Table Of Contents

1. DeMorgan Laws
   - compliments
   - mutually exclusive
2. Modern Definition of Probability
   - Additional Rules
   - Conditional Probability
3. Multiplication Rule
   - Total Probability Rule
4. Systems
   - Probability Part 3 - Counting
     ↳ Multiplication Principle
     ↳ Permutations, Combinations
5. Random Variables
6. Properties of a PMF
   - Properties of a CDF
7. Mean, Expected Value, Variance, Computational Formula, Standard Deviation
8. Binomial Distribution
   - Bernoulli Trials
   - Geometrics Model
9. Poisson Approximation
   - negative binomial distribution
   - random variables pt 3 - Continuous Models
     ↳ properties of a PDF
10. Poisson Process with rate λ
11. Erlang Distribution, Normal Distribution, Probability Density
12. STATISTICS
    - Point Estimation
    - Sample Mean Formula
    - Sample Variance Formula
13. Confidence Intervals, Estimators, calculating n. Case I, II, III
14. Descriptive Statistics, Interquartile Range (IQR), Boxplots
15. Hypothesis Testing (Test Statistic, Critical Region Type II, I Errors
16. P-Value
17. Simple Linear Regression, Covariance, Cauchy-Swartz Inequality

MAT 2377 3x — May 3rd 2011

Probability and Statistics for Engineers

- Read notes on the web
- Probability Part I:

Random Experiment:

Definition: A process where we cannot predict the outcome with certainty

Definition: A random experiment has a <u>sample space</u> S which is the set of all possible outcomes

Definition: An event is a subset of possible outcomes

A few special events:

(a) the sample space S is called the certain event

(b) $\emptyset$ = empty set
called the impossible event

Let x be an outcome, that belongs to S ($x \in S$)

Let A be an event that is $A \subseteq S$
subset of

Let x be the observed outcome

Terminology: if $x \in A$, then we say that A has occured

Example 1: Example of a few random experiments

1. Selection of a component and verification of its compliance
   $S = \{\text{compliant, non-compliant}\}$

2. Lifetime of a computer
   $S = \{t \mid t \geq 0\}$

3. # of calls to a system during a fixed interval of time
   $S = \{0, 1, 2, \ldots\}$

4. Selection of 2 tools without replacement among a box of 3 tools; say $\{A, B, C\}$

$S = \{AB, AC, BA, BC, CA, CB\}$

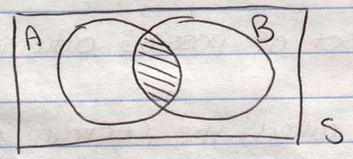
5. Selection of 3 tools without replacement among $\{A, B\}$

$S = \{AAA, AAB, ABA, BAA, ABB, BBA, BAB, BBB\}$

<u>Operation on Events:</u>

a) Intersection of A and B

$A \cap B$ occurs $\equiv$ A occurs <u>and</u> B occurs

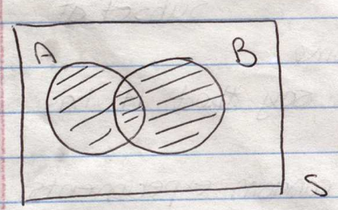


$A \cap B$

$A_1 \cap A_2 \cap \ldots \cap A_n$ occurs $\equiv$ all events $A_1, A_2, \ldots, A_n$ occurs

b) Union of A and B

$A \cup B$ occurs $\equiv$ A occurs <u>or</u> B occurs <u>or</u> A and B occur (both)

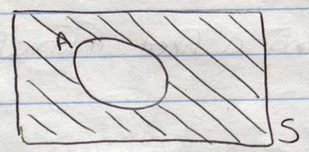


$A \cup B$

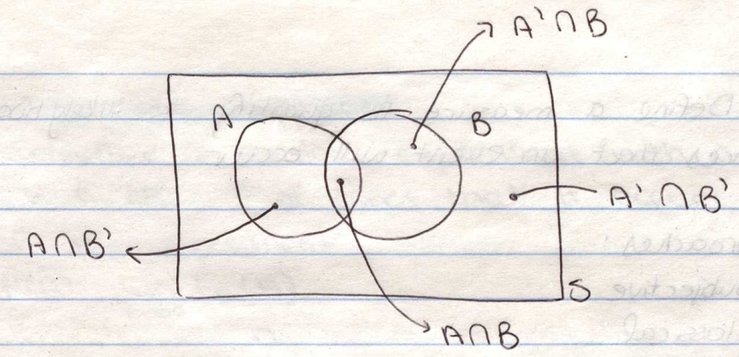
$A_1 \cup A_2 \cup \cdots \cup A_n$ occurs $\equiv$ at least one of $A_1, \ldots, A_n$ occurs

c) Complement of A

A' occurs $\equiv$ A does not occur



A'



* but = and

<u>De Morgan Laws:</u>

a) $(A_1 \cup A_2 \cup \ldots \cup A_n)'$ occurs $\equiv$ none of $A_1, \ldots, A_n$ has occured

$\equiv A_1' \cap A_2' \cap \ldots \cap A_n'$ occurs

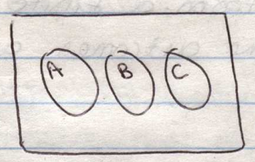
The complement of the union is the intersection of the complements

b) $(A_1 \cap A_2 \cap \ldots \cap A_n)'$ occurs $\equiv$ at least one has <u>not</u> occured

$\equiv A_1' \cup A_2' \cup \ldots \cup A_n'$

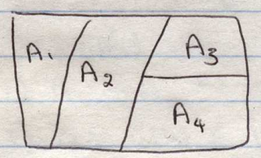
The compliment of the intersection is the union of the compliment

Definition: We say that $A_1, \ldots, A_n$ are mutually exclusive if $A_i \cap A_j = \emptyset$ when $i \neq j$



A, B, C are mutually exclusive

Defintion: We say that $A_1, \ldots, A_n$ are exhaustive if $A_1 \cup A_2 \cup \ldots \cup A_n = S$



A_1, A_2, A_3, A_4 are mutually exclusive and they are also exhaustive

Goal: Define a measure to quantify the likelyhood or the chances that an event will occur

3 approaches:
(i) Subjective
(ii) Classical
(iii) frequency approach

(i) Subjective Approach

Consider an event A. We associate a number between 0 and 1 to this event, denoted $P(A)$. The closer to 1 means more likely, closer to 0 means less likely

(ii) Classical approach

Definition: Equally likely model:
Consider a finite sample S with equally likely outcomes
$P(A) = \frac{N(A)}{N(S)}$; where $N(A)$ = # of outcomes in A

Remark: In random selection from a finite collection, we can assume that each of the outcomes are equally likely

Ex: Consider 20 items where 2 are defective. We choose one item at random.
$P(A) = \frac{N(A)}{N(S)} = \frac{2}{20}$
where A = "defective item"

(iii) Frequency Approach

Consider a random experiment with a sample space S, Let $f_n(A)$ = # of times that A occurs in n trials of the experiment

$$P(A) = \lim_{n\to\infty} \frac{f_n(A)}{n}$$

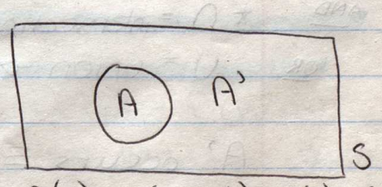
Ex: $\left.\begin{array}{l} n = 100 \text{ trials} \\ f_n(A) = 2 \text{ defective outcomes} \end{array}\right\} P(A) \cong \frac{f_n(A)}{n} = \frac{2}{100} = 0{,}02$

Modern Definition of Probability:
3 axiomes of probability:
Consider a sample space S
A probability is a function P that assigns numbers (real) to events such that:

(i) [Positivity]: $P(A) \geq 0$
(ii) [Certainty]: $P(S) = 1$
(iii) [Additivity]: if $A_1, A_2, \ldots A_n$ are mutually exclusive, then $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots P(A_n)$

<u>Note</u>: All approaches satisfy the 3 axioms, thus any consequence of the axioms will hold true for the approaches

Theorem:
a) $P(\emptyset) = 0$
b) $A_1 \subseteq A_2 \Rightarrow P(A_1) \leq P(A_2)$
c) $0 \leq P(A) \leq 1$
d) $P(A') = 1 - P(A)$



$1 = P(S) = P(A \cup A') = P(A) + P(A')$

May 5

Last lecture:
- → Random experiment
- → Sample Space = S = set of all possible outcomes
- → Event = set of possible outcomes

Recall:
- If A occurs, this means that the observed outcome belongs to A.

Example:
Number of calls at a system in 30 minutes
$S = \{0, 1, 2 \ldots\}$
Let A be the event that there is at most 2 calls
that is $A = \{0, 1, 2\}$
a) If we observe 5 calls, because this is more than 2, A has not occured.

b) If we observe 2 calls, then A has occured

- We discussed operations on events
  $A_1 \cap A_2 \cap \ldots \cap A_n$ occurs
  $\equiv$ all the events $A_1, A_2 \ldots, A_n$ occur
  $A_1 \cup A_2 \cup \ldots \cup A_n$ occurs $\equiv$ at least one of $A_1, A_2 \ldots A_n$ has occured

AND * $\cap$ = intersection → all has occured
OR $\cup$ = union → at least one has occured.

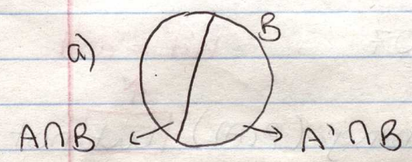
A' occurs $\equiv$ A does not occur
We saw the DeMorgans

- Approaches to Probability
  a) Subjective
  b) Classical
  c) Frequency

- Modern Definition of Probability
  Axioms of Probability
  a) Positivity (all must be ≥ 0)
  b) certainty
  c) Additivity

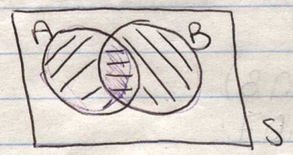
- Theorem
  a) $P(\emptyset) = 0$
  b) $A_1 \subseteq A_2 \Rightarrow P(A_1) \leq P(A_2)$
  c) $0 \leq P(A) \leq 1$
  d) $P(A') = 1 - P(A)$

- Additional Rules

a)


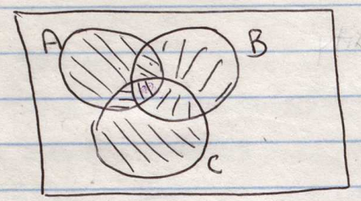
$P(B) = P(A \cap B) + P(A' \cap B)$

$P(B \cap A') = P(B) - P(A \cap B)$ → probability that B will occur but not A

b)


A∪B

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
→ inclusion and exclusion theory.

c) 

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

d) $P(A_1 \cup A_2 \cup \ldots \cup A_n) = 1 - P([A_1 \cup A_2 \cup \ldots \cup A_n]')$
$= 1 - P(A_1' \cap A_2' \cap \ldots \cap A_n)$

by De Morgan.

Example 2 from notes
→ probability of defective etching is 0,12 (12%)
→ probability of a defective slot is 0,29 (29%)
→ probability of both defects is 0,07 (7%)

Let A = defective etching
and B = defective slot
we know $P(A) = 0{,}12$
$P(B) = 0{,}29$
$P(A \cap B) = 0{,}07$

* means and *

but = and

a) probability of defective etching (P(A)) $\boxed{\text{but}}$ not a defective slot

$P(A \cap B') = P(A) - P(A \cap B)$
$= 0{,}12 - 0{,}07$
$= 0{,}05$

b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0{,}12 + 0{,}29 - 0{,}07$
$= 0{,}34$

c) $P(A' \cap B') = 1 - P(A \cup B)$
$= 1 - 0{,}34$
$= 0{,}66$

Conditional Probability

Definition: Let B be an event such that $P(B) > 0$. The conditional probability that A will occur given that B occurs is $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Remark: Consider the equally likely model:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{N(A \cap B)/\cancel{N(S)}}{N(B)/\cancel{N(S)}} = \frac{N(A \cap B)}{N(B)}$$

Example

defective	Factory A	Factory B	Total
yes	25	50	75
no	32	100	132
Total	57	150	207

If we select an item at random from these 207 items, what is the probability that it is defective given that it is from Factory A.

→ Let D = defective

$P(D|A) = \dfrac{25}{57}$

Example 3 from the notes

20 light bulbs such that 2 are defective. Choose 2 bulbs without replacement. Compute the probability that the second is defective given that the first is not defective

D	D'
2	18

(20)

D	D'
2	17

(19)

$$P(D_2 \mid D_1') = \frac{2}{19}$$

Let D_i = i^{th} bulb is defective

Multiplication Rule

Recall: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

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

Note: $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$ by definition

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

Example 4 from notes

20 bulbs such that 2 are defective. Choose 2 without replacement. Probability that both are not defective?

$$P(D_1' \cap D_2') = P(D_2' \mid D_1')\,P(D_1')$$

$$= \frac{17}{19} \cdot \frac{18}{20}$$

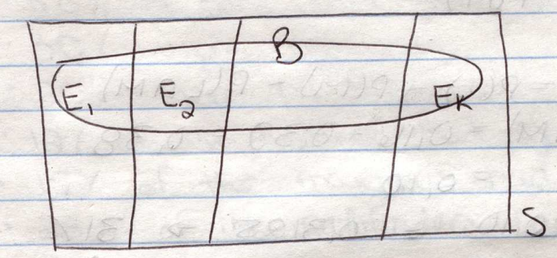
you take 1 out, so you're left with 19, but 2 are still defective, so there are 17 that still work over the total → 17/19

from the initial 20 bulbs, 2 are defective so 18 are left, over the total → 18/20

$= 0.80526$

$= 80\%$

Total Probability Rule



$E_1, \ldots, E_k$ are mutually exclusive and extraustive

Given information:

$P(B|E_1), \ldots, P(B|E_k)$

$P(E_1), P(E_2), \ldots, P(E_k)$

$$P(B) = P(B \cap E_1) + P(B \cap E_2) + \ldots + P(B \cap E_k)$$
$$= P(B|E_1)P(E_1) + \ldots + P(B|E_k)P(E_k)$$

Suppose we want $P(E_i|B)$ → write it out.

$$P(E_i|B) = \frac{P(E_i \cap B)}{P(B)} \quad \text{by definition}$$

$$= \frac{P(B|E_i)P(E_i)}{P(B|E_1)P(E_1) + \ldots + P(B|E_k)P(E_k)} \quad \text{multiplication rule} \quad \text{total prob. rule}$$

The latter is known as Bayes' Theorem

Example 5 from the notes

- 32% contain toxic levels of mercury
- 16% contain toxic levels of lead
- 38% contain at least one of the two

$P(M) = 0.32$

$P(L) = 0.16$

$P(M \cup L) = 0.38$

1. $P(L|M) = \dfrac{P(L \cap M)}{P(M)}$

<u>Note:</u> $P(L \cup M) = P(L) + P(M) - P(L \cap M)$

$\Rightarrow P(L \cap M) = 0{,}16 + 0{,}32 - 0{,}38$

$= 0{,}10$

$\therefore P(L|M) = \dfrac{0{,}10}{0{,}32} = 0{,}3125 \rightarrow 31\%$

the samples that contain mercury also have lead.

2. A = Arsenic

$P(A|M) = 0{,}45$ (given)

We want $P(M \cap A)$

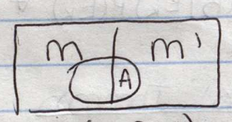
$P(M \cap A) = P(A|M)P(M)$

$= (0{,}45)(0{,}32)$

$= 0{,}144 \rightarrow 14\%$ have both M and A.

3. $P(A|M') = 0{,}02$ (given)

We want $P(A)$



$P(A) = P(A \cap M) + P(A \cap M')$

$= (P(A|M'))(P(M')) + (P(A|M))(P(M))$

$= (0{,}02)(1 - 0{,}32) + (0{,}45)(0{,}32)$

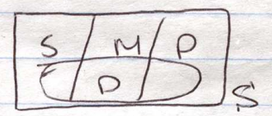
$= 0{,}1576 \rightarrow 15\%$

Example 7 from notes.

- 50% of cars sold were sentras
- 30% were maximas
- 20% were pathfinders
- 12% of the sentras had ignition problems
- 15% of the maximas had problems
- 25% of the pathfinders had problems.

$\rightarrow$ Let D = problems with ignitions

1. we want $P(D)$



$P(D) =$

$P(D|S)P(S) + P(D|M)P(M) + P(D|P)P(P)$

$= (0{,}12)(0{,}5) + (0{,}15)(0{,}3) + (0{,}25)(0{,}2)$

$= 0{,}155 \rightarrow 15{,}5\%$ of these cars have problems with the ignition.

May 10.

- Assignment 1 due May 19.

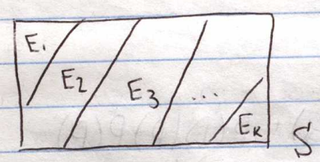
Last Lecture

- Refer to the notes for Probability Part II

Definition: probability that A will occur given that B occurs is $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Multiplication Rule

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



$E_1, \ldots, E_k$ are mutually exclusive (all different) and are exhaustive events

Given: $P(B|E_i)$ for $i = 1, \ldots, k$
$P(E_i)$ for $i = 1, \ldots, k$

Want: $P(B)$

in B and E_1 in B and E_2

$P(B) = P(B \cap E_1) + P(B \cap E_2) + \ldots + P(B \cap E_k)$
$= P(B|E_1)P(E_1) + P(B|E_2)P(E_2) + \ldots P(B|E_k)P(E_k)$

Total Probability Rule

Bayes' Rule (reverses the total prob rule)

$P(E_i|B) = \frac{P(E_i \cap B)}{P(B)}$ → by definition.

$= \frac{P(B|E_i)P(E_i)}{P(B|E_i)P(E_i) + \ldots + P(B|E_k)P(E_k)}$ ← multiplication rule (numerator) ← total prob rule

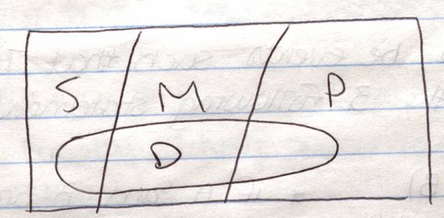
Example 7 ctd

50% were sentras
30% were Maximas
20% were Pathfinders

Furthermore,

12% of sentras had engine problems,
15% of Maximas
25% of Pathfinders had problems.

1. Find the probability that a 1999 Nissan car had a problem with the ignition.



Let D = "problem with the ignition system."

We want $P(D) = P(D \cap S) + P(D \cap M) + P(D \cap P)$
but we're given the conditional prob, so using the multiplication rule, we can figure it out

$P(D) = P(D|S)P(S) + P(D|M)P(M) + P(D|P)P(P)$
$= (0.12)(0.5) + (0.15)(0.3) + (0.25)(0.2)$
$= 0.155$
↳ 15.5% of 1999 Nissans had ignition problems

2. What is the probability that a pathfinder had this defect?

probability it had the defect given it was a pathfinder.

$$P(P|D) = \frac{P(P \cap D)}{P(D)} = \frac{P(D|P)\,P(P)}{P(D)}$$

$$= \frac{(0{,}25)(0{,}2)}{(0{,}155)}$$

$$= 0{,}3226$$

$\Rightarrow$ 32.26% of the 1999 Nissan cars that had ignition problems were pathfinders

---

Definition: Consider 2 events A and B. We say that they are independant if $P(A \cap B) = P(A)P(B)$

Theorem: Let A and B be events such that $P(A) > 0$ and $P(B) > 0$. Then the 3 following statements are equivalent:

1. $P(A \cap B) = P(A)P(B)$ * if A and B are independant
2. $P(A|B) = P(A)$ $\Rightarrow$ prob of A and B occuring $= P(A)$
3. $P(B|A) = P(B)$ $\Rightarrow$ prob of B and A occuring $= P(B)$

Recall:
$P(A|B)$ = conditional probability that A will occur given that B occurs.

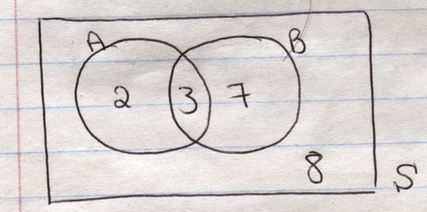
Refer to the last example
$P(D) = 0{,}155$
$P(D|P) = 0{,}25$
Since $P(D) \neq P(D|P)$, then D and P are INDEPENDANT events.

Definition: We say that the events $A_1, \dots, A_k$ are independant if all subcollections $A_{i_1}, A_{i_2}, \dots, A_{i_n}$ we have that the probability of intersection is equal to the product of probabilities
$P(A_{i_1} \cap \dots \cap A_{i_n}) = P(A_{i_1}) \dots P(A_{i_n})$

Remark: If we have a collection of independant events, and we substitute any event by its complement, then we get a new collection of independant events

Example:
If A and B are independant, then
i) A', B are independant
ii) A', B' are independant
iii) A, B' are independant

Example 8 from the notes
Equally likely outcomes with $N(S) = 20$
$N(A) = 5$
$N(B) = 10$
$N(A \cap B) = 3$



Are A and B independant?

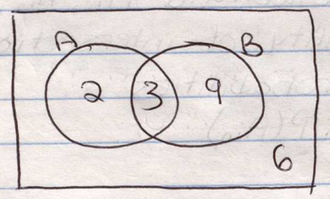
$P(A \cap B) = \frac{N(A \cap B)}{N(S)} = \frac{3}{20}$

$P(A)P(B) = (5/20)(10/20) = \frac{50}{400}$

Since $3/20 \neq 50/400$ then A and B are dependant events

Example 9 from the notes

$N(S) = 20 \quad N(A) = 5 \quad N(B) = 12 \quad N(A \cap B) = 3$



* by looking at the picture, you can <u>not</u> tell if A and B are independant. You must compute the probabilities to find out.

$P(A \cap B) = \frac{3}{20}$

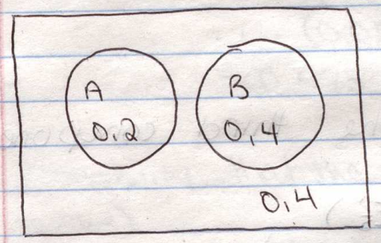
$P(A)P(B) = \left(\frac{5}{20}\right)\left(\frac{12}{20}\right) = \frac{60}{400} = \frac{3}{20}$

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

$\therefore$ A and B are independant.

Note on page 7 from the notes

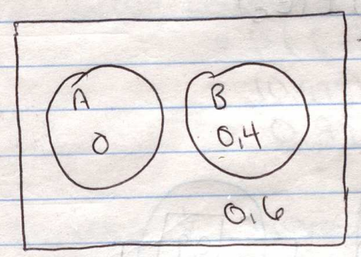
Examples



→ example of <u>mutually exclusive</u>

$P(A \cap B) = P(\emptyset) = 0 \neq P(A)P(B)$

they are not independant but are mutually exclusive.



→ example of mutually exclusive and independant

$P(A \cap B) = 0 = 0(0.4) = P(A)P(B)$

but this case is rare because the Prob that $A = 0$ is very unlikely.

Example 10 from the notes

Probability that a computer works is 0.9. Consider 5 independant components

a) What is the prob that all 5 of the components will work?

Solution: Let E_i = "component i works" for $i = 1, \dots, 5$

We know that $P(E_i) = 0.9$

We want $P(E_1 \cap E_2 \cap E_3 \cap E_4 \cap E_5)$

$= P(E_1)P(E_2)P(E_3)P(E_4)P(E_5)$ (by independance)

$= (0.9)^5$

$= 0.59049$

b) Compute the probability that at least 1 component will work (we're interested in the union)

$\rightarrow P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5)$

$= 1 - P((E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5)')$

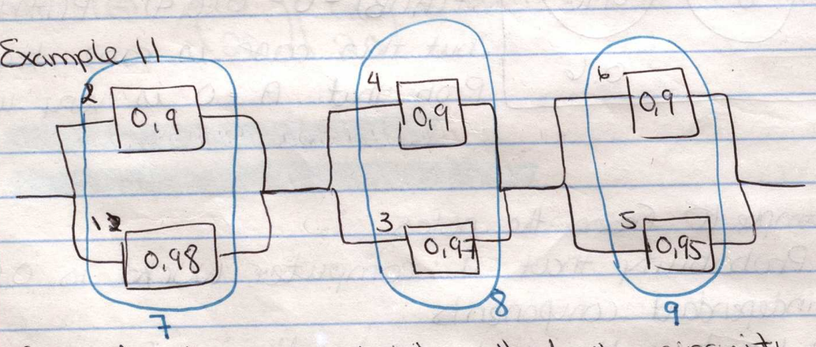
$= 1 - P(E_1' \cap E_2' \cap E_3' \cap E_4' \cap E_5')$ (by deMorgan)

$= 1 - P(E_1')P(E_2')P(E_3')P(E_4')P(E_5')$ (by independance)
$= 1 - (0{,}1)^5$
$= 0{,}99999$

c) Probability that only the third component will work.

$P(E_1' \cap E_2' \cap E_3 \cap E_4' \cap E_5')$
$= P(E_1')P(E_2')P(E_3)P(E_4')P(E_5')$
$= (0{,}1)^4(0{,}9)$
$= 0{,}00009$

Example 11



Compute the probability that the circuit works

$P(E_7) = P(E_1 \cup E_2)$
$= 1 - P[(E_1 \cup E_2)']$
$= 1 - P(E_1' \cap E_2')$
$= 1 - P(E_1')P(E_2')$ by independance.
$= 1 - [(1-0{,}9)(1-0{,}98)]$
$= 1 - (0{,}1)(0{,}02)$
$= 0{,}998$

Similarily
$P(E_8) = P(E_3 \cup E_4) = 1 - P(E_3')P(E_4')$
$= 1 - (0{,}03)(0{,}1)$
$= 0{,}997$

$P(E_9) = P(E_5 \cup E_6) = 1 - P[(E_5' \cap E_6')]$
$= 1 - P(E_5')P(E_6')$
$= 1 - (0{,}05)(0{,}1)$
$= 0{,}995$

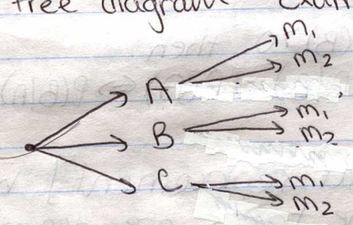
Probability that the circuit works (meaning E_7, E_8, E_9 works)

$P(\text{"circuit works"}) = P(E_7 \cap E_8 \cap E_9)$
$= P(E_7)P(E_8)P(E_9)$ by independance.
$= (0{,}998)(0{,}997)(0{,}995)$
$= 0{,}99003$

series = ↓ chance of the circuit working
parallel = ↑ chance of the circuit working

Probability Part III - Counting

tree diagram: Example 12.



There are 6 possible outcomes:

A & M_1	B & M_2
A & M_2	C & M_1
B & M_1	C & M_2

<u>Multiplication Principle</u>

Consider an experiment which is broken into steps
# of outcomes at step $i = n_i$

k = number of steps
total number of outcomes $= n_1 \times n_2 \times n_3 \times \ldots \times n_k$

Example 13 (refers to ex 12)
3rd step: Choose 1 of 5 units

total number of outcomes $= (3)(2)(5) = 30$

<u>Factorial Notation</u>

$$n! = \begin{cases} 1, & n=0 \\ n(n-1)(n-2)\dots 1 & n \geq 1 \end{cases}$$

May 12/11 → for the midterm - buy the textbook! You need the tables II and III from the appendix

Last lecture: Probability II

Definition: We say that A and B are independant events. If $P(A \cap B) = P(A)P(B)$

Recall: If $P(A) > 0$ and $P(B) > 0$ then
$P(A \cap B) = P(A)P(B) \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$

<u>Example</u>: [Sampling without replacement] We will see that in this case, the trials are dependant. Consider $n=10$ items such that 3 of these items are defective. We sample 2 items at random without replacement. Let E_i = "i^{th} item is defective" for $i = 1, 2$

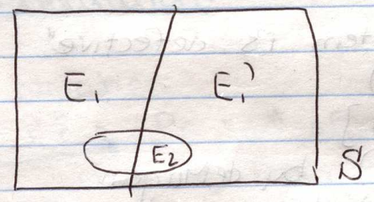
Are E_1 and E_2 independant events?
If so, then $\underbrace{P(E_2|E_1) = P(E_2)}$

we will investigate this equality

$P(E_2|E_1) = 2/9 = 0.222$

Given E_1, which means that for the second trial we have 9 items left such that 2 defective items are left

$$\begin{aligned} P(E_2) &= P(E_2 \cap E_1) + P(E_2 \cap E_1') \\ &= P(E_2|E_1)P(E_1) + P(E_2|E_1')P(E_1') \\ &= (2/9)(3/10) + (3/9)(7/10) \\ &= 0.03 \end{aligned}$$



Since $P(E_2|E_1) \neq P(E_2)$, thus E_1 and E_2 are dependant

b) If we increase the number, $n = 1000$ but we keep the same proportiones for defective items ie $P(E_1) = 0.3$

$P(E_2|E_1) = 299/999 = 0.2993$
and
$$\begin{aligned} P(E_2) &= P(E_2 \cap E_1) + P(E_2 \cap E_1') \\ &= P(E_2|E_1)P(E_1) + P(E_2|E_1')P(E_1') \\ &= (299/999)(300/1000) + (300/999)(700/1000) \\ &= 0.3 \end{aligned}$$

$P(E_2|E_1) + P(E_1)$

$\therefore$ E_1 and E_2 are dependant but the 2 quantities are almost equal. Thus, in this case, the dependance is less important

Remark: When sampling without replacement from a large population, we will ignore the dependance between trials * If you are not given the population size, we assume that it is large *

Example

5% of items are defective, we choose 3 at random. What is the probability that at least 1 is defective?

Solution: Let E_i = " i^{th} item is defective"

We want $P(E_1 \cup E_2 \cup E_3)$

$= 1 - P[(E_1 \cup E_2 \cup E_3)]$

$= 1 - P(E_1' \cap E_2' \cap E_3')$ by deMorgan

$= 1 - P(E_1')P(E_2')P(E_3')$ by independance

$= 1 - (0.95)^3$

$= 0.1426$

Probability Part III - Counting Methods

Recall from last lecture:

Multiplication Principle: We want/have a sequence of steps for our experiment

\# of steps = k

\# of possible outcomes for step $i = n_i$

Then, # of total outcomes $= n_1 \times n_2 \times n_3 \ldots \times n_k$

Factorial Notation:

$$n! = \begin{cases} 1, & n = 0 \\ n(n-1)(n-2)\ldots 1, & n \geq 1 \end{cases}$$

Example

a) $2! = 2 \times 1 = 2$

b) $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$

c) $0! = 1$ → there's only 1 way to choose nothing

Definition: Consider n objects. We select r objects. The selection (order does not matter) is called a combination. If the order does matter, then we have an arrangement of objects, and it's called a permutation.

Notation:

$P^n_r = {}_nP_r$ = # of permutations of size r when choosing from n objects

on calculator

$P^n_r = {}_nP_r = \underset{\text{for first position}}{n}\ \underset{\text{for second position}}{(n-1)} \ldots \underset{\text{for the } r^{th} \text{ position}}{(n-(r-1))}$

$= n(n-1)(n-2)\ldots(n-r+1)$

$= \dfrac{n(n-1)(n-2)\ldots(n-r+1)(n-r)(n-r-1)\ldots 1}{(n-r)(n-r-1)(n-r-2)\ldots 1}$

$= \dfrac{n!}{(n-r)!}$

Combinations

Selection is important, NOT the order.

$C^n_r = {}_nC_r = \binom{n}{r}$

We will show that

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

We will count permutations again

$_nP_r = {}_nC_r \cdot {}_rP_r$

select r objects from n

order the r objects

$$\Rightarrow {}_nC_r = \frac{_nP_r}{_rP_r} = \left[\frac{n!}{(n-r)!}\right] \Big/ \left[\frac{r!}{(r-r)!}\right]$$

*but $0! = 1$

$$\therefore {}_nC_r = \frac{n!}{(n-r)!\,r!}$$

Example 16

Placement of 3 chips. We have access to 5 different chips In how many different ways can we assemble the controller?

Solution:

\# of ways $= 5 \times 4 \times 3$

for the 1st position

for the 2nd position

for the 3rd position

$= 60$ different ways

Alternative solution

$$\# \text{ of ways} = {}_5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = 60$$

Example 17

50 articles where 3 are defective. We select 5 articles at random (we assume it's sampling without replacement)

a) In how many different ways can we choose the 5 articles?

Solution:

\# of ways $= {}_{50}C_5 = \binom{50}{5} = 2\,118\,760$

* selection, select → what you have is important, NOT the order
* arrangement, placement → order IS important

b) In how many different ways can we select the 5 articles, in such a way that exactly 1 is defective?

\# of ways: $\binom{3}{1} \cdot \binom{47}{4}$

$= {}_3C_1 \cdot {}_{47}C_4$

$= (3) \cdot (178\,365)$

$= 535\,095$

c) If we select 5 articles, what is the probability that exactly 1 will be defective?

Solution: probability $= \dfrac{\binom{3}{1}\binom{47}{4}}{\binom{5}{50}}$

$= \dfrac{535\,095}{2\,118\,760}$

$= 0{,}05255$

Example 18

3 types of holes: small, medium, large

In how many ways can we program the machine to obtain 2 small, 2 medium and 3 large holes?

Example of a valid arrangement:

SMSLLLM → order does not matter

By the multiplication principle, # of ways =

$\binom{7}{2} \cdot \binom{5}{2} \cdot \binom{3}{3}$ or ${}_7C_2 \cdot {}_5C_2 \cdot {}_3C_3$

number of ways to choose position for 2 S

number of ways to choose 2 M

number of ways to choose 3 L

$= (21)(10)(1)$

$= 210$

Random Variables

Let S be a sample space. Define a function $X: S \to \mathbb{R}$ that associates real numbers to outcomes.

* Note: X is called a random variable.

We can define events with random variables

Example

- $\{X = 2\} = \{s \in S : X(s) = 2\}$

  all outcomes such that when evaluated, the value = 2

- $\{X \leq 3\} = \{s \in S : X(s) \leq 3\}$

- $A \subseteq \mathbb{R}$ ← subset of numbers

$\{X \in A\} = \{s \in S : X(s) \in A\}$

Classification

Let R_X = range of random variables X

= set of real values for this experiment in terms of X

a) We say that X is discrete if R_X is finite or countable (the range is not ∞)

b) If R_X is an interval of real numbers, then X is continuous.

May 17/11 Last Lecture: Counting from Probability Part III

Multiplication Principle:

k steps and n_i = # of possible outcomes for step i

total # of outcomes $= n_1 \times n_2 \times n_3 \times \ldots \times n_k$

Example

Production of a computer application

Assignment of the job:

Step 1: Choose 1 application from 4

Step 2: Choose 1 program from 3

Step 3: Choose 1 of 3 languages (Java, C++, C#)

# of ways to assign the job $= 4 \times 3 \times 3 = 36$.

a) If the assignment is chosen at random, what is the

probability that Bob will work with Java?

Solution: We know there are 36 different jobs
we define A = " Bob with Java"

$$P(A) = \frac{N(A)}{N(S)} = \frac{4 \times 1 \times 1}{36} = \frac{4}{36}$$

- $_nC_r = \binom{n}{r}$ = # of combinations of size r chosen from n objects

$$= \frac{n!}{(n-r)!\,r!}$$ (*Note: order is not important for combinations)

- $_nP_r$ = # of permutations of size r chosen from n objects

$$= \frac{n!}{(n-r)!}$$ (* Order IS important for permutations)

Random Variables

$X : S \to \mathbb{R}$

a function that assigns real numbers to outcomes is called a Random Variable

Remark: Let R_X be the range of the random variable. If R_X is finite or countable (0,1,2,3...) we say that the random variable X is discrete.

ii) When working with a discrete random variable X we will use two tools to specify probabilities

a) A probability mass function (p.m.f) is a function denoted by f, where $f(x) = P(X = x)$ for $x \in R_X$

b) A cumulative distribution is a function denoted by F such that $F(x) = P(X \leq x)$, $x \in \mathbb{R}$

Example 2 in Random Variable notes
p. first component is defective is 0,1
p. second component is defective is 0,2
the components are independant.

Let X = # of defective components

a) Give the range of X

Solution: $R_X = 0, 1, 2$ ← either none are defective, 1, 2

b) Determine the pmf of X

Solution:
probability that X = 0 → $f_X(0) = P(X = 0)$

Define D_i : " i^{th} component is defective" $i = 1, 2$

$f_X(0) = P(X = 0) = P(D_1' \cap D_2')$ ← probability that 1st is not defective and 2nd isn't either

$= P(D_1')\,P(D_2')$
$= (0,9)(0,8)$
$= 0,72$

$f_X(2) = P(D_1 \cap D_2)$
$= P(D_1)\,P(D_2)$ by independance
$= (0,1)(0,2)$
$= 0,02$

1st is defective but not 2nd — 1st isn't defective but 2nd is

$$f_X(1) = P(X=1) = P((D_1 \cap D_2') \cup (D_1' \cap D_2))$$
$$= P(D_1 \cap D_2') + P(D_1' \cap D_2)$$
$$= (0{,}1)(0{,}8) + (0{,}9)(0{,}2) \text{ by independance}$$
$$= 0{,}26$$

Therefore the pmf of X is

$$f_X(x) = \begin{cases} 0{,}72 & x=0 \\ 0{,}26 & x=1 \\ 0{,}02 & x=2 \end{cases}$$

Alternatively as a table:

x	0	1	2
$f_X(x)$	0,72	0,26	0,02

Properties of a p.m.f.

i) $0 \le f(x) \le 1$ for $x \in R_X$ (must be between 0 & 1)

ii) $\sum_{x \in R_X} f_X(x) = 1$ (the certainty axiom: the sum of all x in the range adds up to 1)

Let us go back to part b)

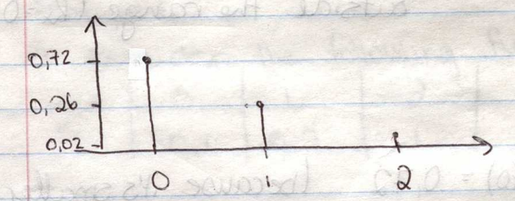
We know $f_X(0) = 0{,}72$, $f_X(2) = 0{,}02$

$$f_X(1) = 1 - (f_X(0) + f_X(2))$$
$$= 1 - (0{,}72 + 0{,}02)$$
$$= 0{,}26$$

Since $\sum_{x \in R_X} f(x) = 1$

Remark: if a function satisfies the above properties (i) and (ii) then it is called a probability mass function.

A stick diagram is a graphical representation of the p.m.f.



d) probability of at least 1 defective

$$= P(\{X=1\} \text{ or } \{X=2\})$$
$$\Rightarrow P(X \geq 1) = P(X=1) + P(X=2)$$
$$= f(1) + f(2)$$
$$= 0{,}26 + 0{,}02$$
$$= 0{,}28$$

e) the p that there are more than 2 defective

$$P(X > 2) = P(\emptyset) = 0$$

f) $P(-0{,}2 < X \le 1{,}5)$ (X is between -0,2 and 1,5)

$$= f(0) + f(1)$$
$$= 0{,}72 + 0{,}26$$
$$= 0{,}98$$

iii) Properties of a pmf again

Let $A \subseteq \mathbb{R}$, then $P(X \in A) = \sum_{x \in A \cap R_X} f(x)$

c) Determine the c.d.f of X

Recall $F_X(x) = P[X \leq x]$ for $x \in \mathbb{R}$

Case I

Let $x < 0$:

$F_X(x) = P[X \leq x] = 0$ (there are no possible masses outside the range ($R_X = 0-2$))

Case II

Let $0 \leq x < 1$

$F_X(x) = P(X \leq x) = f(0) = 0{,}72$ (because it's smaller than 1)

Case III

Let $1 \leq x < 2$

$F_X(x) = P[X \leq x] = f(0) + f(1) = 0{,}98$

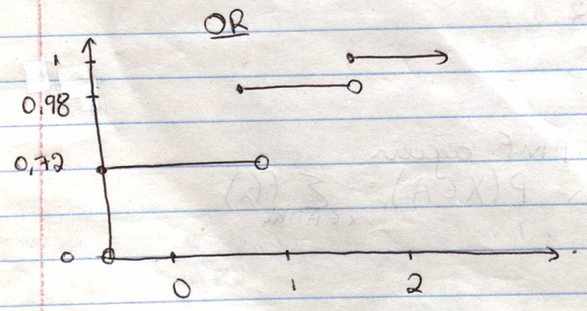
Case IV

Let $2 \leq x$

$F_X(x) = P[X \leq x] = 1$

$$\therefore F_X(x) = \begin{cases} 0 & , x < 0 \\ 0{,}72 & , 0 \leq x < 1 \\ 0{,}98 & , 1 \leq x < 2 \\ 1 & , x \geq 2 \end{cases}$$

OR



Properties of the c.d.f. F:

i) $0 \leq F(x) \leq 1 \quad x \in \mathbb{R}$

ii) a non decreasing function

iii) $\lim_{x \to \infty} F_X(x) = 1$, $\lim_{x \to -\infty} F_X(x) = 0$

Example

a) Consider the following function:

x	0	1	2
f(x)	0,1	0,2	1

Is f a p.m.f.?

$\rightarrow$ since $\sum_{x \in R_X} f(x) \neq 1$, f is not a p.m.f.

b) What about

x	-1	0	4
f(x)	1/2	1/4	1/4

- Since $0 \leq f(x) \leq 1$ for $x \in R_X = \{-1, 0, 4\}$
- $\sum_{x \in R_X} f(x) = 1/2 + 1/4 + 1/4 = 1$

$\therefore$ Therefore, f is a p.m.f.

Example:

Consider the following c.d.f.

$$F(x) = \begin{cases} 0 & , x < -1 \\ 0{,}1 & , 1 \leq x < 0 \\ 0{,}5 & , 0 \leq x < 10 \\ 0{,}75 & , 10 \leq x < 15 \\ 1 & , x \geq 15 \end{cases}$$

Give the p.m.f. of X

→ The range of X is:

$R_x = \{-1, 0, 10, 15\}$

So $f(x) = \begin{cases} 0.1, & x = -1 \\ 0.5 - 0.1 = 0.4, & x = 0 \\ 0.75 - 0.5 = 0.25, & x = 10 \\ 1 - 0.75 = 0.25, & x = 15 \end{cases}$

Definition: Consider a discrete random variable X. The expectation (or expected value) of $h(X)$ is

$$E[h(X)] = \sum_{x \in R_x} h(x) f_x(x)$$

Definition: The mean (or expected value) of X is denoted by μ or μ_x or $E[X] = \sum_{x \in R_x} x f_x(x)$

Note: μ is a weighted average of the value in R_x

- μ is a center of mass
- it is a measure of centered tendancy

Interpretation: If we repeat the experiment a large number of times the $E[X]$ is the value that we expect to see on average [Law of large Numbers]

Definition: The variance of X is:

$$\sigma^2 = \sigma_x^2 = V[X] = E[(X - \mu_x)^2]$$

It is a measure of dispersion.

$$\sigma_x^2 = [(X - \mu_x)^2] = \sum_{x \in R_x} (x - \mu_x)^2 f_x(x)$$

Computational Formula

$$\sigma_x^2 = \left[\sum_{x \in R_x} x^2 f_x(x)\right] - \mu_x^2$$

Definition: The Standard Deviation of X is:

$$\sigma = \sigma_x = \sqrt{V[X]}$$

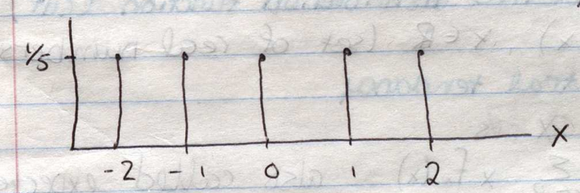
Note: σ_x is a measure of dispersion in the original units of measurement.

Example 4

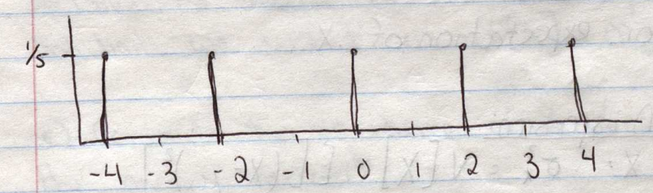
x	$f_x(x)$
-2	1/5
-1	1/5
0	1/5
1	1/5
2	1/5

y	$f_y(y)$
-4	1/5
-2	1/5
0	1/5
2	1/5
4	1/5

Compare σ_x and σ_y



$$\mu_x = \sum_{x \in R_x} x f_x(x) = (-2)(1/5) + (-1)(1/5) + (0)(1/5) \cdots + (2)(1/5) = 0$$



$$\mu_y = \sum_{y \in R_y} y f_y(x) = (-4)(1/5) + (-2)(1/5) + \ldots + (4)(1/5) = 0$$

$$\sigma_X^2 = \left[\sum_{x \in R_X} x^2 f_X(x)\right] - \mu_X^2$$

$$= \left[(-2)^2 \cdot 1/5 + (-1)^2(1/5) + \ldots + (2)^2(1/5)\right] - 0^2$$

$$= 2$$

$$\sigma_Y = \left[\sum_{y \in R_Y} y^2 f_Y(y)\right] - \mu_Y^2$$

$$= \left[(-4)^2(1/5) + (-2)^2(1/5) + \ldots + (4)^2(1/5)\right] - 0^2$$

$$= 8$$

$\therefore \sigma_X = \sqrt{2} = 1.4142$ so the standard deviation of

$\sigma_Y = \sqrt{8} = 2.828$ Y is larger than that of X

$\sigma_Y > \sigma_X$

May 19/11 **Random Variables**

Let X be a discrete random variable, then it's range is R_X and is finite and countable.

X has a prob. mass function f: (p.m.f)

$f_X(x) = P(X = x)$, $x \in R_X$

X has a cumulative distribution function (cdf) F:

$F_X(x) = P(X \leq x)$, $x \in \mathbb{R}$ (set of real numbers)

Measure of central tendancy

The mean of X is:

$\mu_X = E(X) = \sum_{x \in R_X} x f_X(x)$ also called expected value of X or expectation of X.

Measures of Dispersion:

Variance of X: $\sigma_X^2 = V[X] = E[(X - \mu_X)^2]$

$$= \sum_{x \in R_X} (x - \mu_X)^2 \cdot f_X(x)$$

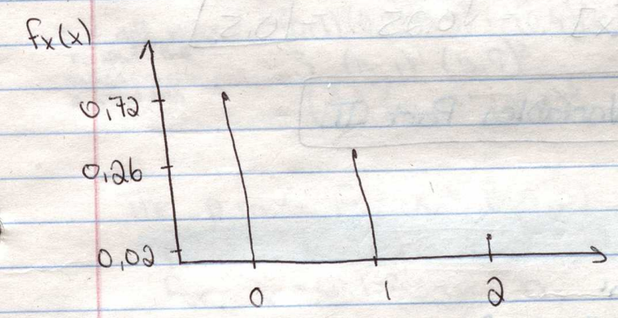
$$= \left[\sum_{x \in R_X} x^2 f_X(x)\right] - \mu_X^2$$

Standard Deviation (how dispersed the masses are on the line) of X is:

$\sigma_X = \sqrt{V[X]}$

Example 3 - refer to example 2.

where $f_X(x) = \begin{cases} 0.72, & x = 0 \\ 0.02, & x = 1 \\ 0.26, & x = 2 \end{cases}$



a) Give the expected # of defective components

Recall: X = # of defective components

Solution: We want $E[X] = \sum_{x \in R_X} x f_X(x)$

$$= 0(0.72) + 1(0.26) + 2(0.02)$$

$$\boxed{= 0.3}$$

b) Give the mean of X: (the expected value of X)

Solution: We want $\mu_X = E[X] = 0.3$

c) Compute the variance of X (expected deviation from the mean squared)

Solution:

$$\sigma_X^2 = \left[\sum_{x \in R_X} x^2 f_X(x)\right] - \mu_X^2$$

$$= \left[0^2[0{,}72] + 1^2(0{,}26) + 2^2(0{,}02)\right] - (0{,}3)^2$$

$$= 0{,}26 + 0{,}08 - 0{,}09$$

$$\boxed{= 0{,}25}$$

d) Compute the standard deviation of X

Solution: $\sigma_X = \sqrt{V[X]} = \sqrt{0{,}25} = \boxed{0{,}5}$

Random Variables Part II

Discrete Models

Binomial Distribution:

Definition: A **Bernoulli trial** is an experiment with two possible outcomes: **success** and **failure**

Notation: $p = P(\text{"success"})$

Definition: A **binomial experiment** is n independant Bernoulli trials, where the probability of success is fixed at p.

A **binomial random variable** X is the number of successes in a binomial experiment

Notation: $X \sim B(n,p)$

→ means that X has a binomial distribution with parameters n and p

Example

$X \sim B(3; 0{,}1)$

number of trials $n=3$

probability of success $p=0{,}1$

a) Find $P(X=1)$

$$P(X=1) = P(\{SFF, FSF, FFS\})$$

$$= P(\{SFF\}) + P(\{FSF\}) + P(\{FFS\})$$

$$= (0{,}1)(0{,}9)(0{,}9) + (0{,}9)(0{,}1)(0{,}9) + (0{,}9)(0{,}9)(0{,}1)$$

# of diff ways to put the success in 3 positions

$$= 3\,(0{,}1)^1(0{,}9)^2$$

$$= \binom{3}{1}(0{,}1)^1(0{,}9)^2$$

The p.m.f. for $X \sim B(n,p)$ is:

$$f_X(x) = P[X=x] = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, 1, 2, \ldots, n$.

The c.d.f for $B(n,p)$ is given in table II

The mean of $X \sim B(n,p)$ is $\mu_X = E[X] = np$

The variance is $\sigma_X^2 = V[X] = np(1-p)$

Example 6

70% of the chips receive a thick enough coating

Compute the probability that among $n=15$ chips:

a) Exactly 10 will have a thick enough coating

Solution:

X = # of chips that have a thick enough coating

So $X \sim B(15, 0{,}7)$

We want: $P(10=X) = f_X(10)$

$$= \binom{15}{10}(0{,}7)^{10}(0{,}3)^5$$

$= (3003)(0.7)^{10}(0.3)^5$
$= 0.2061$

Alternatively:
$P(X=10) = P(X \le 10) - P(X \le 9)$
$= F_X(10) - F_X(9)$
$= 0.4845 - 0.2784$ (values from table II)
$= 0.2061$

Table II
$\boxed{n=15}$

X	P → $\boxed{0.7}$
0	
1	
2	
⋮	
9	$\boxed{0.2784}$
10	$\boxed{0.4845}$

b) $P(X \le 2)$ at most have an even coating of 2
$= F_X(2) = 0.0000$ (from table II)

with the pmf
$P(X \le 2) = f_X(0) + f_X(1) + f_X(2)$
$= \binom{15}{0}(0.7)^0(0.3)^{15} + \binom{15}{1}(0.7)^1(0.3)^{14} + \binom{15}{2}(0.7)^2(0.3)^{13}$
$= 0.0000$

c) P("at least 14 with a thick enough coating") → 14 or more
$= P(X=14) + P(X=15)$
$= 0.0353$

OR Alternatively
$P(X \ge 14) = 1 - P(X < 14)$

$= 1 - P(X \le 13)$ → this is our c.d.f. at 13.
$= 1 - F_X(13)$
$= 1 - 0.9647$ → value from table II
$= 0.0353$

Example 7: Refer to example 6.

a) We want
$E[X] = np$
$= 15(0.7)$
$= 10.5$

b) We want the standard deviation
$\sigma_X = \sqrt{V[X]}$
$= \sqrt{np(1-p)}$
$= \sqrt{(15)(0.7)(0.3)}$
$= 1.7748$

Geometric Model

Reminder from calculus:
Geometric Series:
$a + ax + ax^2 + ax^3 + \ldots$
$=$ converges to $\frac{a}{1-x}$ if $|x| < 1$

Let X = # of independant Bernoulli trials required for success

Note: $p = P(\text{"success"})$

It's p.m.f. is $f_X(x) = P(X=x)$ for $x = 1, 2, 3, \ldots$
$= P($"first $x-1$ trials are failures and x^{th} is a success"$)$

$\Rightarrow P(X=x) = (1-p)^{x-1}\, p$ for $x = 1, 2, 3 \ldots$

It's c.d.f is:

$$F_X(x) = P(X \le x)$$
$$= 1 - P(X > x)$$
$$= 1 - P(\text{"first } x \text{ trials are failures"})$$
$$= 1 - (1-p)^x$$

$\therefore F_X(x) = 1 - (1-p)^x$ for $x = 1, 2, 3 \ldots$

It's mean is $\mu_X = E[X] = 1/p$

It's variance is $\dfrac{1-p}{p^2} = V[X] = \sigma_X^2$

Example 8

We have boxes of 10. A box is returned only if there is more than 1 defective. Assume 5% of the components are defective

a) probability that the 10th box recieved is the first returned box?

Solution: Compute the probability that a box will be returned.

$p = P(\text{"a box is returned})$

$= P(X > 1)$ where X = # of defective in a box of 10.

$\sim = B(10, 0.05)$

$= 1 - P(X \le 1)$

$= 1 - \left[\binom{10}{0}(0.05)^0(0.95)^{10} + \binom{10}{1}(0.05)^1(0.95)^9\right]$

$= 0.08614$

We want $P(Y=10)$ where Y = # of boxes received to observe one that is returned

$Y \sim$ geometric with $p = 0.08614$

$P[Y=10] = f_Y(10) = (1-p)^9\, p = 0.03829$

b) Expected # of boxes received when we first return a box

$= E[Y] = 1/p = 11.609$

c) More than 10 boxes will be received before one is returned

We want $P[Y > 10]$

$= 1 - P(Y \le 10)$

$= 1 - F_Y(10)$

$= 1 - \left[1 - (1-p)^{10}\right]$

$= 0.46625$

May 24/11 · Assignment 2 due June 2nd.

Last lecture: Random Variables Part II

Discrete Models

<u>Recall</u>: A Bernouilli trial is a random experiment with 2 possible outcomes: success and failure.

<u>Notation</u>: $p = P(\text{"success"})$

<u>Remark</u>: For the binomial distribution, geometric distribution and the negative binomial (later today), we are working with <u>independant Bernouilli trials</u> and $p = P(\text{"success"})$ is fixed (probability does not change from trial to trial.)

<u>Scenario 1</u>:

fixed number of trials, that is $n = \#$ of trials

Let $X = \#$ of successes

$\sim B(n,p)$

← binomial distribution with parameters n and p

<u>Scenario 2</u>:

not a fixed number of trials

Let $Y = \#$ of trials required to observe a success

Then $Y \sim$ geometric distribution with parameter p

<u>Example</u>:

Let $X \sim B(20, 0.4)$ → means there are 20 trials, the prob of having a success is 40%.

a) Compute $P(X=8)$

b) Compute $P(9 \le X \le 15)$

<u>Solutions</u>:

a) $P(X=8) = f_X(8)$

*where f_X is the p.m.f of X

$$f_X(8) = \binom{20}{8}(0.4)^8(0.6)^{12}$$

← number of failures

↑ # of successes

↑ % of failure

8 successes out of 20 trials

↑ % of success

$f_X(8) = 0.1797$

$$\boxed{\therefore P(X=8) = 0.1797}$$

Note: if $X \sim B(n,p)$ its p.m.f is

$f_X(x) = P(X=x)$

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

for $x = 0, 1, \dots, n$

b) $P(9 \le X \le 15)$

$= P(X=9) + P(X=10) + P(X=11) + P(X=12) + P(X=13) + P(X=14) + P(X=15)$

$= P(X \le 15) - P(X \le 8)$

$= F_X(15) - F_X(8)$ ← find these values at the back of book, tbl II

In table II

$n = 20$ $\quad p = 0.4$

$x = 8 \longrightarrow 0.5956$

$x = 15 \longrightarrow 0.9997$

so...

$= 0.9997 - 0.5956$

$= 0.4041$ ← much more efficient than computing the individual masses.

Definition: Consider independant Bernouilli trials with $p = P(\text{"success"})$. Let $X = \#$ of trials required to observe r successes. We say that X has a negative binomial distribution with parameters p and r.

It's pmf is:

$f_X = P(X = x)$, $x = r, r+1, r+2, \ldots$

$= P($"x^{th} trial is a success" and " the first $x-1$ trials give $r-1$ successes"$)$

$$= \underbrace{\binom{x-1}{r-1} p^{r-1} (1-p)^{x-r}}_{r-1 \text{ successes among } x-1 \text{ trials}} \cdot \underset{\text{last trial is a success}}{p}$$

$$\boxed{\therefore f_X(x) = \binom{x-1}{r-1} p^{r} (1-p)^{x-r} \quad \text{for } x = r, r+1, \ldots}$$

← p.m.f. for a negative binomial

It's mean is

$$\mu(x) = E[X] = \frac{1}{p} \cdot r = \frac{r}{p}$$

(expected # of trials for r successes)

It's variance is

$$\sigma_X^2 = V[X] = \frac{r(1-p)}{p^2}$$

<u>Example 9</u> in the notes

2.5% of boxes are returned to the manufacturer. Give the probability of the 10th box received by the supplier is the 3rd box returned to the manufacturer.

→ Think of a box as a Bernoulli trial since it is either returned or not.

Consider a returned box as a success

Let X = # of boxes required to observe 3 returned boxes.

We want $P(X = 10)$

Note: $X \sim$ negative binomial with $p = 0.025$ and $r = 3$

$$P(X=10) = f_X(10) = \binom{9}{2} (0.025)^3 (0.975)^7$$

← prob of 3 successes in 10 trials

$$= {}_9C_2 (0.025)^3 (0.975)^7$$

$$= (36)(0.025)^3 (0.975)^7$$

$$= 0.00047$$

$$\boxed{\therefore P(X=10) = 0.00047}$$

b) Consider 10 boxes. What is the probability that among the 10 boxes, we have 3 returned boxes?

→ X = # of returned boxes among $n = 10$ boxes then $X \sim B(10, 0.025)$

We want $P(X=3)$

$$P(X=3) = \binom{10}{3} (0.025)^3 (0.975)^7$$

$$= {}_{10}C_3 (0.025)^3 (0.975)^7$$

$$= (120)(0.025)^3 (0.975)^7$$

$$= 0.00157$$

$$\boxed{\therefore P(X=3) = 0.00157}$$

Poisson Approximation to the binomial Distribution.

Definition: Let X be a discrete random variable with the following prob. mass function (p.m.f).

$f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ for $x = 0, 1, 2 \ldots$ where λ is a parameter such that $\lambda > 0$

Is this a p.m.f.?

→ we need i) $0 \le f(x) \le 1$ for all $x \in R_X$

ii) $\sum_{x \in R_X} f(x) = 1$

Note: that $e^{-\lambda}\frac{\lambda^x}{x!} > 0$

Recall: (from Calculus II)

$e^{\lambda} = \sum_{x=0} \frac{\lambda^x}{x!}$ ← Maclauren series for e^{λ}

$\Rightarrow \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = 1$ and furthermore $0 \le e^{-\lambda}\frac{\lambda^x}{x!} \le 1$

for all $x = 0, 1, 2, \ldots$

We can show that if $X \sim$ Poisson with parameter λ

the mean is:

$\mu(x) = E[X] = \lambda$

the variance is:

$\sigma_X^2 = V[X] = \lambda$

Poisson Approximation of the binomial $B(n, p)$

Assume that n is large and p is small

Equate means: $\lambda = np$

We get:

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \approx e^{-\lambda}\frac{\lambda^x}{x!}$

Rule of Thumb

Approximation is good if $n \geq 20$ and $p \le 0.05$

Very good if $n \geq 100$ and $np \le 10$

Proof:

Let $n \to \infty$ and $p \to 0$ such that $\lambda = np$ is a constant

$\binom{n}{x} p^x (1-p)^{n-x}$ (replace all p by λ/n)

$= \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$

$= \frac{n!}{x!(n-x)!} \frac{\lambda^x}{n^x} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^x}$

$= \frac{\lambda^x}{x!} \cdot \frac{n!}{(n-x)! \cdot n^x} \cdot \frac{(1-\lambda/n)^n}{(1-\lambda/n)^x}$

$\to e^{-\lambda}\frac{\lambda^x}{x!}$ as $n \to \infty$

since $\left(1 - \frac{\lambda}{n}\right)^{-x}$ converges to 1 as $n \to \infty$

and $\frac{n!}{(n-x)! \cdot n^x} = \frac{n(n-1)(n-2)\ldots(n-(x-1))}{n \; n \; n \ldots n}$

$\to 1$ as $n \to \infty$

Recall from Calculus:

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

We can show that

$\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$ as $n \to \infty$

Example 10 from notes

3% of drivers get at least 1 parking ticket per year. Approximate the probability that among 80 drivers

a) 3 will get at least 1 parking ticket in a year

Solution: Let X = # of drivers among $n = 80$ that get at least 1 ticket in a year

$X \sim B(n=80, p=0.03)$

Since n is large and p is small ie $n \geq 20$ and $p \leq 0.05$ then the poisson approximation is good

Use $\lambda = np = (80)(0.03)$

$= 2.4$ ← on average, 2.4 drivers of 80 get a ticket.

We want

$$P(X=3) \approx e^{-2.4}\frac{(2.4)^3}{3!} = 0.2090$$

$$\boxed{\therefore P(X=3) = 0.2090}$$

b) We want at least 3 people will get a ticket per year

$P(X \geq 3) = 1 - P(X < 3)$

$= 1 - [P(X=0) + P(X=1) + P(X=2)]$

$$\approx 1 - e^{-2.4}\left[\frac{(2.4)^0}{0!} + \frac{(2.4)^1}{1!} + \frac{(2.4)^2}{2!}\right]$$

$= 0.4303$

$$\boxed{\therefore P(X \geq 3) = 0.4303}$$

c) $$P(2 \leq X \leq 5) \approx e^{-2.4}\left[\frac{(2.4)^2}{2!} + \frac{(2.4)^3}{3!} + \frac{(2.4)^4}{4!} + \frac{(2.4)^5}{5!}\right]$$

$= 0.6559$

$$\boxed{\therefore P(2 \leq X \leq 5) = 0.6559}$$

Definition: Consider a process of "changes" or "occurances" that occur in an interval. Assume that:

a) number of changes in distinct ~~or~~ and disjoint intervals are independant

Let α = rate of the changes

b) Consider a very small interval of length h, the $P(\text{"exactly 1 change"}) \approx \alpha h$

c) For the same small interval $P(\text{"more than 1 change"}) \approx 0$

then we say that the changes are modeled as a Poisson process with rate α.

May 26/11 Last lecture:

We introduced

- the negative binomial distribution
- Poisson distribution.

Recall: a Bernoulli trial is an experiment with 2 outcomes. $p = P(\text{"success"})$

Consider an independant Bernoulli trial with a probability of success p.

Let X = # of trials required to observe r successes.

Then X has a negative binomial distribution.

Poisson Distribution

A discrete random variable X. the prob. mass function (p.m.f.)

$$f_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{for } x = 0, 1, 2 \ldots \text{ where } \lambda > 0$$

Then we say that X has a Poisson distribution with parameter λ

Note: λ is the mean of X in this case, that $\mu(x) = E[X] = \lambda$

Poisson Approximation to the binomial $B(n,p)$ where n is large and p is small, then:

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \approx e^{-\lambda} \frac{\lambda^x}{x!}$$

where $\lambda = np$

Poisson Process

Consider a process of "changes" in an interval of real numbers, that satisfy 3 axioms:

a) # of changes for disjoint intervals are independant

b) for a very small interval of length h (almost 0) then the prob. of exactly 1 change is:

$P(\text{"exactly one change"}) \approx \alpha h$

We will interpret α as the rate of the process.

c) For the same small interval

$P(\text{"more than one change"}) \approx 0$

We say that such a process is called a Poisson process with rate $\alpha > 0$

Theorem:

Let X = # of changes in a fixed length interval, say t, for a Poisson Process of rate α. Then $X \sim$ Poisson with mean $\mu_X = \lambda = \alpha t$

Example 11 from Random Variables II

Calls occur according to a Poisson Process. On average, there are 6 calls per hour. And we want to know the probability that:

a) prob that there are no calls in one hour

Solution: Let X = # of calls in one hour.

Then $X \sim$ Poisson with mean $\mu = \alpha t = (6)(1) = 6$

We want

$$P[X=0] = f_X(0) = e^{-6} \frac{6^0}{0!} = \boxed{0.0025}$$

$\therefore$ 0.0025 prob that there will be no calls in 1 h.

b) there are at most 2 calls in 30 minutes

Solution: Let X = # of calls in 30 mins (0.5 hours)

Then $X \sim$ Poisson with mean $\mu = \alpha t = (6)(\frac{1}{2}) = 3$

We want

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$
$$= e^{-3}\left[\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right]$$
$$\boxed{= 0.4232}$$

$\therefore$ the prob that there are at most 2 calls in 30mins is 0.4232.

c) prob. of 10 calls in 2 hours

Solution: X = # of calls in 2 hours

$X \sim$ Poisson with mean $\mu = \alpha t = (6)(2) = 12$

We want:

$$P(X=10) = e^{-12}\frac{12^{10}}{10!} = \boxed{0.1048}$$

$\therefore$ the prob that there is 10 calls in 2 hours is 0.1048.

d) $P(X=5)$ where $X \sim$ Poisson with $\mu = \alpha t = (6)(1) = 6$

$$= e^{-6}\frac{6^5}{5!} = \boxed{0.1606}$$

$\therefore$ prob of 5 calls in one hour is 0.1606.

Example 12

An error occurs at a rate of 1 error per 35 000 cm. Assume that the errors occur as a Poisson Process.

a) Let X = # of errors in 52 500cm. Give the distribution of X.

X is counting changes for a poisson process with a rate of $\alpha = \frac{1}{35000}$ errors/cm. in an interval of length $t = 52\,500$ cm.

Thus $X \sim$ Poisson with mean $\mu = \alpha t = \left(\frac{1}{35000}\right)(52\,500) = \boxed{1.5}$

$\therefore$ on average we should see at least 1.5 errors

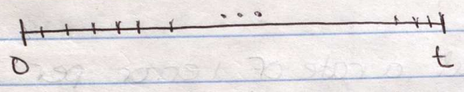
b) prob of at least 1 error in 52 500cm.

We want

$$P[X \geq 1] = 1 - P[X < 1]$$
$$= 1 - P(X=0)$$
$$= 1 - e^{-1.5}\frac{(1.5)^0}{0!}$$
$$\boxed{= 0.7769}$$

$\therefore$ the prob that at least 1 error will occur in 52 500cm is 0.7769.

Why does the count of changes for a Poisson Process of rate α? (in an interval of length t follow)



consider n subintervals where n is very large.
the length of a subinterval is:

$h = \frac{t}{n}$ (which is very small)

In the subinterval, we have an (approximate) Bernouilli trial where a success is a change:

$p = P(\text{"1 change"}) \approx \alpha h$

Let X = # of changes in the interval of length t
$X \sim$ binomial (approximately) where we have $B(n, p)$
and $p = \alpha h = \frac{\alpha t}{n}$ As $n \to \infty$, $p \to 0$

$$P(X = x) \approx \binom{n}{x} p^x (1-p)^{n-x} \longrightarrow e^{-\lambda} \frac{\lambda^x}{x!}$$

where $\lambda = np = \frac{n\alpha t}{n} = \alpha t$

$\therefore X \sim$ Poisson with mean $\mu = \alpha t$

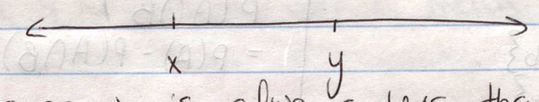
Random Variables - Part III: Continuous Models

Consider a random variable X. Its c.d.f. (cumulative distribution function) is:

$F_X(x) = P[X \le x]$, $x \in \mathbb{R}$

F_X is always a non-decreasing function
Consider $x < y$ where $x \in \mathbb{R}$, $y \in \mathbb{R}$

$\{X \le x\} \subseteq \{X \le y\} \Rightarrow P(X \le x) \le P(X \le y)$
subset



because x is always less than y, $P(X \le x) \le P(X \le y)$
$\Rightarrow F_X(x) \le F_X(y)$

$\therefore F_X$ is non-decreasing.

Definition:
Let X be a random variable with a c.d.f. F_X.
If F_X is a continuous function then we say that X is a <u>continuous random variable</u>

Definition:
It's <u>probability density function</u> is ~~$f_X(x) = F'(x)$~~

$$f_X(x) = \begin{cases} F'(x), & \text{if } F'(x) \text{ exists} \\ 0, & \text{otherwise} \end{cases}$$

<u>Properties of a probability density function</u>
a) $f(x) \ge 0$ (always non negative)
b) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (sample space prob equals one)

Remark: A function that satisfies (a) and (b) is called a p.d.f.

Remark: $P(a < X \le b)$
$= P(X \le b) - P(X \le a)$
$= F_X(b) - F_X(a)$
$= \int_a^b f_X(x)\,dx$ ← by the fundamental theorem of Calculus

Define

$A = \{X \le b\}$

$B = \{X \le a\}$

$\{a < X \le b\} = A \cap B'$

$P(A \cap B')$
$= P(A) - P(A \cap B)$

Remark: the area under the density is 0.

$P(X = x) = 0$

Thus, $P(a \le X \le b) = P(a < X \le b) = P(a < X < b) = P(a \le X < b)$

all of these events have the same probability but the area under the density is 0.

$= \int_a^b f_X(x)\,dx = F_X(b) - F_X(a)$

Example 1

Consider a Poisson Process with rate λ

Let X be the length required to observe a change

Show that

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

Solution:

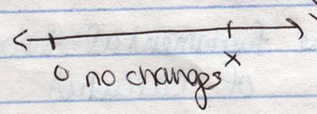
Let $x < 0$:

$P(X \le x) = 0$

Let $x \geq 0$:

$P(X \le x) = 1 - P(X > x)$

$\{X > x\}$ occurs $\Longleftrightarrow$ first change occurred after x



Let Y = # of changes in the interval $[0, x]$

then $Y \sim$ Poisson with a mean $\mu = \alpha t = \lambda x$

Note: $\{X > x\} = \{Y = 0\}$

$\therefore F_X(x) = 1 - P(X > x)$

$= 1 - P(Y = 0)$

$= 1 - e^{-\lambda x} \dfrac{(\lambda x)^0}{0!}$

$\boxed{= 1 - e^{-\lambda x}}$

Example 2

X = time to complete a task (in hours) has the p.d.f.

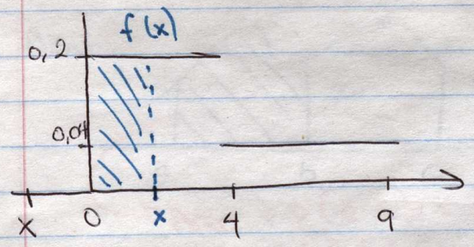
$$f(x) = \begin{cases} 0.2, & 0 < x < 4 \\ 0.04, & 4 < x < 9 \\ 0, & \text{otherwise} \end{cases}$$

a) Give the c.d.f. for X.

Note:

$F_X(x) = P(X \le x)$ $\quad x \in \mathbb{R}$

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



Let $x < 0$: $F(x) = 0$

Let $0 < x < 4$:

$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x 0.2\,dt = 0.2t \Big|_0^x$

$\boxed{= 0.2x}$

Let $4 < x < 9$.

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

$$= \int_0^4 0.2\,dt + \int_4^x 0.04\,dt$$

$$= 0.2t\Big|_0^4 + 0.04t\Big|_4^x$$

$$= 0.2(4) - 0.2(0) + 0.04(x) - 0.04(4)$$

$$= 0.8 + 0.04x - 0.016$$

$$= 0.8 + 0.04(x-4)$$

Let $x > 9$.

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt = 1$$

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ 0.2x, & 0 \le x \le 4 \\ 0.8 + 0.04(x-4), & 4 \le x \le 9 \\ 1, & x \ge 9 \end{cases}$$

Last Lecture:

Consider a Poisson process with rate α.

Let X = # of changes of the Poisson Process in a fixed length interval, say of length $= t$. Then $X \sim$ Poisson distribution with mean $\mu = \alpha t$

Recall: $f_X(x) = e^{-\alpha t}\dfrac{(\alpha t)^x}{x!}$, $x = 0, 1, 2 \ldots$

We also introduced continuous random variables

Definition: If the cdf. F_X of a random variable X is a continuous function, then X is said to be continuous

Definition: Let $f_X(x) = \begin{cases} F'(x), & \text{if it exists} \\ 0, & \text{otherwise} \end{cases}$

then f_X is the p.d.f. of X.

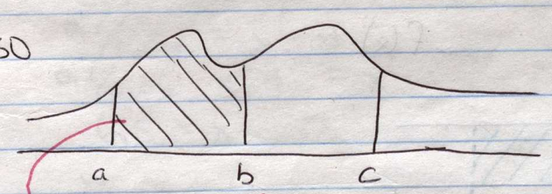
p.d.f. = probability density function.

Basic Properties:

a) $f_X(x) \ge 0$ for all x

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

Also



Area $= P(a < X < b)$

$= \int_a^b f(x)\,dx = F(b) - F(a)$

Recall: $P(X = c) = 0$ and $P(a < X < b) = P(a \le X \le b) =$

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

* but = and

Example 2

$X =$ time to complete a task (in hours) with the p.d.f.

$$f_X(x) = \begin{cases} 0{,}2 & , 0 < x < 4 \\ 0{,}04 & , 4 < x < 9 \end{cases}$$

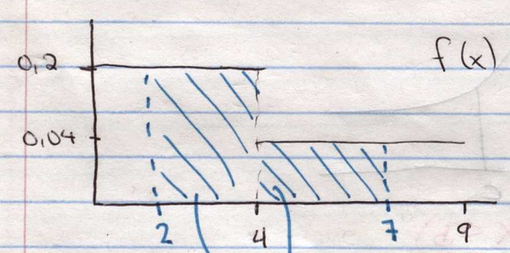
a) It's cumulative distribution function is:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(x)\,dx$$

$$F_X(x) = \begin{cases} 0 & , x \leq 0 \\ 0{,}2x & , 0 \leq x \leq 4 \\ 0{,}04(x-4) + 0{,}8 & 4 \leq x \leq 9 \\ 1 & x \geq 9 \end{cases}$$

b) We want the prob that it will be greater than 2 but less that 7.

$$P(2 < X < 7) = \int_2^7 f_X(x)\,dx$$
$$= \int_2^4 0{,}2\,dx + \int_4^7 0{,}04\,dx$$
$$= 0{,}2x\Big|_2^4 + 0{,}04x\Big|_4^7$$
$$= [0{,}2(4) - 0{,}2(2)] + [0{,}04(7) - 0{,}04(4)]$$
$$= 0{,}52$$



Area $= P(2 < X < 7)$
$= F(7) - F(2)$

* if you have access to the c.d.f. you can use it

$$= [0{,}04(7-4) + 0{,}8] - [0{,}2(2)]$$
$$= 0{,}52$$

c) Determine a time x (for a value x) such that 90% of the tasks are completed at most in x hours

We need to solve:

$$0{,}9 = P(X \leq x)$$

time to complete the task — at most x hours

but $P(X \leq x) = F_X(x)$

That is we want to solve $0{,}9 = F_X(x)$

Recall: $F_X(x)$ is a non decreasing function

$F_X(0) = 0$
$F_X(4) = 0{,}2(4) = 0{,}8$
$F_X(9) = 1$

but we want 0,9. so we know that our value is between 4 and 9

$\therefore 4 < x < 9$

Thus $0{,}9 = F_X(x)$

$$0{,}9 = 0{,}04(x-4) + 0{,}8$$
$$0{,}1 = 0{,}04(x-4)$$
$$0{,}1/0{,}04 = x - 4$$
$$0{,}1/0{,}04 + 4 = x$$
$$6{,}5 = x$$

$\therefore$ 90% of the tasks were completed in at most 6,5 hours

d) Determine the mean and variance of X

Definition: The expectation of $h(x)$ where X is a continuous random variable is

$$E[h(x)] = \int_{-\infty}^{\infty} h(x) f_X(x)\,dx$$ ← gives the average of $h(x)$ values

Definition: The mean of a continuous random variable is:

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

and its variance is:

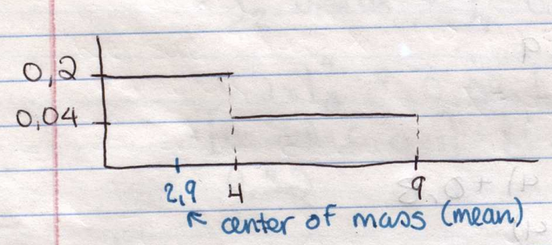
$$\sigma_X^2 = V[X] = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)\,dx$$

Computational Formula

$$\sigma_X^2 = E[X^2] - \mu_X^2 = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \mu_X^2$$

... back to our example..

d) Determine the mean and variance of X



- the mean is:

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

$$= \int_0^4 0{,}2x\,dx + \int_4^9 0{,}04x\,dx$$

$$= 0{,}1x^2\Big|_0^4 + 0{,}02x^2\Big|_4^9$$

$$= [(0{,}1)(4)^2 - (0{,}1)(0)^2] + [(0{,}02(9)^2 - (0{,}02)(4)^2]$$

$$= 2{,}9$$

- It's variance is

$$\sigma_X^2 = E[X^2] - \mu_X^2$$

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \mu_X^2$$

$$= \int_0^4 0{,}2x^2\,dx + \int_4^9 0{,}04x^2\,dx - \mu_X^2$$

$$= 0{,}2\,\frac{x^3}{3}\Big|_0^4 + \frac{0{,}04}{3}x^3\Big|_4^9 - (2{,}9)^2$$

$$= 4{,}7233$$

Theorem:

Let X be a random variable (can be discrete or continuous, the theorem holds true for both). Consider $Y = aX + b$

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

however, $V[Y] = a^2 V[X]$

thus $\sigma_Y = \sqrt{V[Y]}$

$= \sqrt{a^2 V[X]}$

$= |a|\,\sigma_X$ ← a multiple of the standard deviation.

back to example 2..

e) cost is \$1500/hour

$\therefore C = 1500X$

What is the expected value of the cost?

Solution:

We want the expectation of our cost $E[C]$

$E[C] = E[1500X]$

$= 1500\,E[X]$

$= 1500(2{,}9)$

$= 4350$ ∴ on average the cost of the task would be \$4350.

f) Compute the standard deviation (not variance because that would mean squared \$, and we can't deal with that)

$$\sigma_c = \sqrt{V[C]} = \sqrt{V[1500x]}$$
$$= \sqrt{(1500)^2\, V[x]}$$
$$= 1500\,\sigma_x$$

$$\Rightarrow \sigma_c = 1500\,\sigma_x$$
$$= 1500\sqrt{4,7233}$$
$$= 3259,9732$$

Recall: $\sigma_x^2 = E[X^2] - \mu_x^2$
$= \boxed{4,7233}$

<u>Poisson Process with rate λ:</u>

Let x = length of the interval required to observe a change.

We say that X has an exponential distribution with parameters λ * $\lambda = pt$

Refer to Example 1 for its cdf. that is:

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

Its p.d.f is $f_X(x) = F'(x)$

$$f_X(x) = -e^{-\lambda x}(-\lambda)$$
$$= -\lambda e^{-\lambda x}, \quad x > 0$$

<u>Note:</u> Its mean is $E[X] = 1/\lambda$ and its variance is $\sigma_x^2 = 1/\lambda^2$

<u>Example 3</u>

errors on a tape occur at a rate of 1 error per 35 000 cm as a Poisson process

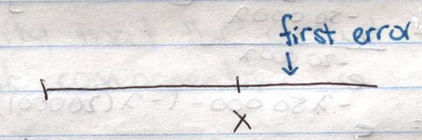
a) determine the length x of the tape that is 90% certain that there are no errors

Our rate is $\lambda = 1/35\,000$ errors per cm.

Let x = length (in cm) required to observe an error

$\therefore X \sim$ exponential with $\lambda = 1/35\,000$

We want
$0,9 = P(X > x)$



Solve: $0,9 = P(X > x)$
$$= 1 - P(X \leq x)$$
$$= 1 - F_X(x)$$
$$= 1 - [1 - e^{-\lambda x}]$$
$$= e^{-\lambda x}$$

$$\ln(0,9) = \ln e^{-\lambda x}$$
$$\ln(0,9) = -\lambda x$$
$$-\frac{\ln(0,9)}{\lambda} = x$$
$$-\frac{\ln(0,9)}{(1/35\,000)} = x$$
$$36687,61805 \text{ cm.}$$

b) No errors in 20 000 cm. What is the prob of no errors on the next 30 000 cm?

Solution:

We want

$$P[X > 50\,000 \mid X > 20\,000]$$

$$= \frac{P[X > 50\,000 \text{ and } X > 20\,000]}{P[X > 20\,000]}$$

$$= \frac{P[X > 50\,000]}{P[X > 20\,000]}$$

$$= \frac{1 - P(X \le 50\,000)}{1 - P(X \le 20\,000)}$$

$$= \frac{1 - F_X(50\,000)}{1 - F_X(20\,000)}$$

$$= \frac{1 - [1 - e^{-\lambda(50000)}]}{1 - [1 - e^{-\lambda(20\,000)}]}$$

$$= \frac{e^{-50\,000\lambda}}{e^{-20\,000\lambda}}$$

$$= e^{-\lambda 50\,000 - (-\lambda(20000))}$$

$$= e^{-\lambda 30\,000}$$

$$= 0.4244$$

Note: this quantity is $P(X > 30\,000)$

Lack of Memory property of the exponential:

Let $s > 0$ and $t > 0$

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

Erlang: Consider a Poisson process with rate λ

Let X = length required to observe r changes

$\sim$ Erlang with parameters r and λ

- Its mean is $E[X] = \frac{r}{\lambda}$
- Its variance is $V[X] = \frac{r}{\lambda^2}$

June 2nd

Midterm on Tuesday after the study break.

2 parts: a) 2 short answer questions

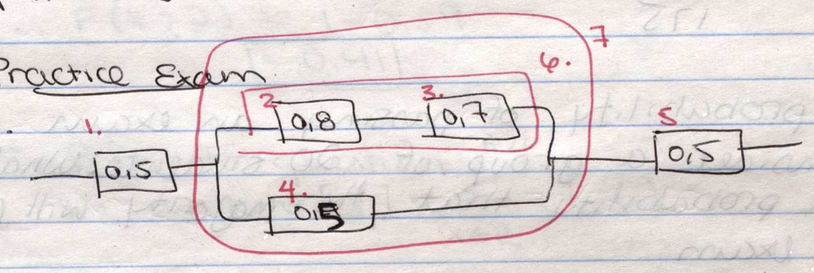
b) 7 multiple choice

Office Hours: Wed 5-7pm Friday 1-3pm

Material covered on the midterm: Everything up to pdf and c.d.f. for a continuous random variable.

Practice Exam

1.



What is the probability that the circuit is functional? Assume independant components.

Define: A_i = " component i works " for $i = 1, \dots, 5$

$$P(A_6) = P(A_2 \cap A_3) = P(A_2)P(A_3) \text{ by independance}$$

$$= (0.8)(0.7)$$

$$= 0.56$$

$$P(A_7) = P(A_6 \cup A_4)$$

$$= 1 - P[(A_6 \cup A_4)']$$

$$= 1 - P(A_6' \cap A_4') \quad \text{by DeMorgan}$$

$$= 1 - P(A_6')P(A_4') \quad \text{by independance}$$

$$= 1 - (1 - 0.56)(1 - 0.5)$$

$$\boxed{0.78}$$

P(" circuit works ")

$$= P(A_1 \cap A_7 \cap A_5)$$

$$= P(A_1)P(A_7)P(A_5)$$

$$= (0.5)^2 (0.78) = 0.195$$

2.

wife	husband ≤ 30000	husband $> 30\,000$
$\leq 30\,000$	425	400
$> 30\,000$	65	110

$P(H > 30\,000 \mid W > 30\,000)$

$= \frac{110}{175} = 0{,}6286.$

3. probability of passing an exam is 0,8. Consider a group of 20 students. What is the probability that the majority will pass the exam.

Let X be the number of students among $n = 20$ that pass the course.

Note: The students can be considered as independant Bernoulli trials, where "success" = "passed the exam"
and so $p = P(\text{"success"}) = 0{,}8$
thus $X \sim B(20, 0{,}8)$

We want

$P[X > 10] = 1 - P(X \leq 10)$
$= 1 - F_X(10)$
$= 1 - 0{,}0026$
$\boxed{= 0.9974}$

From Table II

$n = 20$ $\quad \boxed{p = 0{,}8}$

$\boxed{x = 10}$ $\quad 0{,}0026 \rightarrow F_X(10)$

4. We are given a c.d.f. for X

We want

$P(X > 2) = 1 - \boxed{P(X \leq 2)} \leftarrow$ c.d.f. at 2.
$= 1 - F_X(2)$

We are given $F_X(2) = 0{,}59$

$\therefore P(X > 2) = 1 - 0{,}59$
$\boxed{= 0{,}41}$

5. Refer to question #4.

We want $E[X] = \sum_{x \in R_X} x f_X(x)$

$$F_X(x) = \begin{cases} 0 & , x < 0 \\ 0{,}17 & 0 \leq x < 1 \\ 0{,}4 & 1 \leq x < 2 \\ 0{,}59 & 2 \leq x < 3 \\ 0{,}72 & 3 \leq x < 4 \\ 0{,}8 & 4 \leq x < 5 \\ 1 & x \geq 5 \end{cases}$$

$R_X = \{0, 1, 2, 3, 4, 5\}$

$$f_X(x) = \begin{cases} 0{,}17 & x = 0 \\ 0{,}4 - 0{,}17 = 0{,}23 & x = 1 \\ 0{,}59 - 0{,}40 = 0{,}19 & x = 2 \\ 0{,}72 - 0{,}59 = 0{,}13 & x = 3 \\ 0{,}8 - 0{,}72 = 0{,}08 & x = 4 \\ 1 - 0{,}8 = 0{,}20 & x = 5 \end{cases}$$

$E[X] = 0(0{,}17) + 1(0{,}23) + 2(0{,}19) + 3(0{,}13) + 4(0{,}08) + 5(0{,}2) = \boxed{2{,}32}$

x	0	1	2	3	4	5
$F_X(x)$	0,17	0,4	0,59	0,72	0,8	1

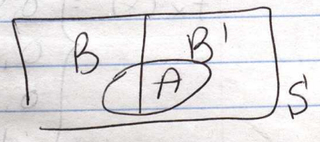
6. 0,2 cracks per km Poisson Process
What is the probability of at most 1 crack in 10 km?

Let X = # of cracks in 10km
$\sim$ Poisson with mean
$\mu = \alpha t = (0,2)(10) = 2$

We want
$P(X \le 1) = P(X=10) + P(X=1)$
$= e^{-2}\frac{2^0}{0!} + e^{-2}\frac{2^1}{1!}$
$= 0,4060$
$= e^{-2}(1+2) = 3e^{-2}$

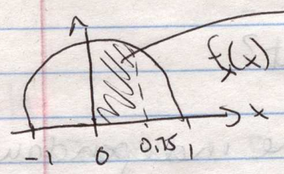
7. A = "A get the job"
B = "B will submit a bid"

We know that
$P(A|B') = 4/5$
$P(B) = 1/3$
$P(A|B) = 1/3$



We want
$P(A) = P(A \cap B) + P(A \cap B')$
$= P(A|B)P(B) + P(A|B')P(B')$
$= (1/3)(1/3) + (4/5)(2/3)$
$= 0,6444$

8. $f_X(x) = 3/4(1-x^2) \quad -1 < x < 1$
$P(0 < x < 0,75) = ?$



Area $= P(0 < X < 0,75)$
$= \int_0^{0,75} f_X(x)\,dx$
$= \int_0^{0,75} 3/4(1-x^2)\,dx$
$= 3/4\left(x - x^3/3\right)\Big|_0^{0,75}$
$= 3/4\left(0,75 - \frac{(0,75)^3}{3}\right) - 0$

$\boxed{= 0,457}$

Remark:
$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$

$$F_X(x) = \begin{cases} 0, & x \le -1 \\ \int_{-1}^{x} 3/4(1-t^2)\,dt = 3/4\left(x - x^3/3\right), & -1 \le x \le 1 \\ 1 \end{cases}$$

for $-1 \le x \le 1$
$F_X(x) = \int_{-1}^{x} 3/4(1-t^2)\,dt$
$= 3/4\left(x - \frac{x^3}{3}\right) - 3/4\left(-1 - \frac{(-1)^3}{3}\right)$

$P(0 < X < 0,75)$
$= F(0,75) - F(0)$
$= 0,457.$

9. prob = 0,1 = P("battery is fully charged"
We want the prob that the third battery is the first one that is fully charged

X = # of batteries required to observe a fully charged battery $\sim$ geometric with $p = 0,1$

We want $P(X=3) = (1-p)^2 p$
$\boxed{= 0.081}$

10. Are A and B independant?

By definition A and B are independant if $P(A \cap B) = P(A)P(B)$

$P(A \cap B) = ?$

Addition Rule:

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

$0.65 = 0.3 + 0.5 - P(A \cap B)$

$\Rightarrow P(A \cap B) = 0.8 - 0.65 = 0.15$

Compare $P(A \cap B) = 0.15$

to $P(A)P(B) = (0.3)(0.5) = 0.15$

both quantities are the same.

$\therefore$ A & B are independant

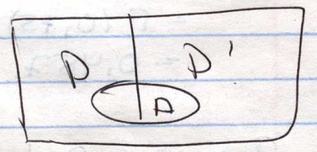
11. 2% of welds are defective

D = "weld is defective"

A = "signal is positive, ie a defective"

$P(A|D) = 0.9$

$P(A|D') = 0.05$



We want $P(D|A) = ?$

$P(D|A) = \dfrac{P(D \cap A)}{P(A)}$ by definition

$= \dfrac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D')P(D')}$ ← multiplication rule ← total prob rule

$P(D|A) = \dfrac{(0.9)(0.02)}{(0.9)(0.02) + (0.05)(0.98)}$

$= 0.2687$

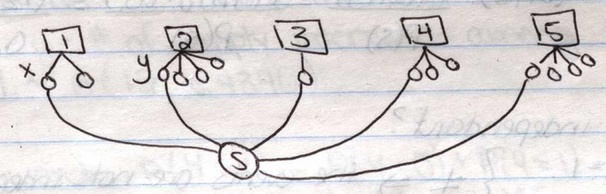
June 16/11

Question 1 From the Midterm.

Step 1: Choose 1 of 5 doors

Step 2: Choose a pothole

In each case, that is behind each door, there is one and only one pothole to the safe haven.



a) P("reach the safe haven")

Let S = "reaching a safe haven"

and D_i = "door i" $i = 1,2,3,4,5$

We know:

$P(D_i) = 1/5$, $i = 1,2,3,4,5$

$P(S|D_1) = 1/2$ $\quad P(S|D_4) = 1/3$

$P(S|D_2) = 1/4$ $\quad P(S|D_5) = 1/4$

$P(S|D_3) = 1$

$P(S) = P(S|D_1)P(D_1) + P(S|D_2)P(D_2) + P(S|D_3)P(D_3) + P(S|D_4)P(D_4) + P(S|D_5)P(D_5)$

$= (1/2)(1/5) + (1/4)(1/5) + (1)(1/5) + (1/3)(1/5) + (1/4)(1/5)$

$= 0.46667$

If we use an equally likely model P("safe haven") =

$$= \frac{N(\text{safe haven})}{N(S)} = \frac{5}{14}$$

*Note: Not all pathways are equally likely

$P(\text{"path x"}) = 1/5 \cdot 1/2 = 1/10$

$P(\text{"path y"}) = 1/5 \cdot 1/4 = 1/20$

b) probability of choosing door one given that we reach the safe haven

We want $P(D_1|S) = \frac{P(D_1 \cap S)}{P(S)} = \frac{P(S|D_1)P(D_1)}{P(S)} = \frac{(1/2)(1/5)}{0.46667} = 0.2143$

c) Are they independant?

$\left.\begin{array}{l} P(S|D_3) = 1 \\ P(S) = 0.4667 \end{array}\right\} \neq \Rightarrow$ the events are not independant they are dependant.

Alternative Solution

$P(S \cap D_3) = P(S|D_3)P(D_3)$
$= (1)(1/5) = 1/5$

but $P(S)P(D_3) = (0.4667)(1/5) \neq 1/5$

Since $P(S \cap D_3) \neq P(S)P(D_3) \Rightarrow$ these events are dependant

Question 2

The train arrives at a rate of 10 trains/hour, Poisson Process

a) X has a poisson distribution with $\mu = \alpha t = 10 \cdot \frac{m}{60} = \frac{1}{6} m$

b) $P(X > 1) = 1 - P(X \leq 1)$
$= 1 - [P(X=0) + P(X=1)]$

$\mu = \frac{m}{6} = \frac{4}{6} = \frac{2}{3}$

$\Rightarrow 1 - e^{-2/3}\left(\frac{2/3^0}{0!} + \frac{2/3^1}{1!}\right)$

$= 0.1443$

c) $P(X=0)$ in 5 minutes

$\mu = \frac{5}{6} = 0.8333$

$P(X=0) = e^{-5/6}\frac{(5/6)^0}{0!} = 0.4346$

d) Consider 5 minute intervals (12 of them)
Let y = # of empty intervals among 12
$Y \sim B(12, 0.4346)$

We want $P(Y \leq 1) = P(Y=0) + P(Y=1)$

$= \binom{12}{0}(0.4346)^0(1-0.4346)^{12} + \binom{12}{1}(0.4346)^1(1-0.4346)^{11}$

$= 0.0109$

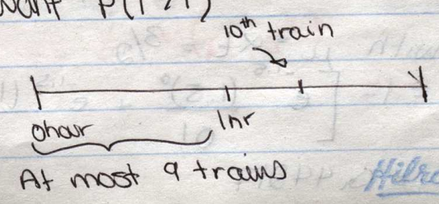
Added Question to introduce the Erlang Distribution:

e) What is the probability that we must wait more than an hour to observe 10 trains?

Solution:

Let T be the time to wait to observe 10 trains (in hr

We want $P(T > 1)$



Let X = # of trains in 1 hour
So $X \sim$ Poisson with mean $\mu = 10$

$$\therefore P(T > 1) = P(X \leq 9)$$
$$= e^{-10}\left(\frac{10}{0!} + \frac{10^1}{1!} + \frac{10^2}{2!} + \dots + \frac{10^9}{9!}\right)$$

Erlang Distribution:

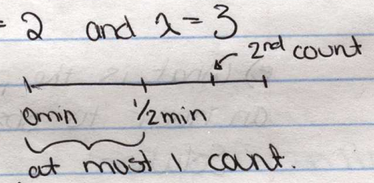
T = length of interval required to observe r changes in a poisson process with rate λ

$T \sim$ Erlang distribution with parameters r and λ

$E[T] = r/\lambda \qquad V[T] = r/\lambda^2$

Example 4 from the notes
Poisson Process with 3 counts per minute on average.
Probability that the wait for 2 counts in at most 30 sec.

Let T = waiting time for 2 counts in minutes
$T \sim$ Erlang distribution with $r = 2$ and $\lambda = 3$



We want $P(T \leq 1/2) = 1 - P(T > 1/2)$
$= 1 - P(X \leq 1)$

where $X \sim$ Poisson with $\mu = \alpha t = 3/2$
$$= 1 - \left[e^{-1.5}\frac{(1.5)^0}{0!} + e^{-1.5}\frac{(1.5)^1}{1!}\right]$$
$= 0.44217$

b) mean time between counts?
Let T = time required for 1 count (in minutes
$\sim$ Exponential with $\lambda = 3$
$=$ Erlang with $r = 1$; $\lambda = 3$

We want $\mu_T = E[X] = 1/\lambda = 1/3 = 0.3333$

c) T = waiting time (in min) for 4 particles $\sim$ Erlang with $r = 4$ and $\lambda = 3$

We want $E[T] = r/\lambda = 4/3 = 1.3333$ minutes

d) σ_T where T is defined in (c)
$= \sqrt{V[T]}$
$= \sqrt{r/\lambda^2}$
$= \sqrt{4/9}$
$= 0.666669$

Normal Distribution:

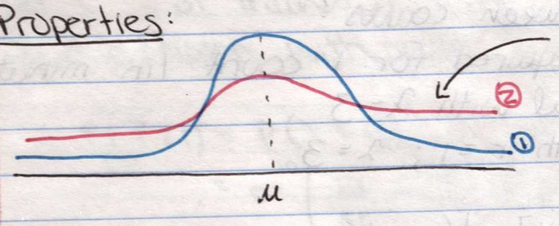
Let X be a continuous random variable with a p.d.f
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \quad \text{for } -\infty < x < \infty$$

the parameters are $\sigma > 0$ and $\mu \in \mathbb{R}$

Note: $E[X] = \mu$
$V[X] = \sigma^2$

we say that X has a normal distribution with mean μ and variance σ^2
Notation: $X \sim N(\mu, \sigma^2)$

Properties:



bell shaped symetric density about $X = \mu$

$\sigma_1 < \sigma_2$

Definition: The Standard normal distribution has the p.d.f.

$$\phi(z) = \frac{1}{2\pi^2} e^{z^2/2} \quad \text{where } -\infty < z < \infty$$

$Z \sim N(0,1)$

standard normal.

It's c.d.f is: $\Phi(z) = P[Z \leq z]$, $z \in \mathbb{R}$ Note: Refer to table 3 in book

Example 5

Z has a standard normal distribution ie: $Z \sim N(0,1)$

1. $P(0.53 < Z < 2.06)$

$= \Phi(2.06) - \Phi(0.53)$

$= 0.9803 - 0.7019$

$= 0.2784$

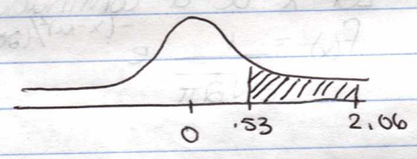


Table III

Z	0.03	0.06
0.5	0.7019	
2.0		0.9803

2. $P(-2.63 < Z \leq 0.53)$

$= \Phi(0.53) - \Phi(-2.63)$

$= 0.7019 - 0.0043$

$= 0.6976$

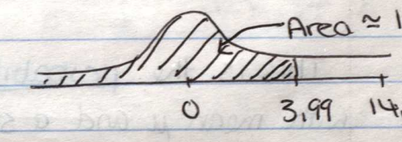
3. $P(Z > 14.5) = 1 - P(Z \leq 14.5)$

$= 1 - \Phi(14.5)$

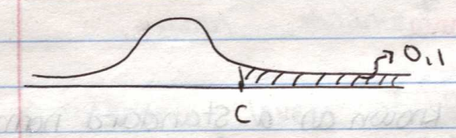
$\simeq 1 - 1$

$\simeq 0$

Z	0.09
3.9	0.999967



4. Find c such that $P(Z < c) = 0.10$



$P(Z > c) = 0.1$

$\Leftrightarrow \Phi(c) = 0.9$

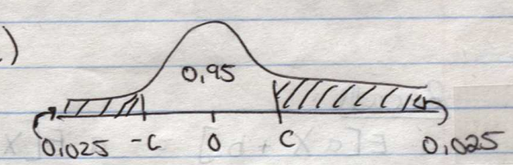
$\therefore c = 1.28$

Z	0.08
1.2	0.899727

(close to 0.9)

5. Find c such that

$0.95 = P(-c < Z < c)$



$\Rightarrow \Phi(c) = 0.975$

$\Rightarrow c = 1.96$

Z	0.6
1.9	0.975002

almost 0.975

June 21/11 Last Class:

X = distance required to observe r changes in a Poisson Process of rate $\lambda \sim$ Erlang distribution with parameters r and λ

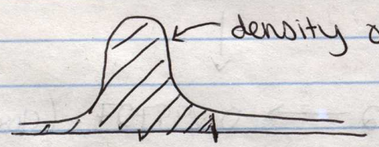
Normal Distribution

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

This is the probability density of a normal distribution with mean μ and a standard deviation σ

Notation: $X \sim N(\mu, \sigma^2)$
mean, variance

If $Z \sim N(0,1)$, this is known as a standard normal.



$\Phi(z) = P[Z \le z]$ = Area under the p.d.f from $-\infty$ to z

Recall:

$E[aX+b] = a\,E[X] + b$

$V[aX+b] = a^2\,V[X]$

Standardization of a random variable X, where $E[X] = \mu$ and $V[X] = \sigma^2$, is $Z = \frac{X-\mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$

$\therefore E[Z] = 1/\sigma\, E[X] - \mu/\sigma = \mu/\sigma - \mu/\sigma = 0$

$V[Z] = V[1/\sigma X - \mu/\sigma] = (1/\sigma)^2 V[X] = \sigma^2/\sigma^2 = 1$

Theorem: If $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$

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

Consequence:

$$P[X \le x] = P\left[\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right]$$
$$= P\left[Z \le \frac{x-\mu}{\sigma}\right] = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Example 6 (from the notes)

X = width (in cm)

$\sim N(2; (0{,}007)^2)$ * they give you the std. dev, to turn it into variance, square it.
mean, variance

a) tolerance levels of $2.0 \pm 0{,}012$. What percentage do not conform to the tolerance levels?

We want:

$P(X < 2 - 0{,}012 \text{ or } X > 2 + 0{,}012)$ these are mutually exclusive, disjoint events.

$= 1 - P(2 - 0{,}012 \le X \le 2 + 0{,}012)$

$= 1 - \left[\Phi\left(\frac{(2+0{,}012)-2}{\sigma = 0{,}007}\right) - \Phi\left(\frac{2-0{,}012-2}{0{,}007}\right)\right]$

$= 1 - [\Phi(1{,}71) - \Phi(-1{,}71)]$

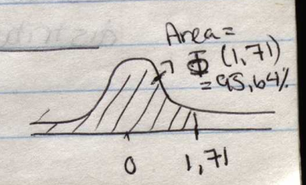
$= 1 - [0{,}9564 - 0{,}0375]$

$= 0{,}0811$

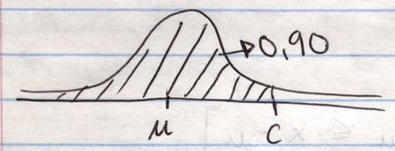
$= 8{,}11\%$

table 3 for $N(0,1)$

z	0,01
1,7	0,9564



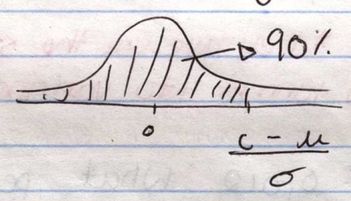
b) The width (in cm) of 90% of the items is below what value? Denote this value by c.



We want c such that:

$$0{,}9 = P[X \le c] = P\left[\frac{X-\mu}{\sigma} \le \frac{c-\mu}{\sigma}\right]$$

$$= P\left[Z \le \frac{c-\mu}{\sigma}\right] = \Phi\left(\frac{c-\mu}{\sigma}\right)$$



z	
1,2	0,899727

0,08

closest to 0,9

$$\therefore \frac{c-\mu}{\sigma} = 1{,}28$$

$\Rightarrow c = 1{,}28\sigma + \mu$ since $\mu = 2$ and $\sigma = 0{,}007$

$= 2{,}00896.$

Normal Approximation to the binomial and Poisson distribution.

Theorem: [DeMoivre - Laplace]

Let $X \sim B(n,p)$ and let $Z_n = \frac{X - np}{\sqrt{np(1-p)}}$

As $n \to \infty$, we get $P[Z \le z] \to \Phi(z)$ where Φ is the c.d.f. of $N(0,1)$

Rule of Thumb: if $np \ge 5$ and $n(1-p) \ge 5$ then the approximation is good.

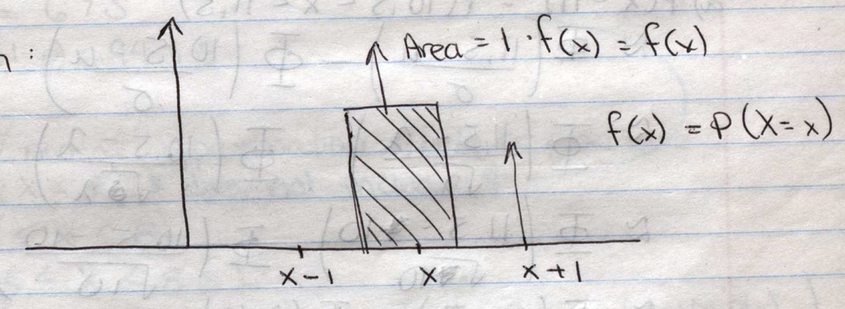
Theorem: Let $X \sim \text{Poisson}(\lambda = \mu)$ and $Z_\lambda = \frac{X-\lambda}{\sqrt{\lambda}}$

As $\lambda \to \infty$, then $P[Z_\lambda \le z] \to \Phi(z)$

Rule of Thumb: if $\lambda > 5$, then the approximation is good.

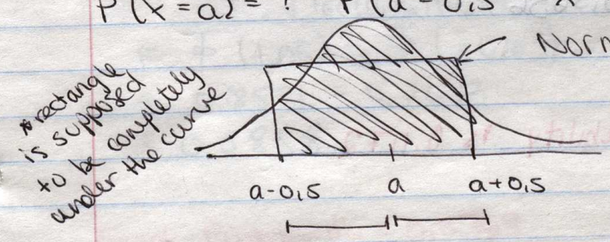
Continuity Correction:

histogram:



$P(X=a) = ?$ $P(a-0{,}5 < X < a+0{,}5)$ ← continuity correction

Normal density

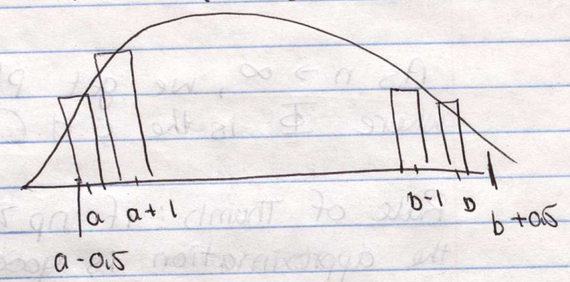


* rectangle is supposed to be completely under the curve

$P(X=a) = P(a-0.5 < X < a+0.5)$

$\approx \Phi\left(\frac{a+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a-0.5-\mu}{\sigma}\right)$

$P(a \le X \le b)$



$\approx P(a-0.5 \le X \le b+0.5)$

$\approx \Phi\left(\frac{b+0.5-\mu}{\sigma}\right) - \Phi\left(\frac{a-0.5-\mu}{\sigma}\right)$

<u>Example 7</u> from notes

$X \sim$ Poisson with $\lambda = 10$, Approximate

a) $P(X=11)$ b) $P(9 \le X \le 11)$

a) $P(X=11) = P(10.5 < X < 11.5)$

$\approx \Phi\left(\frac{11.5-\mu}{\sigma}\right) - \Phi\left(\frac{10.5-\mu}{\sigma}\right)$

$\approx \Phi\left(\frac{11.5-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{10.5-\lambda}{\sqrt{\lambda}}\right)$

$\approx \Phi\left(\frac{11.5-10}{\sqrt{10}}\right) - \Phi\left(\frac{10.5-10}{\sqrt{10}}\right)$

$\approx \Phi(0.47) - \Phi(0.16)$

$= (0.6808 - 0.5636)$

$= 0.1172$

Note: the exact probability is 0.1172

b) $P(9 \le X \le 11)$

$= P(8.5 < X < 11.5)$

$\approx \Phi\left(\frac{11.5-\mu}{\sigma}\right) - \Phi\left(\frac{8.5-\mu}{\sigma}\right)$

$\approx \Phi\left(\frac{11.5-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{8.5-\lambda}{\sqrt{\lambda}}\right)$

$\approx \Phi\left(\frac{11.5-10}{\sqrt{10}}\right) - \Phi\left(\frac{8.5-10}{\sqrt{10}}\right)$

$\approx \Phi(0.47) - \Phi(-0.47)$

$= 0.6808 - 0.3192$

$= 0.3616$

Note: The exact probability is 0.3640

<u>Example 8</u>

$X \sim B(100; 0.06)$

Approximate $P(5 \le X < 9)$

Here $np = 6 > 5$

and $n(1-p) = 94 > 5$

thus the normal approximation should be good

$P(5 \le X < 9)$ * does not include 9

$= P(5 \le X \le 8)$

$= P(4.5 < X < 8.5)$

$\approx \Phi\left(\frac{8.5-(100)(0.06)}{\sqrt{100(0.06)(0.94)}}\right) - \Phi\left(\frac{4.5-(100)(0.06)}{\sqrt{100(0.06)(0.94)}}\right)$

$\approx \Phi(1.05) - \Phi(-0.63)$

$= 0.8531 - 0.2643$

$= 0.5888$

Exact Probability $= 0.5769$

Point Estimation - beginning of statistics.

Consider a random variable X. Often we obtain repeated measurements of X:

The observed values are $x_1, x_2, \ldots, x_n$

Random Sample

These are observed values of a set of random variables: denoted as $X_1, X_2, \ldots, X_n$ (assume independant variables)

We will assume that $X_1, \ldots, X_n$ are independant and identically distributed.

The common distribution is called the population

The population has parameters.

Parameters of Interest:

a) μ = mean of the population

b) σ = standard deviation of the population

c) p = probability of success (in probability)
= population proportion (in statistics)

Goal: Estimate parameters

Consider a function of the random sample $X_1, X_2, \ldots, X_n$ say $\hat{\Theta} = h(X_1, \ldots, X_n)$

called a statistic

If $\hat{\Theta}$ is used to estimate a parameter Θ, then $\hat{\Theta}$ is called a point estimator of Θ. The observed value of a point estimator $\hat{\Theta}$, that is $\hat{\theta} = h(x_1, \ldots, x_n)$ is called a point estimate of Θ (it's just a number)

Common Statistics

a) The sample mean is: $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

It's observed value $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ is a point estimate of μ.

b) The Sample Variance:

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{(n-1)}$$

it is an estimator of σ^2

Note: $S^2 = \dfrac{\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{X}^2}{(n-1)}$

$= \dfrac{\left(\sum_{i=1}^{n} X_i^2\right) - \frac{1}{n}\left(\sum X_i\right)^2}{(n-1)}$

Note: $\sum_{i=1}^{n} (X_i - \bar{X}) = 0$

$\Rightarrow X_n - \bar{X} = \sum_{i=1}^{n-1} (X_i - \bar{X})$

So we say that we have $n-1$ degrees of freedom.

c) $\hat{p} = \dfrac{X}{n} = \dfrac{\#\text{ of successes}}{n}$

sample proportion is an estimator for p.

Remark

$S = \sqrt{S^2}$ ← sample standard deviation.

June 23/11 Last lecture: Standardization Theorem.

If $X \sim N(\mu, \sigma^2)$ and $Z = {}^{X-\mu}/_{\sigma} = \dfrac{X - E[X]}{\sqrt{V[X]}}$

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

That is Z has a standard normal distribution.

Recall: Table III has values for the c.d.f Φ of $N(0,1)$

We also discussed normal approximations of the binomial and Poisson distribution

Note: $X \sim B(n,p)$ and $np > 5$ and $n(1-p) > 5$ then X has an approximate normal distribution

If $X \sim$ Poisson with $\mu = \lambda > 5$ then $X \sim N(\lambda, \lambda)$ approx.

Recall: use the continuity correction.

Define: $Y = a_1 X_1 + \dots + a_n X_n$

a) $E[Y] = a_1 E[X] + \dots + a_n E[X_n]$

b) If $X_1, \dots, X_n$ are independant, then $V[Y] = a_1^2 V[X_1] + \dots + a_n^2 V[X_n]$

Consider a random sample $X_1, \dots, X_n$ from a population with mean μ and variance σ^2.

The sample mean is $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1}{n} X_1 + \frac{1}{n} X_2 + \dots + \frac{1}{n} X_n$

Thus $E[\bar{X}] = \frac{1}{n} E[X_1] + \frac{1}{n} E[X_2] + \dots + \frac{1}{n} E[X_n]$

$$= \underbrace{\frac{1}{n}\mu + \frac{1}{n}\mu + \dots + \frac{1}{n}\mu}_{n \text{ terms.}} = \mu$$

$\therefore E[\bar{X}] = \mu$

$$V[\bar{X}] = (1/n)^2 V[X_1] + \dots + (1/n)^2 V[X_n]$$
$$= \underbrace{1/n^2\, \sigma^2 + \dots + 1/n^2\, \sigma}_{n \text{ terms}}$$
$$= n\, 1/n^2\, \sigma$$
$$= \sigma^2/n$$

$\therefore V[\bar{X}] = \sigma^2/n \longrightarrow 0$ as $n \to \infty$

Properties of Estimators. $\hat{\theta}$ is an estimator of θ

a) if $E[\hat{\theta}] = \theta$ then we say that $\hat{\theta}$ is an unbiased estimator for θ. The bias is defined as $\text{bias}(\hat{\theta}) = E[\hat{\theta}] = \theta$

b) A measure of the error of the estimation is called the standard error and is $\sigma_{\hat{\theta}} = \sqrt{V[\hat{\theta}]}$

Remark: $\sigma_{\hat{\theta}}$ is a good measure of error for the unbiased estimator

Example:

Consider the sample mean $\bar{X}$. Since $E[\bar{X}] = \mu$, thus $\bar{X}$ is an unbiased estimator. The standard error of the mean is:

$$\sigma_{\bar{X}} = \sqrt{V[\bar{X}]} = \sqrt{\sigma^2/n} = \sigma/\sqrt{n}$$

Note: Often $\sigma_{\hat{\theta}}$ involves unknown parameters. So we give the (estimated standard error $\hat{\sigma}_{\hat{\theta}}$

For $\bar{X}$, it is $\hat{\sigma}_{\bar{X}} = s/\sqrt{n}$ where s = sample standard deviation

Estimating a proportion p:

The estimation is: $\hat{p} = X/n = \#\text{ of successes}/n$ where $X \sim B(n,p)$

Thus $E[\hat{p}] = E\left[\frac{1}{n} X\right] = \frac{1}{n} E[X]$
$= \frac{1}{n}(np)$
$= p$

$\therefore \hat{p}$ is an unbiased estimator for p

It's standard error is:

$$\sigma_{\hat{p}} = \sqrt{V[\hat{p}]} = \sqrt{V\left[\frac{1}{n} X\right]} = \sqrt{\left(\frac{1}{n}\right)^2 V[X]} = \sqrt{\left(\frac{1}{n}\right)^2 np(1-p)}$$
$$= \sqrt{\frac{p(1-p)}{n}}$$

$\therefore$ The estimated standard error is $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Example 3: Refer to example 1
Give the standard error of the mean from (a) and of the proportion from (b)

Solution:
b) $\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.6)(0.4)}{5}} = 0.2191$

a) $\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 291.9704/\sqrt{5} = 130.573$ hours

We will skip the discussion on the mean squared error (MSE)

Sampling Distribution

Definition: the distribution of $\hat{\theta}$ is called a sampling distribution.

Theorem: [Closure of normality under linear combinations]
if $X_1, \dots, X_n$ are independant and $X_i \sim N(\mu_i; \sigma^2)$ for $i = 1, 2, \dots, n$
then $Y = a_1X_1 + \dots + a_nX_n \sim N(E[Y]; V[Y])$

Consequence: If the population is $N \sim (\mu; \sigma^2)$ then $\bar{X} \sim N(\mu; \frac{\sigma^2}{n})$
What if the population is not normal?
→ Central Limit Theorem (C.L.T)
$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ approximately (as $n \to \infty$) for large n (say $n \geq 30$)

Example 4: The compressive strength of concrete has a mean of $\mu = 2500$ psi and $\sigma = 50$ psi.

Let $\bar{x}$ be the mean of 40 observations. What is $P(\bar{x} < 2490)$?

By the C.L.T:
$\bar{x} \sim N(\mu, \sigma^2/n)$ approximately, so
$P(\bar{x} < 2490) \approx \Phi\left(\frac{2490 - 2500}{50/\sqrt{40}}\right) = \Phi(-1.26) = 0.1038$
↖ table III

Comparing means from 2 independant populations:
$\bar{X}_i$ is a sample mean from the ith population with mean μ_i and variance σ_i^2. n_i = i^{th} sample size.

$\bar{X}_i \sim N(\mu_i, \sigma^2_i/n_i) \Rightarrow \bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

Example 5
$\mu_1 = 5000$ $\sigma_1 = 40$ $n_1 = 20$
$\mu_2 = 5050$ $\sigma_2 = 30$ $n_2 = 15$

Compute $P(\bar{X}_1 - \bar{X}_2 > 10)$
$= 1 - P(\bar{X}_1 - \bar{X}_2 \leq 10)$
$= 1 - \Phi\left(\frac{10 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right)$

Remark: if $z > 3.99$ then $\Phi(z) \approx 1$

$= 1 - \Phi\left(\frac{10 - (-50)}{\sqrt{\frac{40^2}{20} + \frac{30^2}{15}}}\right)$

$= 1 - \Phi(5.07$

$\therefore P(\bar{x}_1 - \bar{x}_2 > 10) \approx 1 - 1 = 0$

Confidence Intervals. (Interval Estimation)

Estimation of μ
Case 1 · σ is known
· normal population or $n \geq 30$

$\bar{x} \sim N(\mu, \sigma^2/n)$

$= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Example 3: Refer to example 1
Give the standard error of the mean from (a) and of the proportion from (b)

Solution:
b) $\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.6)(0.4)}{5}} = 0.2191$

a) $\hat{\sigma}_{\bar{x}} = s/\sqrt{n} = 291.9704/\sqrt{5} = 130.573$ hours

We will skip the discussion on the mean squared error (MSE)

Sampling Distribution

Definition: the distribution of $\hat{\theta}$ is called a sampling distribution.

Theorem: [Closure of normality under linear combinations]
if $X_1, \dots, X_n$ are independant and $X_i \sim N(\mu; \sigma^2)$ for $i = 1, 2, \dots, n$
then $Y = a_1X_1 + \dots + a_nX_n \sim N(E[Y]; V[Y])$

Consequence: If the population is $N \sim (\mu; \sigma^2)$ then $\bar{X} \sim N(\mu; \frac{\sigma^2}{n})$
What if the population is not normal?
→ Central Limit Theorem (C.L.T)
$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$ approximately (as $n \to \infty$) for large n (say $n \geq 30$)

Example 4: The compressive strength of concrete has a mean of $\mu = 2500$ psi and $\sigma = 50$ psi.

Let $\bar{x}$ be the mean of 40 observations. What is $P(\bar{x} < 2490)$?

By the C.L.T:
$\bar{x} \sim N(\mu, \sigma^2/n)$ approximately, so
$P(\bar{x} < 2490) \approx \Phi\left(\frac{2490 - 2500}{50/\sqrt{40}}\right) = \Phi(-1.26) = 0.1038$ ← table III

Comparing means from 2 independant populations:
$\bar{X}_i$ is a sample mean from the ith population with mean μ_i and variance σ_i^2. n_i = i^{th} sample size.

$\bar{X}_i \sim N(\mu_i, \sigma_i^2/n_i) \Rightarrow \bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2; \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$

Example 5
$\mu_1 = 5000$ $\sigma_1 = 40$ $n_1 = 20$
$\mu_2 = 5050$ $\sigma_2 = 30$ $n_2 = 15$

Compute $P(\bar{X}_1 - \bar{X}_2 > 10)$
$= 1 - P(\bar{X}_1 - \bar{X}_2 \leq 10)$
$= 1 - \Phi\left(\frac{10 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\right)$

Remark: if $z > 3.99$ then $\Phi(z) \approx 1$

$= 1 - \Phi\left(\frac{10 - (-50)}{\sqrt{\frac{40^2}{20} + \frac{30^2}{15}}}\right)$

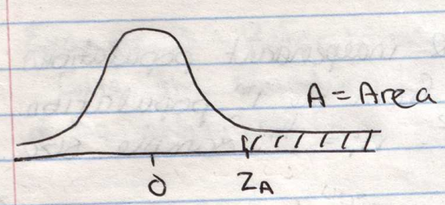
$= 1 - \Phi(5.07)$
$\therefore P(\bar{x}_1 - \bar{x}_2 > 10) \approx 1 - 1 = 0$

Confidence Intervals. (Interval Estimation)

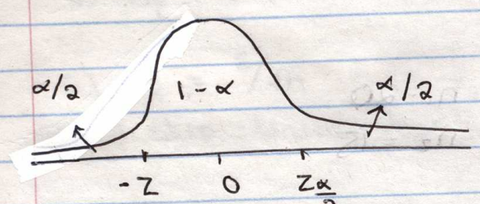
Estimation of μ
Case 1 · σ is known
· normal population or $n \geq 30$

$\bar{x} \sim N(\mu, \sigma^2/n)$

$= \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$



z_A is an upper quantile of $N(0,1)$ such that $P(Z > z_A) = A$



area to the right is $\alpha/2$

$$\Rightarrow 1-\alpha = P\left(-z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right)$$

$$= P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

A $(1-\alpha)100\%$ confidence interval for μ is

$$\bar{x} \pm z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

Note: $1-\alpha$ = confidence level.

June 08/11 Last Lecture: Computed the section on point estimation and sampling distributions.

Example 7

X_1, X_2, X_3 is a random sample of size $3 = n$. from a population with mean μ and variance σ^2. Show that both of the following estimators:

$$\hat{\mu}_1 = \frac{X_1 + X_2}{2} \quad \text{and} \quad \hat{\mu}_2 = X_2$$

are unbiased for estimating μ.

Recall: $\hat{\theta}$ is an unbiased estimator for $\hat{\theta}$ if $E[\hat{\theta}] = \theta$

$$E[\hat{\mu}_1] = E\left[\frac{X_1 + X_3}{2}\right] = \frac{1}{2}E[X] + \frac{1}{2}E[X_3] = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

$\therefore \hat{\mu}_1$ is unbiased

b) Which one is best for estimating μ?

Solution: $V[\hat{\mu}_1] = E[(\hat{\mu}_1 - E[\hat{\mu}_1])^2]$

$= E[(\hat{\mu}_1 - \mu)^2]$ definition of variance

$$V[\hat{\mu}_1] = V\left[\frac{1}{2}X_1 + \frac{1}{2}X_3\right] = \left(\frac{1}{2}\right)^2 V[X_1] + \left(\frac{1}{2}\right)^2 V[X_3]$$

$$= \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{1}{2}\sigma^2$$

$V[\hat{\mu}_2] = V[X_2] = \sigma^2$

$\therefore V[\hat{\mu}_1] < V[\hat{\mu}_2]$

$\therefore \hat{\mu}_1$ is a better estimator

Results from last class:

$\bar{X} = \frac{X_1 + \dots + X_n}{n}$ (sample mean)

$E[\bar{X}] = \mu$ and $V[\bar{X}] = \sigma^2/n$

If the population is normal, then $\bar{X} \sim N(\mu, \sigma^2/n)$

If $n \geq 30$ then $\bar{X} \sim N(\mu, \sigma^2/n)$ approximately

Also we discussed the (estimated) standard error

$\hat{\sigma}_{\bar{X}} = s/\sqrt{n}$, s = sample standard deviation

$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, $\hat{p}$ = sample proportion $= \frac{\#\text{ of successes}}{\#\text{ of trials}}$

Confidence Intervals: (Interval Estimation)

Recall: an upper quantile of order A for $N(0,1)$ is z_A. It satisfies $P(Z > z_A) = A$

Estimation of μ:

Case 1

- σ is known
- normal population or large sample ($n \geq 30$)

A $(1-\alpha)100\%$ confidence interval for μ is

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Terminology:
- $1-\alpha$ = confidence level or coefficient
- α = error rate
- $z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = E$ = maximum error of the estimate at a level of confidence of $1-\alpha$

Example 1

$n = 10$ observations with $\bar{x} = 64.46$. Assume a normal population with $\sigma = 4.5$.

a) Construct a 95% C.I. for μ.

Here $1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

A 95% C.I. for μ is:

$$\bar{x} \pm z_{0.025} \frac{\sigma}{\sqrt{n}} = 64.46 \pm 1.96\left(\frac{4.5}{\sqrt{10}}\right) = [61.67, 67.25)$$

from the table in the notes online

b) a 98% CI. for μ is:

$$\bar{x} \pm z_{0.02} \frac{\sigma}{\sqrt{n}} = 64.46 \pm 2.326\left(\frac{4.5}{\sqrt{10}}\right) = [61.15, 67.77]$$

Note: We use the length of the CI as a measure of precision

c) Compare the precision from a) and b)

95% CI is more precise when compared to a 98% CI.

Interpretation of Confidence

If we repeat a large number of times the collection of getting n observations for each sample, we compute a CI. About $(1-\alpha)100\%$ of the CI's will contain μ.

In practice, we want high confidence and good precision. We fix $1-\alpha$, often $1-\alpha = 0.95$, and then with the sample size n, we control the precision.

To be $(1-\alpha)100\%$ confident that the maximum error is at most E when estimating μ with $\bar{x}$, we need:

$$n \geq \left[\frac{z_{\alpha/2}\,\sigma}{E}\right]^2$$

If σ is unknown, we use s from another related study.

Example 2

$\sigma = 10$. We want to be 95% confident that the error is less than 0.3 when estimating μ. Find n.

Solution: $n \geq \left[\frac{z_{0.025}\,\sigma}{E}\right]^2 = \left[\frac{(1.96)(10)}{0.3}\right]^2 = 4268.44$

we need $n = 4269$ observations

Case 2

- σ is unknown
- large sample ($n \geq 40$)

Theorem: $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim N(0,1)$ approximately for large n

So we get an (approximate $(1-\alpha)\%$ for μ is $\bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

where s = sample standard deviation for $n \geq 40$

Example 4

$n = 45$ $\bar{x} = 24.5$ and $s = 2.2$

A 95% CI for μ is $\bar{x} \pm z_{0.025} \frac{s}{\sqrt{n}}$

$= 24.5 \pm (1.96) \frac{2.2}{\sqrt{45}}$

$= [23.56, 25.14]$

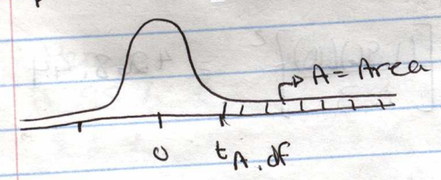
Case 3

- σ is unknown
- population is normal

Theorem: normal population

$\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t$ distribution with $r = n-1$ degrees of freedom

quantile $t_{A,df}$



$A(1-\alpha)100\%$ CI for μ when the population is normal is:

$\bar{x} \pm t_{\frac{\alpha}{2},(n-1)} \frac{s}{\sqrt{n}}$

Note: When $n \to \infty$, $z_{\frac{\alpha}{2}} = t_{\frac{\alpha}{2},\infty}$

Example 5

$n = 10$

$\bar{x} = 64.46$

$s = 4.7$

Assume that the population is normal. A 95% CI for μ is $\bar{x} \pm t_{0.025,9} \frac{s}{\sqrt{10}}$

$= 64.46 \pm 2.262 \left(\frac{4.7}{\sqrt{10}}\right)$

$= [61.10, 67.82]$

table V for t:

v	$t_{0.025,v}$
9	2.262
∞	1.96

to find z_A, we can use table V because $t_{A,v} = z_A$

For example:

$z_{0.025} = t_{0.025,\infty} = 1.96$

Remark: If v is large, then use $r = \infty$ degrees of freedom

Estimating a probability p:

Recall:

X = # of successes $\sim B(n,p)$

$\hat{p} = \frac{X}{n}$ ← estimater of p.

Note: $\dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \sim N(0,1)$ approximately when n is large

An (approximate) $100\%\,(1-\alpha)$ CI for p is:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

<u>Example 6</u>

$n = 300$ lenses where $x = 11$ are defective. Construct 95% CI for p = proportion of defective lenses.

A 95% CI for p is:

$$\hat{p} \pm z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \frac{11}{300} \pm (1.96) \sqrt{\frac{11/300\,(1 - 11/300)}{300}}$$

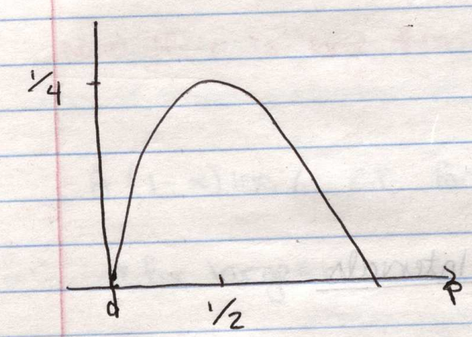
$$= [0.015, 0.058]$$

where $\hat{p} = 11/300$, $n = 300$, $z_{0.025} = 1.96$

To be $(1-\alpha)100\%$ confident that the error is less than E when estimating p, we need

$$n \geq \left[\frac{z_{\alpha/2}}{E}\right]^2 p(1-p)$$

We either use $\hat{p}$ from another related study or notice that:



$$p(1-p) = p - p^2$$

If we use $n \geq \left[\dfrac{z_{\alpha/2}}{E}\right]^2 \cdot \dfrac{1}{4}$ we are controlling for the error

<u>Example 7</u>

We want to be 95% confident that the error is less than 0.02 when estimating p. Find n

$$n \geq \left[\frac{z_{0.025}}{E}\right]^2 \cdot \frac{1}{4}$$

$$= \left[\frac{1.96}{0.02}\right]^2 \cdot \frac{1}{4}$$

$$= 2401$$

$\therefore$ take $n = 2401$.

Note: $\dfrac{\hat{p}-p}{\sqrt{\dfrac{p(1-p)}{n}}} \sim N(0,1)$ approximately when n is large

An (approximate) $100\%\,(1-\alpha)$ CI for p is:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

<u>Example 6</u>

$n = 300$ lenses where $x = 11$ are defective. Construct 95% CI for p = proportion of defective lenses.

A 95% CI for p is:

$$\hat{p} \pm z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \frac{11}{300} \pm (1.96) \sqrt{\frac{11/300\,(1 - 11/300)}{300}}$$

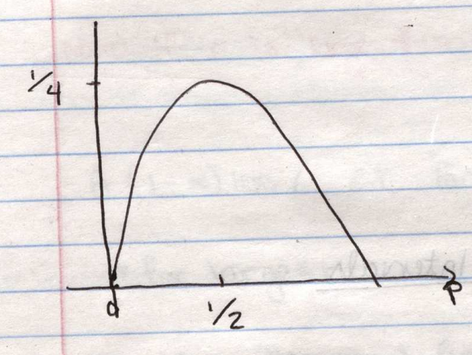
$$= [0.015, 0.058]$$

where $\hat{p} = 11/300$, $n = 300$, $z_{0.025} = 1.96$

To be $(1-\alpha)100\%$ confident that the error is less than E when estimating p, we need

$$n \geq \left[\frac{z_{\alpha/2}}{E}\right]^2 p(1-p)$$

We either use $\hat{p}$ from another related study or notice that:



$$p(1-p) = p - p^2$$

If we use $n \geq \left[\dfrac{z_{\alpha/2}}{E}\right]^2 \cdot \dfrac{1}{4}$ we are controlling for the error

<u>Example 7</u>

We want to be 95% confident that the error is less than 0.02 when estimating p. Find n

$$n \geq \left[\frac{z_{0.025}}{E}\right]^2 \cdot \frac{1}{4}$$

$$= \left[\frac{1.96}{0.02}\right]^2 \cdot \frac{1}{4}$$

$$= 2401$$

$\therefore$ take $n = 2401$.

June 30/11 Last lecture: <u>Confidence Intervals</u>

Estimation of μ:
3 cases.

<u>Case I</u>
- σ is known
- normal population or n is large (ie $n \geq 30$)

A $(1-\alpha)100\%$ CI for μ is $\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

<u>Case II</u>
- σ is unknown
- n is large ($n \geq 40$)

A $(1-\alpha)100\%$ CI for μ is $\bar{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

where s = sample standard deviation

<u>Case III</u>
- σ is unknown
- population is normal

A $(1-\alpha)100\%$ CI for μ is $\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$

where $t_{\frac{\alpha}{2}, n-1}$ is an upper quantile from t distribution with $r = n-1$ degrees of freedom of order $\frac{\alpha}{2}$

Note: If r is large, then use $t_{\frac{\alpha}{2}, \infty} = Z_{\frac{\alpha}{2}}$

A $(1-\alpha)100\%$ CI for p is $\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

* for large n.

We also discussed finding a sample size n to control the error of the estimate at a certain level of confidence

Descriptive Statistics

<u>Goal</u>: Describe the data from a random sample.
ie: independant observation from a random variable X

<u>Notation</u>
Random sample of size n: $x_1, x_2, x_3 \ldots x_n$

<u>Assumption</u>
We have numerical observations (ie: quantitative)

We will describe:
a) the center of the distribution
b) the dispersion of the distribution
c) the shape of the distribution

a) Measures of the center:
i) the sample mean (that is the average) is:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

* the average is very sensitive to extreme values

ii) median is another measure of the center

Percentile: Reorder $x_1, x_2, \dots, x_n$ in increasing order, to get the order statistics:

$$y_1 \le y_2 \le y_3 \le \dots \le y_n$$

the kth percentile is a value such that about $k\%$ of the values are at most equal to this value.

Think of n values as splitting the real lines into (n+1) bins.

Relative position of the kth percentile

$$= (n+1)\,k\% = \underset{\text{integer part}}{r} + \underset{\text{fractional part}}{a/b}$$

If $a/b = 0$, then the kth percentile is y_r.

If $a/b \neq 0$, then the kth percentile is

$$y_r + a/b\,(y_{r+1} - y_r)$$
$$= (1 - a/b)\,y_r + a/b\,y_{r+1}$$

Note: median = 50th percentile

Example

n = 8 observations. Describe the center?

sample mean: $\bar{x} = \dfrac{x_1 + \dots + x_8}{8}$

$= 74.004375$ mm

Compute the median?

It's position is: $(n+1)50\% = 9(50\%) = 4.5$

$\therefore$ the median is $y_4 + 0.5(y_5 - y_4)$

$= 74.003 + 0.5(74.004 - 74.003) = \boxed{74.0035 \text{ mm}}$

b) Describing the dispersion

sample variance $= s^2$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}^2}{n-1} = \frac{\left(\sum x_i^2\right) - \frac{1}{n}\left(\sum x_i\right)^2}{n-1}$$

sample standard deviation

$= s = \sqrt{s^2}$

Other measures of dispersion:

Range = max − min = $y_n - y_1$

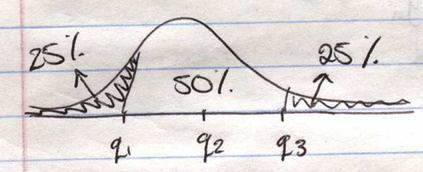
* ↑ not resistent to extreme values

Inter quartile range: IQR

$IQR = q_3 - q_1$

$= 3^{rd}$ quartile − 1st quartile

$= 75^{th}$ percentile − 25th percentile



* this IS resistant to extreme values

Example 1 (ctd)

Describe the dispersion of n = 8 diameters

sample variance $= s^2$

$$= \frac{\left(\sum x_i^2\right) - 8\bar{x}^2}{8-1} = 2.1696 \times 10^{-5}$$

$\Rightarrow$ sample standard deviation: $s = \sqrt{s^2} = 0.0047$ mm

The range is $y_n - y_1$

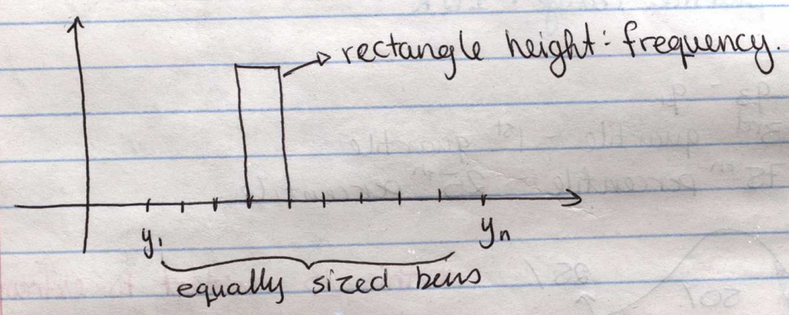
$= 74.015 - 74 = 0.015$

The IQR $= q_3 - q_1$
$= 74.005 - 74.00125$
$= 0.00375$ mm

Position of $q_1 = (n+1)\,25\%$
$= (8+1)\,25\%$
$= 2.25$

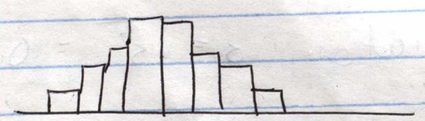
$\Rightarrow q_1 = y_2 + 0.25(y_3 - y_2)$
$= 74.001 + 0.25(74.002 - 74.001)$
$= 74.00125$

c) Shape of a distribution <u>histogram</u>:

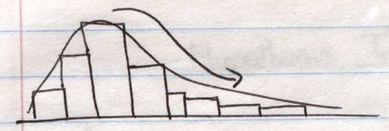


height can either be:
- frequency
- relative frequency
- density $= \dfrac{\text{relative frequency}}{\text{length of bin}}$

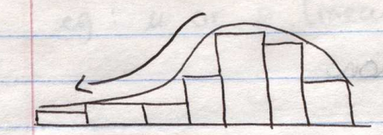
Describing the symmetry:



approximately symmetric

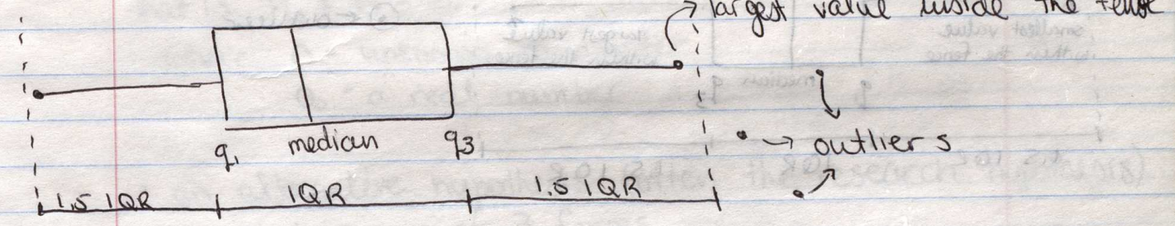


atypical on the <u>right</u>
(skewed to the right)



atypical on the <u>left</u>
(skewed to the left)

Box Plot



July 5/11 Assignment 4 due July 12 (Tuesday)
Office Hours this week:
Wednesday 5pm - 7pm
Friday 9am - 10:30am.

Last Lecture: Descriptive Statistics
Goal: describe the data from a random sample of size n

Center of the distribution
→ sample mean
→ sample median

Dispersion or Variability (sensitive to extreme values)
→ sample range = max − min

→ interquartile range

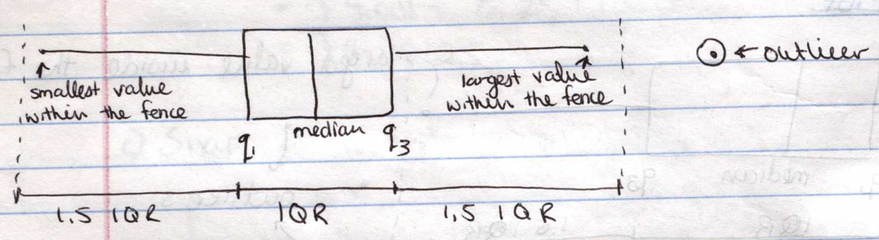
$IQR = q_3 - q_1$

$= 75^{th}$ percentile $- 25^{th}$ percentile

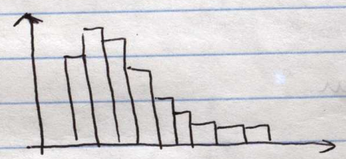
→ sample standard deviation

Finding Outliers

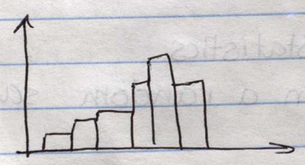
- Using a boxplot:



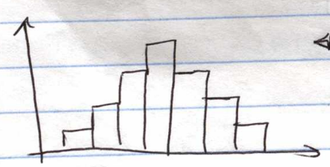
- Also using a histogram to describe the shape of the distribution



← skewed to the right.



← skewed to the left.



← approximately symmetric

Hypothesis Testing

Goal: To test hypotheses concerning population parameters

eg: μ or p (mean and probability of success/population proportion)

Step 1:

Write down the hypotheses:

- a null hypothesis H_0 which will be a simple hypothesis, that is $H_0: \theta = \theta_0$

  where θ = unknown parameter

  θ_0 = a real number

- an alternative hypothesis (often the research hypothesis) which will take one of 3 forms

  a) $H_1: \theta > \theta_0$ (right sided)

  b) $H_1: \theta < \theta_0$ (left sided)

  c) $H_1: \theta \neq \theta_0$ (two sided)

These are composite alternatives: ie contains more than one value.

Example 1

$n = 45$ tires

and observe $\bar{x} = 65\ 158.7$ km

Using an old process $\mu = 65\ 000$km

Claim that the new μ is larger than 65 000km.

We want to test

$$\begin{cases} H_0: \mu = 65\ 000\text{km} \\ H_1: \mu > 65\ 000\text{km} \end{cases}$$

$\bar{x}$ is our estimate of the new mean μ.
Since $\bar{x} = 65\,158{,}7$ km $> 65\,000$ km, we have evidence in favor of the alternative H_1

Suppose that we conclude that $H_1: \mu > 65\,000$ is true if $\bar{x} > 65\,000$. If there is no change, ie the new process is not better than the old process, what are the chances that we conclude that $\mu > 65000$

We want $P(\bar{X} > 65\,000 \mid \mu = 65\,000)$

Note: $\bar{X} \sim N(\mu = 65\,000 ; \frac{\sigma^2}{n})$

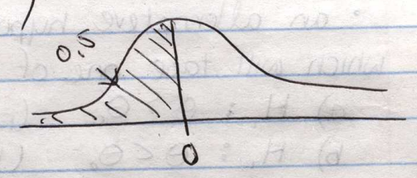
so $P(\bar{x} > 65\,000 \mid \mu = 65\,000)$

$$= 1 - \Phi\left(\frac{65\,000 - 65\,000}{\sigma/\sqrt{n}}\right)$$

$$= 1 - \Phi(0)$$

$$= 1 - 0{,}5$$

$$= 0{,}5$$



We need a better way to assess the strength of the evidence

New Terminology

- Test Statistic: a statistic used to test H_0 against H_1
- Critical Region: a set of values for the test statistic such that if the observed value falls in this critical region, then we reject H_0 and accept H_1
- Type I Error: Accept H_1 when H_0 is true.

$\alpha = P(\text{type I error})$

$\alpha = P(\text{fall in CR.} \mid H_0)$

Note: α = significance of the test. Usually want this to be small.

- Type II Error: failing to reject H_0 when H_1 is true

$\beta(\theta_1) = P(\text{not falling in the CR.} \mid \theta = \theta_1 \in H_1)$

Example 2

$$\begin{cases} H_0: \mu = 65\,000 \text{ km} \\ H_1: \mu > 65\,000 \text{ km} \end{cases}$$

Recall: $n = 45$.
$\sigma = 1000$ km

Use the following critical region:
Reject H_0 if $\bar{x} > 65\,400$

a) Compute $\alpha = P(\text{type I error})$

Solution:

$$\alpha = P(\text{reject } H_0 \mid \mu = 65\,000)$$

$$= P(\bar{X} > 65\,400 \mid \mu = 65\,000)$$

$$= 1 - \Phi\left(\frac{65\,400 - 65\,000}{\sigma/\sqrt{n} = 1000/\sqrt{45}}\right)$$

$$= 1 - \Phi(2{,}68)$$

$$= 1 - 0{,}9963 \quad \text{(from table III)}$$

$$= 0{,}0037$$

b) If $\mu = 66\,000$ km $\in H_1$

$$\beta(66\,000) = P(\text{not rejecting } H_0 \mid \mu = 66\,000)$$

$$= P(\bar{X} \leq 65\,400 \mid \mu = 66\,000)$$

$$= \Phi\left(\frac{65\,400 - 66\,000}{1000/\sqrt{45}}\right)$$

$$= \Phi(-4{,}02)$$

$$\approx 0$$

Remark: In practice, we fix α beforehand, eg: $\alpha = 5\%$, and then we get the corresponding critical region

there are 3 cases:

Case I

- normal population or large sample $n \geq 30$
- σ is known

Test statistic is:

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

$Z_0 \sim N(0,1)$ if $H_0 : \mu = \mu_0$

Case II

- σ is unknown
- large sample ($n \geq 40$) so the standard mean is approx normal

We use : $Z_0 = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$

$Z \sim N(0,1)$ if $H_0 : \mu = \mu_0$

Case III

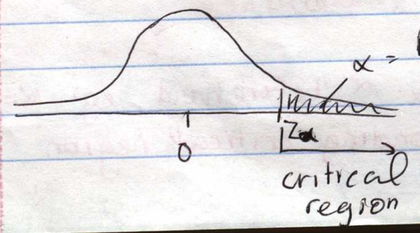
- σ is unknown
- normal population

We use : $T_0 = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$

$T_0 \sim t$ distribution with $r = n-1$ degrees of freedom.

Consider the case where $Z_0 \sim N(0,1)$ if $\mu = \mu_0$

Consider $H_1 : \mu > \mu_0$



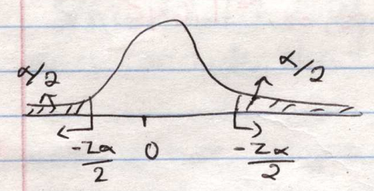
$\alpha = P(\text{type 1 error})$ (prob of alpha falling in the CR)

CR: we reject H_0 and accept that $\mu > \mu_0$ if $Z > Z_\alpha$

For $H_1 : \mu < \mu_0$

we reject H_0 if $Z_0 < -Z_\alpha$

For $H_1 : \mu \neq \mu_0$



We reject H_0 if $Z_0 < -Z_{\alpha/2}$ or if $Z_0 > Z_{\alpha/2}$

Equivalently : Reject H_0 if $|Z_0| > Z_{\alpha/2}$

Note: if $T_0 \sim t(n-1)$ under $\mu = \mu_0$, we use the C.R.:

i) $t_0 > t_{\alpha, n-1}$ if $H_1 : \mu > \mu_0$

ii) $t_0 < -t_{\alpha, n-1}$ if $H_1 : \mu < \mu_0$

iii) $|t_0| > t_{\alpha/2, n-1}$ if $H_1 : \mu \neq \mu_0$

Example 3

Test $\begin{cases} H_0 : \mu = 65 \\ H_1 : \mu > 65 \end{cases}$

Note: measured in 1000 km.

Test the hypotheses at $\alpha = 5\%$

We are given: $n = 50$, $\sum x_i^2 = 163\,910.1$ and $\sum x_i = 2\,859.3$

Solution:

- σ is unknown

its estimate is

$$s = \sqrt{\frac{(\sum x_i^2) - [(\sum x_i)^2/n]}{n-1}} = 2.8535$$

Joanna

Also, n is large, ie ($n \geq 40$) so the test statistic is:

$$Z_0 = \frac{\bar{X} - 65}{s/\sqrt{50}}$$

CR. we will reject H_0 if $Z_0 > Z_\alpha = Z_{0.05} = 1.645$

Recall: In table V, $Z_{0.05} = t_{0.05, \infty} = 1.645$

The observed value is:

$$Z_0 = \frac{\bar{x} - 65}{s/\sqrt{50}}$$

$$= \frac{\left[\frac{\Sigma x_i}{n}\right] - 65}{\frac{2.3535}{\sqrt{50}}} = -19.36$$

Conclusion:

Since $Z_0 < 1.645$, then we ~~[illegible]~~ can not reject H_0

Example 4

Test $\begin{cases} H_0 : \mu = 6 \\ H_a : \mu \neq 6 \end{cases}$

We select $n = 10$ observations, and observe $\bar{x} = 5.7$, $s = 0.2$. Assume a normal population. Give the conclusion at $\alpha = 5\%$.

Conditions:

- σ is unknown
- normal population

Test Statistic is

$$T_0 = \frac{\bar{x} - 6}{s/\sqrt{10}}$$

CR: We reject H_0 if $|t_0| > t_{0.025, 9} = 2.262$

Observed value is

$$t_0 = \frac{5.7 - 6}{0.2/\sqrt{10}} = -4.74$$

Since $|t_0| > 2.262$, then we can reject H_0 and conclude that $\mu \neq 6$

July 7/11 Last lecture: Hypothesis Testing

Statement of the Hypotheses:

a) Right Sided Test:

$H_0: \mu = \mu_0$ where μ_0 is a real number.

against $H_1: \mu > \mu_0$

b) Left sided test

$H_0: \mu = \mu_0$ against $H_1: \mu < \mu_0$

c) $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ 2 sided test

→ just to see if it's different.

Test Statistic

- $Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ if the population is normal or $n \geq 30$ and σ is known.
- If $n \geq 40$ and σ is unknown, we use $Z_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
- If the population is normal and σ is unknown, we use $T_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t$ with $r = n - 1$ degrees of freedom under H_0

In practice, if we use a critical region, then we fix α = level of significance
$= P(\text{fall in the critical region} \mid \mu = \mu_0)$

Remark: If the observed value of the test statistic falls in the critical region, then we reject H_0 and accept H_1. Otherwise, we fail to reject H_0.

<u>Example 4</u> Use $\alpha = 1\%$ to test $H_0: \mu = 6$
$H_1: \mu \neq 6$

Given: Sample size $= n = 10$, sample mean $\bar{x} = 5.7$ sample standard deviation $\Rightarrow s = 0.2$ and we assume that the population is normal.

The test statistic is:

$$T_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\bar{x} - 6}{s/\sqrt{10}} \sim t \text{ distribution with } n-1 = 9 \text{ degrees of freedom if } H_0 \text{ is true.}$$

Note: This is a 2-sided alternative because all we want to do is show that $\mu \neq 6$.

So the critical region is:
we will reject H_0 if $|t_0| > t_{\frac{\alpha}{2}, n-1} = t_{0.005, 9} = 3.25$

The observed value is

$$t_0 = \frac{\bar{x} - 6}{s/\sqrt{10}} = \frac{5.7 - 6}{0.2/\sqrt{10}} = -4.74$$

Since $|t_0| = 4.74 > 3.25$, thus we can reject H_0. At $\alpha = 1\%$, we can conclude that $\mu \neq 6$.

P-Value Method

Goal: Measure the significance of the evidence.

Definition: p-value = probability of observing a value as extreme (extreme?) as the current value. (ie: in favor of the alternative) assuming that H_0 is true.

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{or} \quad Z_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim N(0,1)$$

under H_0.

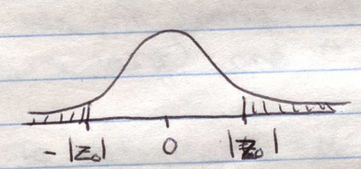
$H_1: \mu > \mu_0$:
$p = P(Z > z_0) = 1 - \Phi(z_0)$

Small p-values are evidence against H_0

We reject H_0 if $p < \alpha$ where α = levels of significance

$H_1: \mu < \mu_0$
$p = P(Z < z_0) = \Phi(z_0)$

$H_0: \mu \neq \mu_0$



$p = P(Z < -|z_0| \text{ or } Z > |z_0|) = 2P(Z > |z_0|)$
$p = 2[1 - \Phi(|z_0|)]$

$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t$ with $r = n-1$

$H_1: \mu > \mu_0 \rightarrow p = P(T > t_0)$
$H_1: \mu < \mu_0 \rightarrow p = P(T < t_0)$
$H_1: \mu \neq \mu_0 \rightarrow p = 2P(T > |t_0|)$

Example 4b refer to example 3.

$H_0: \mu = 65$

$H_1: \mu > 65$ (right sided test)

$\bar{x} = 57{,}89$

$s = 2{,}8506$

$n = 50$ ← large sample

$z_0 = \frac{\bar{x} - 65}{s/\sqrt{50}} = -19{,}38$

$p = P(Z > -19{,}38)$ since it is a right sided test

$= 1 - \underbrace{\Phi(-19{,}38)}_{\sim 0}$

$= 1 - 0 = 1$

Since $p > \alpha = 5\%$ then we do not reject H_0

Note: Only reject H_0 if $p < \alpha$

Example 5: Refer to example 4a

a) Redo Question 4a with a p-value.

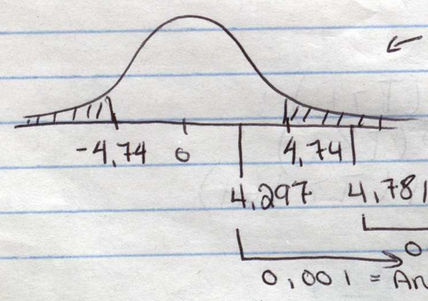
$H_0: \mu = 6$

$H_1: \mu \neq 6$

$t_0 = \frac{\bar{x} - 6}{s/\sqrt{10}} = -4{,}74$

$p = 2P(T > |-4{,}74|)$

$= 2P(T > 4{,}74)$



← t graph with 9 degrees of freedom.

← table V

$\therefore 0{,}0005 < P(T > 4{,}74) < 0{,}001$

and since the p value is twice that, multiply by 2.

$\Rightarrow 0{,}001 < p < 0{,}002$ (range of significance)

b) Redo but make it a one sided test $H_1: \mu < 6$.

$p = P(T < -4{,}74)$ because the t table is symmetric, we

$= P(T > 4{,}74)$ ← can say...

$\therefore 0{,}0005 < p < 0{,}001$

Conclusion for part a: Since the p value is less than $0{,}002$ which is less than $\alpha = 0{,}05$, we can reject H_0 to accept that $\mu \neq 6$.

Part b Since $p < 0{,}001 < 0{,}05 = \alpha$ therefore we can reject H_0 to accept a $\mu < 6$

Controlling Errors of Type II with the sample size

We will assume that the population is normal (or n is large) and σ is known.

To test $H_0: \mu = \mu_0$

We use $Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ where $Z_0 \sim N(0,1)$ if H_0 is true.

BUT if $\mu = \mu_1 \in H_1$, then Z_0 is not a standard normal.

We will restandardize:

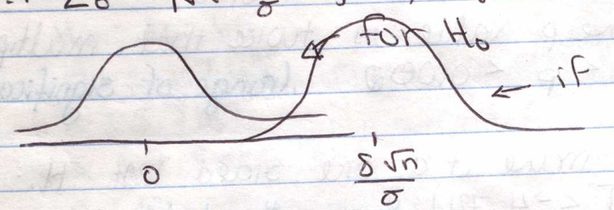
$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - \mu_1 - \mu_0 + \mu_1}{\sigma/\sqrt{n}}$$

$$= \frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} + \frac{(\mu_1 - \mu_0)\sqrt{n}}{\sigma}$$

linear function of the standard normal with some constant, so the standard normal gets shifted.

Hilroy

$\therefore Z_0 \sim N\left(\frac{\delta}{\sigma}\sqrt{n}, 1\right)$ where $\delta = \mu_1 - \mu_0$



Note: We control $\beta(\mu_1)$ with n, at α = level of significance

a) one sided test:

$$n = \left(\frac{Z_\beta + Z_\alpha}{\mu_1 - \mu_0}\right)^2 \sigma^2$$

b) 2 sided test

$$n = \left(\frac{Z_\beta + Z_{\alpha/2}}{\mu_1 - \mu_0}\right)^2 \sigma^2$$

Example 6

$H_0 : \mu = 6$ normal population $n = 10$

$H_1 : \mu \neq 6$ $\sigma = 0.2$

If $\mu = 6.5$, what is the probability of error type 2 if we use a signifacance level of 5%

or $P(\text{error type II} \mid \mu = 6.5) = \beta(6.5)$ if $\alpha = 5\%$

$\beta(6.5) = P[\text{not falling in CR} \mid \mu = 6.5]$

Test Statistic:

$$Z_0 = \frac{\bar{x} - 6}{\sigma/\sqrt{n}} \sim N\left(\frac{(\sqrt{n}(\mu_1 - \mu_0))}{\sigma}, 1\right)$$

$$\sim N\left(\frac{\sqrt{10}(6.5 - 6)}{0.2}, 1\right)$$

$$\sim N(7.906, 1)$$

CR: is $|Z_0| > Z_{\alpha/2} = Z_{0.025}$
$= 1.96$.

$\beta(6.5) = P(|Z_0| \leq 1.96)$

$\beta(6.5) = P(-1.96 \leq Z_0 \leq 1.96)$

$$= \Phi\left(\frac{1.96 - 7.9069}{1}\right) - \Phi\left(\frac{-1.96 - 7.9069}{1}\right)$$

$$= \Phi(-5.9469) - \Phi(-9.8669)$$

Example 7

$H_0 : \mu = 6$

$H_1 : \mu \neq 6$

Find n for $\beta(6.1) = 0.05$ if $\alpha = 5\%$ and $\sigma = 0.2$

Solution:

2 sided test:

$$n = \left(\frac{Z_\beta + Z_{\alpha/2}}{\mu_1 - \mu_0}\right)^2 \sigma^2$$

$$= \left(\frac{Z_{0.05} + Z_{0.025}}{6.1 - 6}\right)^2 (0.2)^2$$

$$= \left(\frac{1.645 - 1.96}{0.1}\right)^2 (0.2)^2$$

$$= 51.98$$

take $n = 52$

$\therefore$ so under these conditions to notice a difference between 6 and 6.5, you would have to choose $n = 52$.

July 12/11 Assignment 4 due Thursday.

Last Class: Hypothesis Testing.

→ p-value Method

p-value = p

= probability of observing a value as extreme as our observed value (ie: in favor of the alternative hypothesis) assuming that the null hypothesis is true.

Interpret: Small p-values are evidence in favor of H_1

Decision Rule: We reject H_0 if $P < \alpha$ where α = level of significance

We also discussed choosing n for controlling our risk of committing an error of type II at a level of significance.

Example:

We want to test

$H_0: \mu = 25$ against

$H_1: \mu \neq 25$.

We select $n = 19$ observations. We will assume that the population is normal. We observe

$\bar{x} = 27.5$ and $s = 5.1$

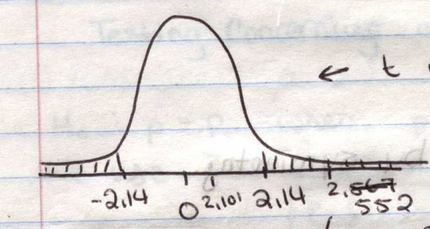
Solution:

- normal population
- σ is unknown

} We will use a t-test statistic.

The observed value of the test statistic is:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{27.5 - 25}{5.1/\sqrt{19}} = 2.14$$



← t with $r = n-1 = 18$ degrees of freedom

p-value $= 2P(T > 2.14)$ since it is a 2-sided test

From table V,

$P(T > 2.101) = 0.025$

$P(T > 2.552) = 0.01$

so

$0.01 < P(T > 2.14) < 0.025$

$\Rightarrow 0.02 < 2P(T > 2.14) < 0.05$

$0.02 < P < 0.05$

If $\alpha = 5\%$, we know that our P value is less than 5%, so we can reject H_0 to accept that $\mu \neq 25$.

Example

Test $H_0: \mu = 125$ against

$H_1: \mu < 125$

From $n = 26$ observations we observe $\bar{x} = 97.5$ ← mean

and $s = 15.0$ ← standard deviation

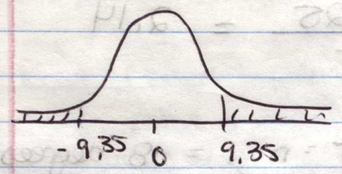
Assume a normal population. Compute the p-value.

Solution: The observed value of the test statistic is:

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{97.5 - 125}{15/\sqrt{26}} = -9.35$$

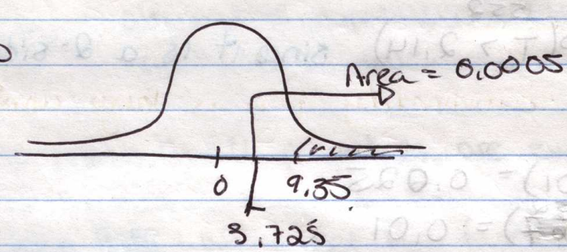
$p = P(T_0 < -9.35)$

where $T_0 \sim t$ with $r = n-1 = 25$ degrees of freedom



so $p = P(T_0 > 9{,}35)$ by symmetry.

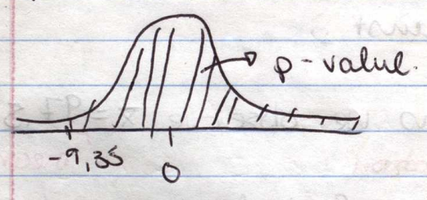
so



so since $P(T > 3{,}725) = 0{,}0005$
$\Rightarrow \quad p < 0{,}0005$

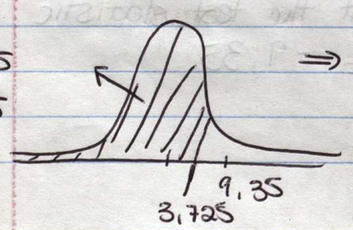
For the same data, suppose we were testing that the mean $H_0: \mu = 125$
$H_1: \mu > 125$

p-value $= P(T_0 > -9{,}35)$ where $T_0 \sim t$ with 25 df.



so from symmetry, $P(T_0 > -9{,}35) = P(T_0 < 9{,}35)$

$1 - 0{,}0005$
$= 0{,}9995$



$\Rightarrow p > 0{,}9995$

If $\alpha = 5\%$, we can not reject H_0 because $p > \alpha$
aka $0{,}9995 > 0{,}05$.

Testing Concerning a Proportion p

$H_0: p = p_0$ where $p_0 \in \mathbb{R}$

We use $z_0 = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$

$Z_0 \sim N(0,1)$ approximately if n is large and H_0 is true.

<u>Example 8</u>

$H_0: p = 0{,}02$
$H_1: p \neq 0{,}02$

Here p = true proportion of the defectives
we observe 8 defectives among $n = 500$ items.
Does the data suggest that $p \neq 0{,}02$ at $\alpha = 5\%$?

Compute the observed value of the z-test statistic.

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \frac{8/500 - 0{,}02}{\sqrt{\dfrac{(0{,}02)(0{,}98)}{500}}} = -0{,}64.$$

$$\begin{aligned}
\text{p-value} &= 2P(Z > |-0{,}64|) \\
&= 2P(Z > 0{,}64) \\
&= 2[1 - \Phi(0{,}64)] \\
&= 2[1 - 0{,}7389] \quad \text{from table III} \\
&= 0{,}5222
\end{aligned}$$

Since the p-value is larger than 5%, we will not reject H_0.

Assessing Normality

- One tool is a histogram. It should not be highly skewed if the data is normal
- A second tool is a quantile-quantile plot (QQ-plot)

Basic Idea:

A sample: $x_1, x_2, x_3, \ldots, x_n$ of size n

Obtain the order statistics:

$$y_1 \leq y_2 \leq \ldots \leq y_n$$

Order of the ith quantile is:

$$p_i = \frac{i - 3/8}{n + 1/4}$$

Obtain the quantile (lower) of the same order, say z_i

$$z_i = \Phi^{-1}(p_i) \quad \text{ie} \quad P(Z \leq z_i) = p_i$$

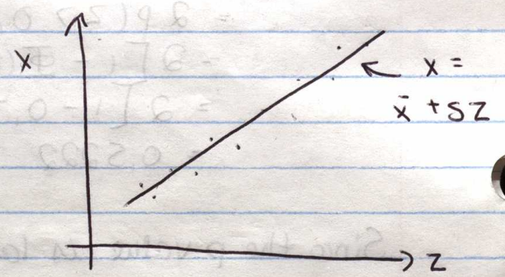
QQ-plot: We plot the points (y_i, z_i) for $i = 1, \ldots, n$.
or (z_i, y_i) for $i = 1, \ldots, n$

Note: $X \sim N(\mu, \sigma^2)$

$$\Rightarrow Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma} X - \frac{\mu}{\sigma} \sim N(0,1)$$

We should see a linear tendancy in the QQ-plot
Consider z-scores on the horizontal axis: we have
$x = \sigma z + \mu$

In practice, we plot the line $x = \bar{x} + sz$



Remark: If there is a systematic deviation away from the line, then we have evidence against normality.

Simple Linear Regression

y = response variable
x = regressor or predictor

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where ε = random error $\sim N(0, \sigma^2)$

$\sigma^2 = V[\varepsilon]$ ← variance of the random error

Note: σ^2 is a constant $\Rightarrow$ variance consistancy assumption

Properties: $E[y|x] = \beta_0 + \beta_1 x + E[\varepsilon]$
$= \beta_0 + \beta_1 x$

$V[y|x] = V[\varepsilon] = \sigma^2$

$y|_x \sim N(\beta_0 + \beta_1 x, \sigma^2)$

Least Squares Criterion:

$$\mathcal{L} = \sum_{i=1}^{n} [y_i - \hat{y}_i]^2$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ for $i = 1, 2, \ldots, n$

$\hat{y}_i$ is called the ith fitted value

Solution:

July 14/11 Final Exam - THURSDAY JULY 21 4-7pm ART 026

4 short answer questions - 16 pts
14 multiple choice questions - 14 pts

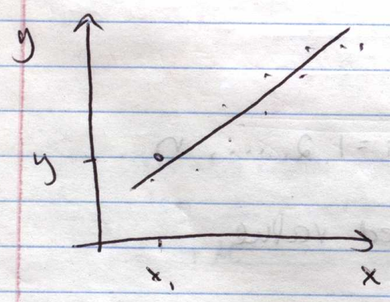
- Open book
- responsible for bringing the tables (II, III, IV)
- non graphing/non programmable calculators only
- make a cheat sheet!
- Office Hours
  Tuesday & Wednesday 5-7pm.
  585 King Edward (B 07 - D.)

Simple Linear Regression

Goal: Describe a response variable variable Y as a linear of a regressor x

Step 1

Produce a scatter plot of y versus x



Using a sample of n paired observations $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ If the tendancy is linear then using a simple linear regression is reasonable

Model:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

where ε = random noise $\sim N(0, \sigma^2)$

Estimating the parameters β_0, β_1 (ie: find the line of best fit)

The fitted line or Estimated line is denoted:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

A Measure of fit

$$\mathcal{L} \; \Sigma_{i=1}^{n} \left[y_i - \hat{y}_i \right]^2 = \Sigma_{i=1}^{n} \left[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x) \right]^2$$

least-squares criterion

We want to minimize $\mathcal{L}$ with respect to $\hat{\beta}_0, \hat{\beta}_1$

The solution: $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xy}}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$S_{xx} = \Sigma_{i=1}^{n} (x_i - \bar{x})^2$$
$$= (\Sigma x_i^2) - n\bar{x}^2$$
$$= (\Sigma x_i^2) - \tfrac{1}{n} (\Sigma x_i)^2$$

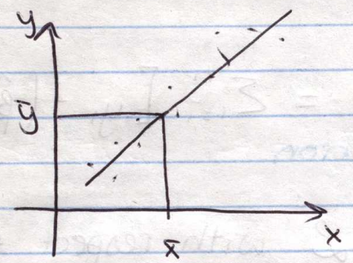
$$S_{yy} = \Sigma_{i=1}^{n} (y_i - \bar{y})^2$$
$$= (\Sigma y_i^2) - n\bar{y}^2$$
$$= (\Sigma y_i^2) - \tfrac{1}{n} (\Sigma y_i)^2$$

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

$$S_{xy} = (\Sigma x_i y_i) - n\bar{x}\bar{y}$$

$$= (\Sigma x_i y_i) - \tfrac{1}{n}(\Sigma x_i)(\Sigma y_i)$$

Note: S_{xy} contains about how x and y vary together



S_{xy} should have the same sign as the slope of the line

Sample Covariance $= \dfrac{S_{xy}}{n-1}$

Covariance: often used as a measure of linear fit

An alternative measured based on S_{xy}:

$$r = \text{correlation} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Note: $d_x = \begin{pmatrix} x_1 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}$ $\quad d_y = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}$

$S_{xy} = d_x \cdot d_y$ ← dot product

$$S_{xx} = \Sigma (x_i - \bar{x})^2 = \|d_x\|^2$$

Cauchy-Swartz Inequality

$$-1 \le \frac{d_x \cdot d_y}{\|d_x\| \|d_y\|} \le 1$$

$\therefore -1 \le r \le 1$ where r = correlation

Note: $r = 1$ or $r = -1$ if the points fall exactly on the line.

Example

$n = 11$

$\Sigma x_i = 2000$ $\quad \Sigma x_i^2 = 532\,000$

$\Sigma y_i = 8{,}35$ $\quad \Sigma y_i^2 = 9{,}1097$

$\Sigma x_i y_i = 2175{,}4$

$S_{xx} = (\Sigma x_i^2) - \tfrac{1}{n}(\Sigma x_i)^2 = \cancel{657{,}21818}\ 168\,363{,}6364$

$S_{xy} = (\Sigma x_i y_i) - \tfrac{1}{n}(\Sigma x_i)(\Sigma y_i) = 657{,}21818$

Estimate of the slope $\hat{\beta}_i = \dfrac{S_{xy}}{S_{xx}} = 0{,}0039$

Estimate of the intercept is $\hat{\beta}_0 = \bar{y} - \hat{\beta}_i \bar{x} = 0{,}05$

The fitted line is $\hat{y} = 0{,}05 + 0{,}0039x$

b) Estimate the mean evaporation coefficient at $x = 140$

We want to estimate

$\mu_{y/x=140} = \beta_0 + \beta_1(140)$

Thus $\hat{\mu}_{y/x=140} = \hat{\beta}_0 + \hat{\beta}_1(140) = 0.5965$

c) Predict the evaporation coefficient at $x=140$

$\hat{y}|_{x=140} = \hat{\beta}_0 + \hat{\beta}_1(140) = 0.5965$

Estimating $\sigma^2 = V[\varepsilon]$

SSE = error sum of squares

$= \sum_{i=1}^{n} [y_i - \hat{y}_i]^2$

$= \sum_{i=1}^{n} e_i^2$ where $e_i = i^{th}$ residual or observed error $= y_i - \hat{y}_i$ = observed response − fitted response

$\hat{\sigma}^2 = \frac{SSE}{n-2} = MSE$ = mean squared error

Note: $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$

Estimate σ^2

Example 3b)

$\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{S_{yy} - \hat{\beta}_1 S_{xy}}{n-2} = \frac{5.7875}{9} = 0.63405 = MSE$

a) Recall from example 1

$(x_2, y_2) = (60, 037)$

Compute the 2nd residual

$e_2 = y_2 - \hat{y}_2 = 0.37 - (0.05 + 0.0039(60))$
$= 0.284$

Correlation:

$r = \frac{S_{xy}}{\sqrt{S_{xx} - S_{yy}}} = \frac{S_{xy}}{S_{xx}} \frac{S_{xx}}{\sqrt{S_{xx} S_{yy}}}$

$= \hat{\beta}_1 \sqrt{\frac{S_{xx}}{S_{yy}}} = 0.55.$

PRACTICE FINAL

#1. The p.m.f. of x is

$f(x) = c(1 + |x-4|)$, $x = 3, 4, 5$

a) Find c we know that the sum of a p.m.f. = 1, so evaluate and solve for c.

$1 = \sum_{x \in R_x} f(x) = c(1+|3-4|) + c(1+|4-4|) + c(1+|5-4|)$
$= 5c$

$1 = 5c$

$1/5 = c$

b) Probability that $x > 4$ given that $x = 4$

$P(x=4 \mid x \geq 4)$

$= \frac{P(\{x=4\} \cap \{x \geq 4\})}{P(x \geq 4)}$

$$= \frac{P(x=4)}{P(x=4)+P(x=5)}$$

$$= \frac{1/5}{1/5 + 2/5} = \boxed{\frac{1}{3}}$$

c) Expected value of x = ?

$$= E[x] = \mu_x = \sum_{x \in R_x} x f_x(x)$$

$$= 3(2/5) + 4(1/5) + 5(2/5)$$

$$= 20/5 = \boxed{4}$$

d) $P(x \geq \mu)$

$$= P(x \geq 4) = P(x=4) + P(x=5)$$

$$= 1/5 + 2/5 = \boxed{3/5}$$

#2. Claim: $\mu = 14$mg.

$n = 25$, $\bar{x} = 14.112$, $s = 0.194$

b) Do we have sufficient evidence at $\alpha = 5\%$ to conclude that $\mu > 14$? Assume normality

We want to test

$H_0: \mu = 14$

$H_1: \mu > 14$ → Right sided

$\left.\begin{array}{l} \cdot \sigma \text{ is unknown} \\ \cdot \text{normal population} \end{array}\right\}$ t test

Observed value of the test statistic is:

$$t_0 = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\bar{x} - 14}{s/\sqrt{25}} = \frac{14.112 - 14}{0.194/5} = 2.89$$

C.R. We will reject H_0 if $t_0 > t_{\alpha, n-1} = t_{0.05, 24} = 1.711$ (table V)

Conclude: Since $t_0 > 1.711$, we can reject H_0

At $\alpha = 5\%$, we have sufficient evidence that $\mu > 14$

c) Construct a 95% confidence interval for μ

The C.I. is:

$$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$$

$$= 14.112 \pm (2.064)\left(\frac{0.194}{\sqrt{25}}\right)$$

$$= 14.11 \pm 0.08$$

$$[14.034, 14.194]$$

#4. $H_0: \mu = 4.5$

$H_1: \mu \neq 4.5$

$\sigma = 1.5$ normal population $n = 25$ observations.

a) $\alpha = 5\%$ $\beta(4) = P[\text{error of type II} \mid \mu = 4]$

$$= P[\text{not rejecting } H_0 \mid \mu = 4]$$

CR. We reject H_0 if $|z_0| > z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$.

$$\beta(4) = P[|z_0| \leq 1.96 \mid \mu = 4]$$

If $\mu = 4$, then $z_0 \sim N\left(\frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma}, 1\right)$

$$= N\left(\frac{\sqrt{25}\,(4-4.5)}{1.5}, 1\right)$$

$$= N(-1.6667, 1)$$

$$\beta(4) = P(-1.96 \le Z_0 \le 1.96 \mid \mu = 4)$$

$$= \Phi\left(\frac{1.96-(-1.6667)}{\sqrt{1}}\right) - \Phi\left(\frac{-1.96-(-1.6667)}{\sqrt{1}}\right)$$

$$= \Phi(3.63) - \Phi(-0.29)$$

$$= 0.9999 - 0.3895$$

$$=$$