$

Now, comparing this last expression to that of the moment-generating function of the Normal distribution, and using Theorem 1, we obtain

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

□

Theorem 24.4 ***(Theorem 4)*** *Let $Y_1, Y_2, \dots, Y_n$ be defined as in Theorem 3, and define $\mathscr{L}_i$ by $\mathscr{L}_i = \frac{Y_i - \mu_i}{\sigma_i}$, $i = 1, 2, \dots, n$. Then $\sum_{i=1}^n \mathscr{L}_i^2 \sim \chi_n^2$.*

Proof. The result follows easily from Theorems 1 & 2 if we show that $\mathscr{L}_i \sim \chi_1^2$. We first need to show that $\mathscr{L}_i = \frac{Y_i - \mu_i}{\sigma_i} \sim \mathrm{N}(0,1)$.

$$\begin{aligned}
m_{\mathscr{L}_i}(t) &= \mathrm{E}\left(e^{t\mathscr{L}_i}\right) = \mathrm{E}\left(e^{t\frac{Y_i-\mu_i}{\sigma_i}}\right) = \mathrm{E}\left(e^{-\frac{t\mu_i}{\sigma_i}} \cdot e^{\frac{t}{\sigma_i}Y_i}\right) \\
&= e^{-\frac{t\mu_i}{\sigma_i}}\,\mathrm{E}\left(e^{\frac{t}{\sigma_i}Y_i}\right) \\
&= e^{-\frac{t\mu_i}{\sigma_i}}\, m_{Y_i}\left(\frac{t}{\sigma_i}\right) \\
&= e^{-\frac{t\mu_i}{\sigma_i}} \cdot e^{\frac{t}{\sigma_i}\mu_i + \frac{\sigma_i^2}{2}\left(\frac{t}{\sigma_i}\right)^2} = e^{\frac{t^2}{2}}
\end{aligned}$$

But the last expression is the moment-generating function of $\mathrm{N}(0,1)$.

Now, $\mathscr{L}_i \sim \mathrm{N}(0,1)$. Define $W_i = \mathscr{L}_i^2$, so

$$\begin{aligned}
F_{W_i}(w) &= F_{\mathscr{L}_i}(w^{1/2}) - F_{\mathscr{L}_i}(-w^{1/2}) \\
&= \int_{-\infty}^{w^{1/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x - \int_{-\infty}^{-w^{1/2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x \\
\frac{\mathrm{d}F_{W_i}(w)}{\mathrm{d}w} &= \frac{1}{2}w^{-1/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-w/2} + \frac{1}{2}w^{-1/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-w/2} \\
&= \frac{1}{\sqrt{2\pi}} w^{-\frac{1}{2}} e^{-\frac{w}{2}}
\end{aligned}$$

Since $\Gamma(1/2) = \sqrt{\pi}$,

$$= \frac{1}{2^{1/2}\Gamma(1/2)} w^{1/2-1} e^{-\frac{w}{2}} \sim \Gamma(\alpha = 1/2, \beta = 2) = \chi_1^2$$

Second method:

$$m_{W_i}(t) = \mathrm{E}\left(e^{tW_i}\right) = \mathrm{E}\left(e^{t\mathscr{L}_i^2}\right) = \int_{-\infty}^{\infty} e^{t\mathscr{L}_i^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathscr{L}_i^2}{2}}\,\mathrm{d}\mathscr{L}_i = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathscr{L}_i^2}{2}(1-2t)}\,\mathrm{d}\mathscr{L}_i$$

Set $b = \left(\frac{1}{1-2t}\right)^{1/2}$, $b > 0$ so $t < 1/2$ and

$$m_{W_i}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathscr{L}_i^2}{2}\left(\frac{1}{b^2}\right)}\,\mathrm{d}\mathscr{L}_i = b\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}b} e^{-\frac{\mathscr{L}_i^2}{2b^2}}\,\mathrm{d}\mathscr{L}_i = b = \left(\frac{1}{1-2t}\right)^{1/2}$$

But this is the moment-generating function of χ_1^2.

Thus, using Theorem 2,

$$m_{\sum_{i=1}^n \mathscr{L}_i^2}(t) = \prod_{i=1}^n m_{\mathscr{L}_i^2}(t) = (1-2t)^{-n/2}$$

and hence Theorem 1 implies that

$$\sum_{i=1}^n \mathscr{L}_i^2 \sim \chi_n^2$$

$\square$

§25 Lec. 25 - Mon. Dec. 2, 2013

Do # 6.52 on page 324:

(a) $Y_i \sim \text{Po}(\lambda_i)$, $i = 1, 2, \dots$ $Y_1 \perp Y_2$

$$m_{Y_1+Y_2} = m_{Y_1}(t) \cdot m_{Y_2}(t) = e^{[\lambda_1(e^t-1)]} \cdot e^{[\lambda_2[e^t-1]]} = \exp\{(\lambda_1+\lambda_2)(e^t-1)\}$$

This implies that $Y_1 + Y_2 \sim \text{Po}(\lambda_1 + \lambda_2)$

(b) $p_{Y_1|Y_1+Y_2}(y_1 \mid Y_1+Y_2 = m) = P(Y_1 = y_1 \mid Y_1+Y_2 = m) = \begin{cases} \frac{P(Y_1=y_1, Y_1+Y_2=m)}{P(Y_1+Y_2=m)} & \text{if } y_1 = 0, 1, 2, \dots, m \\ 0 & \text{if otherwise} \end{cases}$

$$\begin{aligned} \frac{P(Y_1 = y_1, Y_1+Y_2 = m)}{P(Y_1+Y_2 = m)} &= \frac{P(Y_1 = y_1, Y_2 = m - y_1)}{P(Y_1+Y_2 = m)} = \frac{P(Y_1 = y_1)P(Y_2 = m - y_1)}{P(Y_1+Y_2 = m)} \\ &= \frac{\frac{\lambda_1^{y_1} e^{-\lambda_1}}{y_1!} \cdot \frac{\lambda_2^{m-y_1} e^{-\lambda_2}}{(m-y_1)!}}{\frac{(\lambda_1+\lambda_2)^m e^{-(\lambda_1+\lambda_2)}}{m!}} \\ &= \binom{m}{y_1} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{y_1} \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{m-y_1} \\ &= \binom{m}{y_1} \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{y_1} \left[1 - \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)\right]^{m-y_1} \end{aligned}$$

i.e. $Y_1 \mid_{Y_1+Y_2} \sim \text{Bin}\left(m, p = \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$

§25.1 Chapter 7: Sampling Distributions and the Central Limit Theorem (CLT)

§25.1.1 1. Sampling Distribution

Definition 25.1 A statistic is a function of the observable random variables in a sample and known constants.

Statistics are used ot make inferences (estimates or decisions) about unknown population parameters. As a statistic is a function of the random variables observed in the sample, it is itself a random variable. Consequently, we can derive, using the methods discussed in chapter 6, its probability distribution, which we will call the sampling distribution of the statistic.

Definition 25.2 A sample $Y_1, \dots, Y_n$ is called a random sample if the variables are all independent and have a common distribution function.

Example 25.1 $Y_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, $i = 1, 2, \dots, n$. $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ is a statistic and μ and σ^2 are unknown parameters.

Recall the following results from the previous lectures:

1. If $Y_1 \sim \text{N}(\mu, \sigma^2)$, $i = 1, 2, \dots, n$ is a random sample, then

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

2. If $Y_i \sim \mathrm{N}(\mu, \sigma^2)$, $i = 1, 2, \ldots, n$ is a random sample, then

$$\sum_{i=1}^{n} \mathscr{L}_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

Definition 25.3 Let Y_i, $i = 1, 2, \ldots, n$ be n random variables. The joint mgf of Y_i, $i = 1, 2, \ldots, n$ is defined by

$$m_{\underset{\sim}{Y}}(\underset{\sim}{t}) = m_{Y_1,\ldots,Y_n}(t_1, \ldots, t_n) = \mathrm{E}\left(e^{\underset{\sim}{t}^T \cdot \underset{\sim}{Y}} \right) = \mathrm{E}\left(e^{\sum_{i=1}^{n} t_i Y_i} \right)$$

where $\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ and $\underset{\sim}{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}$, $\underset{\sim}{t}^T$ is the transpose of $\underset{\sim}{t}$.

See example 5.154 on page 291 (Assignment # 4).

Theorem 25.2 ***(Theorem 1)*** *Let X_1 and X_2 be two random variables. Then $X_1 \perp X_2$ if and only if $m_{X_1,X_2}(t_1, t_2) = m_{X_1}(t_1) \cdot m_{X_2}(t_2)$.*

Theorem 25.3 ***(Theorem 2)*** *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variable σ^2. Then*

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

has a $\chi^2_{(n-1)}$. Also, $\bar{Y}$ and S^2 are independent random variables.

Sketch of the proof. We first prove that $\bar{Y}$ and S^2 are independent. To do this, we can use Theorem 1 and multivariate transformations (Chapter 6.6, which we did discuss in this class). Then one uses the following identity:

$$\sum_{i=1}^{n} (Y_i - \mu)^2 = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 + n(\bar{Y} - \mu)^2$$

dividing both sides by σ^2, we have

$$\underbrace{\sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma} \right)^2}_{=:U} = \underbrace{\sum_{i=1}^{n} \left(\frac{Y_i - \bar{Y}}{\sigma^2} \right)}_{=:V} + \underbrace{\frac{(\bar{Y} - \mu)^2}{\sigma^2/n}}_{=:W}$$

Then $U = V + W$. Using $Y \perp S^2$, we have $V \perp W$ and hence

$$m_U(t) = m_V(t) \cdot m_W(t)$$

which implies

$$m_V(t) = \frac{m_U(t)}{m_W(t)}$$

On the other hand, $U \sim \chi_n^2$ and $W = \dfrac{(\bar{Y} - \mu)}{\sigma^2/n} \sim \chi_1^2$. Thus,

$$m_V(t) = \frac{\left(\frac{1}{1-2t} \right)^{n/2}}{\left(\frac{1}{1-2t} \right)^{1/2}} = \left(\frac{1}{1 - 2t} \right)^{n - 1/2}$$

It then follows that $V \sim \chi^2_{(n-1)}$. $\square$

This result is important in the design and analysis of experiments.
Two other useful densities:

1. Let $\mathscr{L} \sim \mathrm{N}(0,1)$, $W \sim \chi^2_\nu$ and $\mathscr{L} \perp W$, then

$$T_V = \frac{\mathscr{L}}{\sqrt{\frac{W}{\nu}}}$$

is said to have a t distribution with ν degrees of freedom. This result and the t test are important.

2. Let $W_i \sim \chi^2_{(\nu_i)}$, for $i = 1,2$ and $W_1 \perp W_2$. Then

$$F_{\nu_1,\nu_2} = \frac{W_1/\nu_1}{W_2/\nu_2}$$

is said to have an F distribution which ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom. This result is important when testing a hypothesis like $H_o : \sigma_1 = \sigma_2$.

§26 Lec. 25 - Tues. Dec. 3, 2013

§26.1 Central Limit Theorem (CLT)

Theorem 26.1 ***(Theorem 3)*** *Let Y_n and Y be random variables with mgf's $m_n(t)$ and $m(t)$ respectively. If $\lim_{n\to\infty} m_n(t) = m(t)$ for all t, then the distribution function of Y_n converges to the distribution function of Y as $n \to \infty$.*

Theorem 26.2 ***(Theorem 4, CLT)*** *Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random variables with $\mathrm{E}[Y_i] = \mu$ and $\mathrm{Var}[Y_i] = \sigma^2 < \infty$. Define $U_n = \sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right)$ where $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$. Then the distribution function of U_n converges to the standard normal distribution function as $n \to \infty$.*

Proof. We first notice that $U_n = \sqrt{n}\left(\frac{\bar{Y}-\mu}{\sigma}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathscr{L}_i$ where $\mathscr{L}_i = \frac{Y_i-\mu}{\sigma}$, $\mathrm{E}[\mathscr{L}_i] = 0$ and $\mathrm{Var}[\mathscr{L}_i] = 1$.

$$m_n(t) = m_{U_n}(t) = m_{\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathscr{L}_i}(t) = m_{\sum_{i=1}^n \mathscr{L}_i}\left(\frac{t}{\sqrt{n}}\right) = \left[m_{\mathscr{L}}\left(\frac{t}{\sqrt{n}}\right)\right] \tag{I}$$

as the $\mathscr{L}_i$'s are iid. Next, we note that

$$m_{\mathscr{L}}(t) = \sum_{k=0}^{\infty} \mu'_k \cdot \frac{t^k}{k!}$$

where $\mu'_k = \mathrm{E}[\mathscr{L}^k]$

$$= 1 + 0 \cdot \frac{t^1}{1!} + 1 \cdot \frac{t^2}{2!} + \mu'_3 \cdot \frac{t^3}{3!} + \dots \tag{II}$$

as $\mathrm{E}(\mathscr{L}_i) = 0$ and $\mathrm{Var}(\mathscr{L}_i) = 0$.

Using I and II, we obtain

$$\begin{aligned} m_n(t) &= \left(1 + \frac{\left(t/\sqrt{n}\right)^2}{2!} + \mu'_3 \frac{\left(t/\sqrt{n}\right)^3}{3!} + \cdots\right)^n \\ &= \left[1 + \frac{t^2}{2n} + \frac{t^3}{3! n^{3/2}}\right]^n \end{aligned}$$

Taking the log, we have

$$\log m_n(t) = n \log[1 + x]$$

where $x = \frac{t^2}{2n} + \frac{t^3}{3! n^{3/2}} \mu'_3 + \dots$.

On the other hand,

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \approx x \text{ as } x \to 0$$

Letting $x = \frac{t^2}{2n} + \frac{t^3}{3! n^{3/2}} \mu'_3 + \dots$, we obtain

$$\log m_n(t) = n \log(1 + x) = nx = n \left(\frac{t^2}{2n} + \frac{t^3}{3! n^{3/2}} \mu'_3 + \dots\right) \xrightarrow{n \to \infty} \frac{t^2}{2}$$

Since the other terms tend to zero as $n \to \infty$.

Thus, $\lim_{n \to \infty} \log[m_n(t)] = \frac{t^2}{2}$, which implies

$$\lim_{n \to \infty} m_n(t) = e^{t^2/2}$$

(Recall that $\lim_{x \to x_0} f(g(x)) = f(\lim_{x \to x_0} g(x))$ if $\lim_{x \to x_0} g(x) = \ell$ exists and $f(x)$ is continuous at ℓ.)

But $e^{t^2/2}$ is the mgf for a standard normal random variable. It then follows from Theorem 3 that the distribution function of U_n converges to the distribution function of a standard normal random variable. □

§26.2 3. Application to Binomial Distribution

Recall that if $Y = \sum_{i=1}^n X_i$ where

$$X_i = \begin{cases} 1 & \text{if } i^{\text{th}} \text{ trial results in success} \\ 0 & \text{if otherwise} \end{cases}$$

with $(X_i = 1) = p$ and X_i's are a random sample, then $Y \sim \text{Bin}(n, p)$, and $\frac{Y}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$.

Now CLT:

$$U_n = \sqrt{n}\left(\frac{\bar{X} - p}{\sqrt{pq}}\right) \overset{\text{app}}{\sim} \text{N}(0, 1)$$

Applications

1. This result can be used to find confidence intervals
2. Do ex. 7.73 on page 384

   $Y \sim \text{Bin}(160, 0.95)$

   $$P(Y \leq 155) = P\left(\frac{Y}{160} \leq \frac{155}{160}\right) = P\left(\bar{X} \leq \frac{155}{160}\right) = P\left[\sqrt{160}\left(\frac{\bar{X} - 0.95}{\sqrt{(0.95)(0.05)}}\right) \leq \sqrt{160}\left(\frac{\frac{155}{160} - 0.95}{\sqrt{(0.95)(0.05)}}\right)\right]$$
   $$= P(U_{160} \leq 1.09) = \int_{-\infty}^{1.09} \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathscr{L}^2}{2}}\, \mathrm{d}\mathscr{L} = \Phi(1.09)$$

   Note: in the exam you should write your answer in terms of Φ if you do not have the table for normal probability distribution.

   How large is large enough?

   $$0 \leq p \pm 3\sqrt{pq/n} < 1 \text{ or equivalently } n > q\left(\frac{\text{larger of } p \text{ and } q}{\text{smaller of } p \text{ and } q}\right)$$

§27 Glossary

1.1 Definition – Set . . . 3
1.2 Definition – Disjoint Set . . . 4
1.3 Definition – Experiment . . . 5
1.4 Definition – Countable . . . 5
2.1 Definition – Limit . . . 5
2.2 Definition – Permutation . . . 7
4.1 Definition – Conditional Probability . . . 10
4.2 Definition – Independent . . . 11
5.1 Definition – Partition . . . 14
6.1 Definition – Random Variable . . . 16
6.2 Definition – Discrete . . . 16
7.1 Definition – Expected Value . . . 20
7.2 Definition – Variance, Standard Deviation . . . 21
11.1 Definition – Moment Generating Function . . . 37
12.1 Definition – Distribution Function . . . 38
12.2 Definition – Continuous Random Variable . . . 39
12.3 Definition – Probability Density Function . . . 39
13.1 Definition – Expected Value of a Continuous Random Variable . . . 40
13.2 Definition – Continuous Uniform Distribution . . . 42
15.1 Definition – Normal distribution . . . 43
15.2 Definition – Gamma Distribution . . . 45
16.1 Definition – Chi-Square Distribution . . . 48
16.2 Definition – Exponential distribution . . . 48
17.1 Definition – Beta probability distribution . . . 48
17.2 Definition – Joint probability distribution . . . 50
17.3 Definition – Joint distribution function . . . 51
18.1 Definition – Jointly continuous, Joint probability density function . . . 51
18.2 Definition – Marginal probability distribution . . . 53
19.1 Definition – Independent random variables . . . 56
19.2 Definition – Expected value . . . 56
20.1 Definition – Covariance . . . 57
21.1 Definition – The Multinomial Experiment . . . 60
22.1 Definition – Conditional expectation . . . 61
22.2 Definition – Conditional variance . . . 61
25.1 Definition – Statistic . . . 70
25.2 Definition – Random sample . . . 70
25.3 Definition – Joint moment generating function . . . 71