MAT2377A - Probability and Statistic for Engineers

Prof. Mahmoud Zarepour; Transcribed by Ted Morin

Created January 7th, 2014, Updated Marth 25^{th}, 2014

Sets

Set A set is a collection of items.

$A = \{2, 3, 4\}$

$A = B$ implies that any $x \in A \to x \in B$

Operations on Sets

Let A and B be two sets.

$A \cap B = A$ intersects $B = A$ and $B = \{x : x \in A \text{ and } x \in B\}$

$A \cup B = A$ union $B = A$ or $B = \{x : x \in A \text{ or } x \in B\}$

(1) Example:

$ = {2,3,4}$ and $B = \{3, 4, 5\} \to A \cap B = \{3, 4\}, A \cup B = \{2, 3, 4, 5\}$

Sample Space

Sample Space All the possible outcomes in a random experiment.

$S =$ the set of outcomes in a random experiment.

(2) Example: Flip a coin

$S = \{H, T\}$

(3) Example: Roll a die

$S = \{1, 2, 3, 4, 5, 6\}$

(4) Example: Flip a coin until a heads appears.

$S = \{H, TH, TTH, \cdots\}$

(5) Example: For a circle of radius 1, any point within the circle is what set?

$S = \{(x \times y) : x^2 + y^2 \leq 1\}$

Countable Set A set in which you can list the entries, and an uncountable set is one which will not let you list the entries.

Event

Event Any subset of S (sample space) is called an event.

(6) Example:

Coin is heads: $E = \{H\} \subset \{H, T\}$
So E is an event.

(7) Example:

Die is on even number: $E = \{2, 4, 6\}$

(8) Another example:

$E = \{(x, y) : x^2 + y^2 \leq \frac{1}{4}\}$

Enumeration (Counting)

1. Multiplication Rule.

   If task 1 can be completed in n_1 different ways and task 2 can be completed in n_2 different ways, then both tasks can be completed by $n_1 \times n_2$ ways.

   (9) Example: How many 5 digit numbers are there?

   Task 1: 1,2,3,4,5,6,7,8,9. $n_1 = 9$.
   Task 2: 0-9. $n_2 = 10$.
   ...
   Task 5: 0-9. $n_5 = 10$.
   $9 \times 10^4 = 90,000$

   (10) Example: How many 5 digit numbers with no repeats do we have?

   $n_1 = 9, n_2 = 9, n_3 = 8, n_4 = 7, n_5 = 6$
   $= 9 * 9 * 8 * 7$

   (11) Example: In how many different ways can we write 1,2,3,4,5 in unique order?

   $5 * 4 * 3 * 2 * 1 = 5!$

   (12) Example: How many different orders of these letters is possible, where repeat letters are to be treated as non-unique: INDEPENDENT?

   $11!/(3! * 2! * 3!)$

2. Permutation

   (13) Example: With letters a, b, c, d, write all possible permutations with 2 letters.

ab, ac, ad,
ba, bc, bd,
ca, cb, cd,
da, db, dc.

$P_2^4 = 4 \times 3 = 12$

$P_r^n = ?$

n letters, $\{a_r \cdots a_n\}$, $n \leq r$

r letters.

n = number of ways for completing task 1.

n-1 = number of ways for completing task 2.

$\cdots$ until task r:

$n - (r-1)$ = number of ways for completing task r.

Rule of Permutations $P_r^n = n(n-1) \times \cdots \times (n-r+1)$

Permutations of r letters from n letters. ($r \leq n$)

$P_r^n = n(n-1) \ldots (n-r+1)$

(14) Example

$P_2^4 = 4 \times 3 = 12 \rightarrow \{a, b, c, d\}$

In general for Square Permutations

$P_n^n = n(n-1) \ldots (3)(2)(1) = n!$

In general for Any Permutations

$P_r^n = \frac{n(n-1) \ldots (n-r+1)(n-r)(n-r-1) \ldots 1}{(n-r)(n-r-1) \ldots 1} = \frac{n!}{(n-r)!}$

Combinations

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

Combination # of combination of r letters from n leters.

(15) Example: Let $n = 4, r = 2$
$a, b, c, d \rightarrow ab, ac, ad, bc, bd, cd$
$\binom{4}{2} = 6$
$P_2^4 = 2\binom{4}{2}$

(16) Example
$\binom{4}{3} = ?$
$S = \{a, b, c, d\}$
$abc, abd, acd, bcd \rightarrow \binom{4}{3} = 4$
$P_3^4 = 3!\binom{4}{3}$

General Combination Formula

$\therefore P^n_r = r!\binom{n}{r} \rightarrow \binom{n}{r} = \frac{P^n_r}{r!} = \frac{\frac{n!}{(n-r)!}}{r!} = \frac{n!}{r!(n-r)!}$

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

$P^n_r = \frac{n!}{(n-r)!}$

(17) Example: How many hands of 5 cards do we have?
$\binom{52}{5} = \frac{52!}{5!47!} = \frac{52\times51\times50\times49\times48}{120}$

(18) Example: How many diagonals do we have in a polygone with n sides?
$\binom{5}{2} - 5 = 5$
In general, the solution is:
$\binom{n}{2} - n = \frac{n(n-1)}{2} - n$
$= \frac{n(n-1)-2n}{n} = \frac{n(n-1-2)}{2} = \frac{n(n-3)}{2}$

(19) Example: A city has m by n streets in a grid. How many ways can someone traverse from one corner of the city to the other?
$\frac{(m+n)!}{m!n!} = \binom{m+n}{m} = \binom{m+n}{n}$

(20) Example: How many ways can we write out $(a+b)^5$ expanded?
$(a+b)^5 = a^5 + \frac{5!}{4!1!}a^4b + \frac{5!}{3!2!}a^3b^2 + \frac{5!}{2!3!}a^2b^3 + \frac{5!}{1!4!}ab^4 + b^5$

A General Polynomial Combinational Formula

$$(a+b)^n = \sum_{i=0}^{n}\binom{n}{i}a^ib^{n-i}$$

$(a+b+c)^n = \sum\sum\binom{n}{i,j}a^ib^jc^{n-i-j}$

Properties of Combinations

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

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

(21) Example: How many subsets do we have if the set has n elements?
$S = \{a,b,c\}$
$\{a\},\{b\},\{c\},\{a,b\},\{a,c\},\{b,c\},\{a,b,c\}$
You are effectively doing, in this example, n choose 0, n choose 1, n choose etc until n.
$\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n} = 2^n$
To explain this, we can use the multiplication rule and say, "For task i, choose the i^{th} letter or do not choose it."

Defining Probability

(22) Example: Flip 2 coins.
$S = \{HH, HT, TH, TT\}$
Define E = having at least 1 head.
$E = \{TH, HT, HH\}$
$P(E) = \frac{3}{4}$
Define F = exactly one head = $\{TH, HT\}$
$P(F) = \frac{2}{4}$

(23) Example: Flip a coin until you see the first head.
$S = \{H, TH, TTH, \dots\}$
Event E; Stop before or at 4^{th} trial = $\{H, TH, TTH, TTTH\}$
$P(E) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}$

(24) Example: A point is picked at random from a circle with radius r
$S = \{(x, y) : x^2 + y^2 \leq R^2\}$
Let $E = \{(x, y) : x^2 + y^2 \leq \frac{R^2}{4}\}$
$\frac{R^2}{4} = (\frac{R}{2})^2$
$P(E) = \frac{Area(E)}{Area(S)} = \frac{\pi \times (\frac{R}{2})^2}{\pi \times R^2} = \frac{1}{4}$

Probability

Let S be the sample space of a random experiment. Let P be a set function with the following probabilities:

Assumptions:

i) $P(E) \geq 0$

ii) If E_1, E_2, … are disjoint ($E_i \cap E_j = 0; i \neq j$) then $P(E_1 \cup E_2 \dots) = \sum_i P(E_i)$

iii) $P(S) = 1$

Properties of P

1. $P(\emptyset) = 0$

2. $P(A - B) = P(A) - P(A \cap B)$
Note: $A - B = \{x | x \in A, x \notin B\}$
$P((A - B) \cup (A \cap B)) = P(A)$
$P(A - B) + P(A \cap B) = P(A)$
$P(A - B) = P(A) - P(A \cap B)$

3. If $A \subset B \rightarrow P(A) \leq P(B)$
$B = A \cup (B - A)$
$P(B) = P(A) + P(B - A) \geq 0$
$\therefore P(A) \leq P(B)$

4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
   Proof:
   $A \cup B = A \cup (B - A)$
   $P(A \cup B) = P(A) + P(B - A) = P(A) + P(B) - P(A \cap B)$
5. $0 \leq P(E) \leq 1$
   If $\emptyset \subset E \subset S$
   $0 = P(\emptyset) \leq P(E) \leq P(S) = 1$
   $P(E') = 1 - P(E)$

(25) Example: Union of 3 sets
Let $B \cup C = D$
$P(A \cup B \cup C) = P(A \cup D) = P(A) + P(D) - P(A \cap D)$
$A \cap D = A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$P(A \cap D) = P(A \cap B) + P(A \cap C) - P(A \cap B \cap C)$
$\therefore P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)$

(26) Example: Union of 4 sets
$P(A \cup B \cup C \cup D)$
$= P(A) + P(B) + P(C) + P(D) - P(A \cap B) - P(A \cap C) - P(B \cap C) - P(C \cap D) - P(A \cap D) - P(B \cap D) + P(A \cap B \cap C) + P(A \cap B \cap D) + P(A \cap C \cap D) + P(B \cap C \cap D) - P(A \cap B \cap C \cap D)$

(27) Example: 100 different letters are sent to 100 different addresses at random. What is the probability that *at least* 1 letter reaches its designate address?
Hint: In questions, you can look for key phrases: *at least* suggests union, and *all* suggests intersection.
$E_i = i^{th}$ letter goes to the right address.
$P(E_1 \cup E_2 \cup E_3 \cup \ldots \cup E_100) = P(E_1) + P(E_2) + \ldots + P(E_100) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - \ldots - P(E_9 9 \cap E_100) + P(E_1 \cap E_2 \cap E_3) + \ldots\ldots - P(E_1 \cap \ldots \cap E_100)$
$P(E_i) = \frac{1}{100}$
$i \neq j,\ P(E_i \cap E_j) = \frac{1}{100} \times \frac{1}{99}$
$P(E_i \cap E_j \cap E_k) = \frac{1}{100} \times \frac{1}{99} \times \frac{1}{98}$
$\rightarrow 100 * (\frac{1}{100}) - \binom{100}{2} * \frac{1}{100*99} + \binom{100}{3} * \frac{1}{100*99*98} - \ldots$
$= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots - \frac{1}{100!}$
$= 1 - e^{-1}$ is the probability that 1 letter will reach the correct address.

Conditional Probability

Let S be the sample space for a random experiment.

Let $B \subset S$ (B is an event)

Such that $P(B) > 0$

For any event A we define $P(A|B) = \frac{P(A \cap B)}{P(B)}$

(28) Example: A die is rolled. Let $A =$ result is more than 4, let $B =$ outcome is even.
Find $P(A|B)$
Solution:

- $A = \{5, 6\}$
- $B = \{2, 4, 6\}$

Given that you have a result in B, what are the chances that it is a result in A? Since B has 2, 4, and 6, one third of the time, A will be met as well.

$A \cap B = \{6\}$

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

$P(B) = \frac{3}{6}$

$P(A|B) = \frac{\frac{1}{6}}{\frac{3}{6}}$

$= \frac{1}{3}$

Total Probability Rule

Total Probability Rule Let S be a sample space in a random experiment. Let A be any event ($A \subset S$)

Also we assume $S = E_1 \cup E_2 \cup \ldots \cup E_k$, such that E_i's are disjoint.

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

Proof $A \subset S$

$A \cap S = A$

$A = A \cap (E_1 \cup E_2 \cup \ldots \cup E_k)$

$= (A \cap E_1) \cup (A \cap E_2) \cup \ldots \cup (A \cap E_k)$

$\therefore P(A) = P(A \cap E_1) + P(A \cap E_2) + \ldots + P(A \cap E_k)$

$= \frac{P(A \cap E_1)}{P(E_1)} P(E_1) + \ldots + \frac{P(A \cap E_k)}{P(E_k)} P(E_k)$

$= P(A|E_1)P(E_1) + \ldots + P(A|E_k)P(E_k)$

(29) Example (Polya): There are m white chips and n black chips in the box. You remove one chip from the box without looking at it, and it is put inside. A second chip is then drawn. What is the probability that the second chip is white?

Let $P(A)$ represent the probability that the 2^{nd} chip is white.

Solution:

- $\text{E}_1 = 1^{\text{st}}$ chip is white
- $\text{E}_2 = 1^{\text{st}}$ chip is black.

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

$= \frac{m-1}{m+n-1} \times \frac{m}{m+n} + \frac{m}{m+n-1} \times \frac{n}{m+n}$

$= \dfrac{m}{m+n}$

Slightly more explanation: Probability doesn't measure the physical characteristics of the box, it measures based on information that you have. Not knowing the first chip, the second chip, or the k chip, will not change the chance of the k+1's chip.

(30) Example: You have two boxes, both with m white chips and n black chips, and you take one random chip from the first box and place it in the second box. Now you take a random chip from the second box. What is the probability that this chip is white?

$E_1 = 1^{\text{st}}$ chip is white

$E_2 = 2^{\text{nd}}$ chip is black

$P($ 2^{nd} chip is white$) = P(A|E_1)P(E_1) + P(A|E_2)P(E_2)$

$= \frac{m+1}{m+n+1} \times \frac{m}{m+n} + \frac{m}{m+n+1} \times \frac{n}{m+n}$

$$= \frac{m}{m+n}$$

The total probability rule is useful for tangled experiments, it can help untangle them and find a simple solution.

(31) Example: You roll a die. You now flip a coin the number of times given by the die. What is the probability that no heads were flipped?

$P(\text{no heads}) = ?$

- E_1 = Die is 1
- E_2 = Die is 2
- E_3 = Die is 3
- E_4 = Die is 4
- E_5 = Die is 5
- E_6 = Die is 6

A = no heads

$P(A) = P(A|E_1)P(E_1) + \ldots + P(A|E_6)P(E_6)$

$P(A|E_1) = \frac{1}{2}$

$P(A|E_2) = \frac{1}{4}$

$P(A|E_3) = \frac{1}{8}$

$P(A|E_4) = \frac{1}{16}$

$P(A|E_5) = \frac{1}{32}$

$P(A|E_6) = \frac{1}{64}$

$P(E_i) = \frac{1}{6}$ where $1 \leq i \leq 6$

$P(A) = \frac{1}{2} \times \frac{1}{6} + \ldots + \frac{1}{64} \times \frac{1}{6}$

Baye's Rule

Baye's Rule Let S be the sample space and $S = E_1 \cup \ldots \cup E_k; E_i$'s disjoint and $A \subset C$

$$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{\sum_k P(A|E_k)P(E_k)}$$

Proof:

$P(E_1|A) = \frac{(P(A \cap E_1)/P(E_1))P(E_1)}{P(A)}$

$$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{\sum_k P(A|E_k)P(E_k)}$$

(32) In a transmittion system, you send either 0 or 1. Let $P(\{0\}) = P(\{1\}) = 0.5$

Trans	Receive
0	0
1	1

$P[Rec0|Send0] = 0.99$

$P[Rec1|Send1] = 0.95$

$P[1 \text{ was sent } |Rec1]$

$E_1 = \text{send } 1$

$E_2 = \text{send } 0$

$A = \text{Receive } 1$

Applying the base rule:

$P(E_1|A) = \dfrac{P(A|E_1)P(E_1)}{P(A|E_1)P(E_1) + P(A|E_2)P(E_2)}$

$= \frac{.95\times.5}{.95\times.5+.01\times.5}$

$\therefore = \frac{95}{96}$

(33) Example: 0.01 of people in a population have a certain disease.

$P(+|sick) = 0.98$

$P(-|healthy) = 0.99$

$P(sick|+) =?$

$E_1 = sick$

$E_2 = healthy$

$A =$test is +

$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{P(A|E_1)P(E_1)+P(A|E_2)P(E_2)}$

$\frac{.98\times.01}{.98\times.01+.01\times.99} - \frac{.98}{.98+.99}$

Practice Midterm Questions

(34) There are 5 components connected on a circuit. To get from component 1 to 5, you may take paths through components 2-3 or just component 4.

The components have a certain percentage of chance that they will work:

- 1: 0.5
- 2: 0.8
- 3: 0.7
- 4: 0.5
- 5: 0.5

What is the chance that the entire system functions?

Ai = i$^{\text{th}}$ component works, i=1,2,3,4,5

Independence Two events A and B are independent if

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

$P(A_1 \cap ((A_2 \cap A_3) \cup A_4) \cap A_5)$
$= P(A_1)P(A_5)P((A_2 \cap A_3) \cup A_4)$
$= .5 \times .5(P(A_2 \cap A_3) + P(A_4) - P(A_2 \cap A_3 \cap A_4))$
$= .5 \times .5((0.8)(0.7) + 0.5 - .8 \times .7 \times .5)$
$= 0.195$

To get the probability of none of them working, you'd take the complement of P(none of them work).

(35) Here is a table showing the income of 1000 families.

Income table	Husband (<=30,000)	Husband (> 30,000)	Wife total
Wife (<= $30,000)	425	400	825
Wife (> $30,000)	65	110	175
Husband Total:	490	510	1000 families

What is the probability that a husband makes >30,000 if his wife makes >30,000?

$P(husband > 30,000|wife > 30,000)$
$= \frac{P(husband > 30k \cap wife > 30k)}{P(wife > 30k)}$
$= \frac{\frac{110}{1000}}{\frac{175}{1000}}$
$= \frac{110}{175}$

What is the probability that a husband makes more than 30k?

$P(husband > 30k) = \frac{510}{1000}$

(36) The probability of pipe welds being defective is 0.02. The probability of them being good is 0.98. A device for checking the weld defects has a probability of working on defective welds of 0.90. The device also sends a bad signal on a good weld 5% of the time.

- $P(Defective) = 0.02$
- $P(Good) = 0.98$
- $P(signal\ is\ sent|Defective) = 0.90$
- $P(signal\ is\ sent|Good) = 0.05$
- $P(Defective|Signal\ is\ received) =?$

This is using Baye's Rule.

- $E_1 = Defective$
- $E_2 = Good$
- $A = signalisreceived$
- $P(E_1|A)?$

Solution

$P(E_1|A) = \frac{P(A|E_1)P(E_1)}{P(A|E_1)P(E_1)+P(A|E_2)P(E_2)}$
$= \frac{.9\times.02}{0.9\times0.02+0.05\times0.98}$
$= 0.27$

(37) Let A and B two events.

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

Is A independent from B?

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

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

$P(A \cap B) = 0.15$

$P(A)P(B) = 0.15$

Since $P(A \cap B) = P(A)P(B)$, A and B are independent

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

$= 0.7 + 0.5 - (0.7)(0.5)$

$= 0.85$

DeMorgan's Law

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

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

$P(A' \cup B') = P[(A \cap B)'] = 1 - P(A \cap B) = 1 - 0.15 = 0.85$

Discrete Random Variables

Discrete Random Variables In a random experiment with a sample space S, we map members of S to real numbers.

$S \to R$

w = outcome in S

(38) Roll two dice. What is the probability of their added outcomes?

$S = \{(x, y) : x = 1, 2, \ldots, 6y = 1, 2, \ldots, 6\}$

$\#(S) = 36$

$X(x, y) = x + y$

$X \in \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

x	2	3	4	5	6	7	8	9	10	11	12
$\frac{P(X=x)}{36}$	1	2	3	4	5	6	5	4	3	2	1

(39) In a box containing 10 items there are two defective items.

#	Condition
8	Good
2	Defective

Draw two items at random from this box without replacement.

Let $x = \#$ of defective items in the sample of two items.

Therefore, $X \in \{0, 1, 2\}$

Find $P(x = 0), P(x = 1), P(x = 2)$

Solution:

$$P(x = 0) = \frac{\binom{8}{2}}{\binom{10}{2}} = \frac{28}{45}$$

$$P(x = 1) = \frac{\binom{8}{1} \times \binom{2}{1}}{\binom{10}{2}} = \frac{16}{45}$$

$$P(x = 2) = \frac{\binom{2}{2}}{\binom{10}{2}} = \frac{1}{45}$$

x	0	1	2
$f(x) = P(X = x)$	$\frac{28}{45}$	$\frac{16}{45}$	$\frac{1}{45}$

Note that all the probabilities add up to 1. $f(x)$ is called the probability mass function for the random variable X.

Expected Value (mean)

Expected Value (mean): $E(X) = \sum_{\imath} x\imath f(x\imath) = \mu$

$\mu = 0 * \frac{28}{45} + 1 * \frac{16}{45} + 2 * \frac{1}{45} = \frac{18}{45} = \frac{6}{15} = \frac{2}{5} = 0.4$

Variance and Standard Deviation

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

$\sqrt{E(X - \mu)^2} = \sigma = S.D$

So using the numbers from example (39):

$\sigma^2 = E(X - \mu)^2 = \sum_{x} (x - \mu)^2 f(x)$

$= (0 - 0.4)^2(\frac{28}{45}) + (1 - 0.4)^2(\frac{16}{45}) + (2 - 0.4)^2(\frac{1}{45})$

(40) Example: Mean and Variance

x	$f(x)$
0	0.18
1	0.34
2	0.33
4	0.15

$E(X) = \mu = 0(0.18) + 1(0.34) + 2(0.33) + 4(0.15)$

$= 0.34 + 0.66 + 0.6 = 1.6$

$\sigma^2 = (0-1.45)^2(0.18) + (1-1.6)^2(0.34) + (2-1.6)^2(0.33) + (4-1.6)^2(0.15)$

Properties of Probability Mass Function

i) $f(x) \geq 0$

ii) $\sum_x f(x) = 1$

(41) Let X be a discrete random variable with the probability mass function $f(x)$ given below.

x	0	1	2
$f(x)$	C	2C	3C

Find $E(X)$, σ^2, $P(X \geq \mu + \sigma)$

Solution:

$C + 2C + 3C = 1$

$C = \frac{1}{6}$

x	0	1	2
$f(x)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$

$E(X) = 0(\frac{1}{6}) + 1(\frac{2}{6}) + 2(\frac{3}{6} = \frac{2}{6} + \frac{6}{6} = \frac{8}{6} = \frac{4}{3}$

$\sigma^2 = (0-\frac{4}{3})^2 + (1-\frac{4}{3})^2 + (2-\frac{4}{3})^2(\frac{3}{6})$

$\frac{16}{9} \times \frac{1}{6} + (\frac{1}{9})(\frac{2}{6}) + (\frac{4}{9})(\frac{3}{6}) = \frac{16}{54} + \frac{2}{54} + \frac{12}{54} = \frac{30}{54} = \frac{5}{9}$

$\sigma = \frac{\sqrt{5}}{3}$

$P(x \geq \mu + \sigma) = P(x \geq \frac{4}{3} + \frac{\sqrt{5}}{3}) = 0$

Properties for mu (μ) and sigma squared (σ^2)

1) $\sigma^2 = E(x^2) - (E(x))^2$

Proof:

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

$= \sum_x (x^2 + \mu^2 - 2\mu x) f(x)$

$= \sum_x x^2 f(x) + \sum_x \mu^2 f(x) - \sum_x 2\mu x f(x)$

$= E(x^2) + \mu^2 \sum_x f(x) - 2\mu \sum_x x f(x)$

$= E(x^2) + \mu^2 - 2\mu * \mu$

$= E(x^2) - \mu^2$

(42) Flip a coin 3 times

Let x=# of heads

Find $E(x)$, σ^2, σ

Solution:

$\{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$

x	0	1	2	3
$P(x = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$E(X) = \mu$
$= 0\frac{1}{8} + 1\frac{3}{8} + 2\frac{3}{8} + 3\frac{1}{8}$
$= 1.5$
$E(x^2) = 0^2\frac{1}{8} + 1^2\frac{3}{8} + 2^2\frac{3}{8} + 3^2\frac{1}{8} = \frac{3}{8} + \frac{12}{8} + \frac{9}{8} = \frac{24}{8} = 3$
$\sigma^2 = E(x^2) - \mu^2 = 3 - 1.5^2 = 3 - 2.25 = 0.75$
$\sigma = \sqrt{0.75}$

Cumulative Distribution Function

Cumulative Distribution Function Let x be a random variable with probability mass function f(x),

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

Note, given F you can find f. There is a one-to-one relationship. $F \leftrightarrow f$

(43) In example (42), find the cumulative distribution function.

Solution:

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

$P(X \leq x)$	x
0	$x < 0$
$\frac{1}{8}$	$0 \leq x < 1$
$\frac{4}{8}$	$1 \leq x < 2$
$\frac{7}{8}$	$2 \leq x < 3$
$\frac{8}{8}$	$3 \leq x$

(44) Find the relevant probabilities.

$$F_x(x) = \begin{cases} 0 \text{ if } x < 0 \\ 0.17 \text{ if } x \leq x < 1 \\ 0.4 \text{ if } 1 \leq < 2 \\ 0.59 \text{ if } 2 \leq x < 3 \\ 0.72 \text{ if } 3 \leq x < 4 \\ 0.8 \text{ if } 4 \leq x < 5 \\ 1 \text{ if } x \geq 5 \end{cases}$$

$F(3) = P(x \leq 3) = 0.72$

$$\begin{aligned} P(2 < x \leq 3) &= F(3) - F(2) \\ &= 0.72 - 0.59 \\ &= 0.13 \end{aligned}$$

$P(0 < x < 0.75) =$ total amount of jumps in$F = 0$

Solution: Write PMF:

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

x	0	1	2	3	4	5
$f(x)$	0.17	0.23	0.19	0.13	0.08	0.2

$E(x) = \sum_x xf(x) = 0 \times 0.17 + 1 \times 0.23 + 2 \times 0.19 + 3 \times 0.13 + 4 \times 0.08 + 5 \times 0.2$

$= 2.23$

Okay, what about σ^2?

$$\begin{aligned} E(x^2) &= \sum_x x^2 f(x) \\ &= 0^2 \times 0.17 + 1^2 \times 0.23 + 2^2 \times 0.19 + 3^2 \times 0.13 + 4^2 \times 0.08 + 5^2 \times 0.2 \\ &= 8.44 \end{aligned}$$

Uniform Distance on Finite Sets

Uniform Distance on Finite Sets If x is a discrete random variable uniformly distributed on $\{1, 2, \ldots k\}$, a die with k sides is rolled.

$P(X = x) = \begin{cases} \frac{1}{k} & x = 1, 2, \ldots, k \\ 0 & \text{elsewhere} \end{cases}$

What is the average of x?

$E(x) = \frac{k+1}{2}$ which is equivalent to $E(x) = \sum_x xf(x)$

What is the variance of x?

$$\begin{aligned} \sigma^2 &= E(x^2) - (E(x))^2 \\ &= E(x^2) - (\tfrac{k+1}{2})^2 \end{aligned}$$

$E(x^2) = \frac{1^2+2^2+\cdots+k^2}{k}$

$\sum_{i=1}^{n}[(i+1)^3 - i^3] = (2^3 - 1^3) + (3^3 - 2^3) + \cdots + ((n+1)^3 - n^3)$

$\sum_{i=1}^{n}[(i+1)^3 - i^3] = (n+1)^3 - 1$

$\sum_{i=1}^{n}[i^3 + 3i^2 + 3i + 1 - i^3] = 3\sum_{i=1}^{n} i^2 + 3\sum_{i=1}^{n} i$

$3\sum_{i=1}^{n} i^2 + 3\frac{n(n+1)}{2} + n$

$\sum_{i=1}^{n} = \frac{(n+1)^3 - 1 - \frac{3n(n+1)}{2} - n}{3} = \frac{n(n+1)(2n+1)}{6}$

$E(x^2) = \frac{1^2+\cdots+k^2}{k} = \frac{\frac{k(k+1)(2k+1)}{6}}{k} = \frac{(k+1)(2k+1)}{6}$

$\sigma^2 = \frac{(k+1)(2k+1)}{6} - \frac{(k+1)^2}{4} = \frac{k^2-1}{12} = \sigma^2$

Bernoulli's Trials

Bernoulli's Trials Let $S = \{0, 1\}$, where 0 is failure, and 1 is success.

$P(x = 1) = P, P(x = 0) = 1 - p = q$

Probability mass function for Bernoulli's random variable.

x	0	1
$f(x)$	1-p	p

$$f(x) = \begin{cases} p^x(1-p)1-x & x = 0, 1 \\ 0 & \text{elsewhere} \end{cases}$$

We have $E(x) = 0(1 - p) + 1 \times p = p$

$E(x^2) = 0^2 \times (1 - p) + 1^2 \times p = p$

$\sigma^2 = p - p^2 = p(1 - p)$

Binomial Distribution

Binomial Distribution Repeat Bernoulli's experiment independently n times.

Y = # of successes in n trials.

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

$E(Y) = E(x_1 + \cdots + x_n) = E(x_1) + \cdots + E(x_n) = np$

Probability mass function for Y:

P(y=k) = ?

k = 0,1,2,. . . ,n

What is the chance that you experience k successes?

That means that you have k successes, and n-k successes.

$$\begin{aligned} P(Y = k) &= \tfrac{n!}{k!(n-k)!} \\ &= \tbinom{n}{k} p^k (1-p)^{n-k} k = 0, 1, 2, \ldots, n \end{aligned}$$

(45) In a multiple choice test, each question has 5 possible answers, but only one is correct. If the test has 10 questions, what is the probability that a student who answers questions at random scores:

For this example, you may use the table at pages 739-741 in the textbook.

i) %80 or more?

n=10 (10 questions)

p=0.2 (chance of getting it right at random)

q=0.8 (chance of getting it wrong at random)

$P(Y \geq 8) = P(Y = 8) + P(Y = 9) + P(Y = 10)$

$= \binom{10}{8} 0.2^8 0.8^2 + \binom{10}{9} 0.2^9 0.8^1 + \binom{10}{10} 0.2^1 0$

$= 1 - P(Y \leq 7)$

$1 - 0.999 = 0.001$

ii) %50 or more?
$P(Y \geq 5) = 1 - P(Y \leq 4) = 1 - 0.9672 = 0.0328$

iii) Exactly %10?
$= \binom{10}{1} 0.2^1 0.8^9$

Result If X and Y are independent,

1) $Var(x \pm y) = Var(x) + Var(y)$
2) $Var(cX) = c^2 Var(x)$
$E[cX)^2] - (E(cX))^2 = c^2E(x^2) - c^2(E(x))^2 = c^2\sigma^2$

Geometric Distribution

Geometric Distribution In a sequence of Bernoulli's trials, let X = # of trials we need to have to see the first success.
$P(\{S\}) = p, X \in \{1, 2, 3, \dots\}$ $P(\{F\}) = q = 1 - p, P(X = k) = P(FF \cdots FS) = pq^{k-1}$

k	1	2	3	4	...
$P(X = k)$	p	pq	pqˆ2	pqˆ3	...

Notice that if $S = 1 + q + q^2 + \cdots$
$qS = q + q^2 + q^3 + \cdots$
$S - qS = 1$
$S(1 - q) = 1$
$S = \frac{1}{1-q} = \frac{1}{p}$

(46) The probability of manufacturing a defective item is p = 0.05. Find:

i) The probability that the fifth manufactured item is the first defective item.
Solution:
- p = 0.05
- q = 0.95 $\therefore P(X = 5) = 0.95^4 \times 0.05$

ii) The probability that the first defective item appears at or after 10^{th} manufactured item.
Solution:
$P(X \geq 10) = 1 - P(X \leq 9) = pq^9 + pq^10 + pq^11 + \cdots$
$= pq^9(1 + q + q^2 + \cdots)$
$= \frac{pq^9}{p} = q^9 = 0.95^9$

iii) Given that the first ten produced items are not defective, what is the probability that the fifteenth manufactured item is defective?
Solution:
$P(x = 15 | x \geq 11) = \frac{P(A \cap B)}{P(B)} = \frac{P(x=15)}{P(x \geq 11)} = \frac{pq^14}{q^10} = pq^4 = (0.05)(0.95)^4$
$\{15\} = \{15\} \cap \{11, 12, 13, \dots\}$
$\therefore It'sthesameasthesolutiontoi), becausethetestsareindependent.$

Finding Expected Value and Variance for Geometric Distance

We know that $P(X=k) = pq^{k-1} \quad k = 1, 2, \ldots$

$E(x) = \sum_{k=1}^{\infty} kpq^{k-1} = P[1 + 2q + 3q^2 + 4q^3 + \cdots]$

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

$E[x(x-1)] = \sum_{k=1}^{\infty} k(k-1)pq^{k-1} = \sum_{k=2}^{\infty} k(k-1)pq^{k-1}$

$= pq\sum_{k=2}^{\infty} k(k-1)q^{k-1} = pq[2 \times 1 + 3 \times 2q + 4 \times 3 \times q^2 + \cdots]$

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

$= pq\frac{2(1-q)}{(1-q)^4} = 2pq\frac{1}{(1-q)^3} = \frac{2pq}{p^3} = \frac{2q}{p^2}$

$E(x) = \frac{1}{p}$

$E(x^2) - E(x) = \frac{2q}{p^2}$

$E(x^2) = \frac{2q}{p^2} + \frac{1}{p}$

$E(x^2) - (E(x))^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q+q+p-1}{p^2} = \frac{q}{p^2}$

Bertrand Paradox

Is this even a paradox?

Take a game, where you flip a coin and you get the number of dollars back, doubling for each successful tails, according to this table:

H	TH	TTH	TTTH	TTTTH	…
1	2	4	8	16	…

What is the expected pay for the game?

$\frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 4 + \frac{1}{16} \times 8 + \cdots = E(prize)$

$= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots = \infty$

Idea: Let's say there's a millionaire who won't pay more than 2^{30} rolls.

H	TH	TTH	TTTH	TTTTH	⋯	$T^{30}H$	$T^{30}H$	⋯
1	2	4	2^3	2^4	⋯	2^{30}	2^{30}	⋯

$E(prize) = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 4 \times \frac{1}{8} + \cdots + 2^30 \times \frac{1}{2^31} + 2^30(\frac{1}{2^32} + \frac{1}{33} + \cdots)$

$= 15 + 2^30 \times \frac{1}{2^32}(1 + \frac{1}{2} + \frac{1}{4} + \cdots)$

$15 + \frac{1}{4}(\frac{1}{1-\frac{1}{2}})$

$= 15.5$

Negative Binomial Distribution

In Bernoulli's trials, let x = # of trials to see the r$^{\text{th}}$ success.

k = r, r+1, r+2, ...

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

$= \binom{k-1}{r-1} p^r q^{k-1}$, where $k = r, r+1, \ldots$

$Y_1 = \#$ of trials to see the first success.

$Y_2 = \#$ of trials to see the second success after you see the first success.

...

Y_r

$x = Y_1 + Y_2 + \cdots + Y_r$

$E(x) = \frac{r}{p}$, $var(x) = \frac{r(q)}{p^2}$

Poisson Distribution

In Bernoulli's trials with n experiements P=P({S}), let $np = \lambda$ we can write:

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} (\lambda/n)^k (1 - (\lambda/n))^{n-k}$

There is a break here because I was late for class.

$P(X = k) \to \frac{e^{-\lambda}\lambda^k}{k!} = g(k)$

$k = 0, 1, 2, \ldots$

(47) Is $g(k) = \frac{e^{-\lambda}\lambda^k}{k!}$ a proper probability mass function?
$k = 0, 1, 2, \ldots$

(a) Notice that $e^{-\lambda}\frac{\lambda^k}{k!} \geq 0$.

(b) $$1 = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}(\sum_{k=0}^{\infty}\frac{\lambda^k}{k!})$$
$$= e^{-\lambda}e^{\lambda} = e^0 = 1$$

Mean and Variance for Poisson Distribution

Mean: $E(x) = \mu = \sum_{k=0}^{\infty} k\frac{e^{\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty}\frac{e^{\lambda}\lambda^k}{(k-1)!}$

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

Variance for a Poisson Distribution

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

$E(x(x-1)) = \lambda^2 = E(x^2) - E(x) \rightarrow E(x^2) = \lambda^2 + \lambda$

$\sigma^2 = Var(x) = E(x^2) - (E(x))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$

Traffic

Between time 0 and time t, you want to know how many customers will arrive.

N(t) = #customers arrive in [o, t].

$P[N(t) = k] =?$

We are going to divide the time into sub-intervals that are very small, so that P becomes smaller.

Conversely, because there are so many sub intervals, n is very large.

$P[N(t) = k] = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$

(48) A 911 agent receives 5 calls/hour.

i) Calculate the probability that she receives 2 calls or more in 30 minutes.
In 30 minutes, $\lambda = 2.5$ (average # calls per 30 minutes)
$P(x \geq 2) = 1 - P(x=0) - P(x=1) = 1 = \frac{e^{-\lambda}\lambda^0}{0!} - \frac{e^{-\lambda}\lambda^1}{1!}$

ii) Find the probability that the agent is occupied more than average in the next 3 hours.
$\lambda = 15$

$$P(x > 15) == 1 - P(x \leq 15) = 1 - \sum_{k=0}^{1} 5e^{-}15\frac{15^k}{k!}$$

Hypergeometric Distribution

(49) From a regular deck of cards (52 cards), 5 cards are drawn at random.

i) Find the probability that you have 3 aces.
How many ways can you draw 5 cards from 52 cards?
$\frac{\binom{4}{3}\binom{2}{48}}{\binom{52}{5}}$

ii) Find the probability that all of them are the same suit.
$\frac{\binom{13}{5} \times 4}{\binom{52}{5}}$

iii) Find the probability that you have 4 Kings.
$\frac{\binom{4}{4} \times \binom{48}{1}}{\binom{52}{5}}$

(50) You have a box with m white chips and n black chips. You pick out k. What is the probability that X = x?

$x = \#$ white chips

$P(X = x) = \frac{\binom{M}{x}\binom{N}{k-x}}{\binom{m+n}{k}}$

In the same model, if we replace each chip after noticing the color...

$P(X = x)\binom{k}{x}\frac{m}{m+n}$

Continuous Random Variables

Continuous Random Variables If a random experiment takes many values in an interval then we need to define continuous random variable.

Examples are:

1. Height of a randomly selected student.
2. Age of a person who is diagnosed with skin cancer.
3. The temperature of a certain futures date.

(51) A point is picked at random from a circle of radius R. Let D = the distance of the selected point to the center.

$S = (x, y) : x^2 + y^2 \leq R^2$

$D = \sqrt{x^2 + y^2}, D \in [0, R]$

$P[D \leq x] = F(x)$

F is the cumulative distribution function for random variable D.

$$P[D \leq x] = P[(x, y) : x^2 + y^2 \leq x^2] = \frac{\pi x^2}{\pi R^2} = F(x)$$

$$F(x) = \begin{cases} 0 \text{ if } x < 0 \\ \frac{x^2}{R^2} \text{ if } 0 \leq x \leq R \\ 1 \text{ if } x \geq R \end{cases}$$

Continuous Type A random variable x is of continuous type if its cumulative distribution function, $F(x) = P[X \leq x]$, is continuous everywhere.

$\lim_{x \to a} F(x) = F(a)$

$$F(x) = \begin{cases} 2 \text{ if } x \neq 0 \\ 0 \text{ if } x = 0 \end{cases}$$

$\lim_{x \to 0^-} F(x) = 2$

$\lim_{x \to 0^+} F(x) = 2$

$F(0) = 0$

Definition If x is a continuous random variable with a differentiable cumulative distribution function F, then f(x) = F'(x) is called the probability density function for the random variable X.

Properties:

i) $f(x) \geq 0$ because F is nondecreasing.

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

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

$F(+\infty) = P(X \leq \infty) = 1$

$F(-\infty) = P(x \leq -\infty) = 0$

$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$

$P(-\infty < x < +\infty) = F(\infty) - F(-\infty) = 1$

$\int_{-\infty}^{\infty} F'(x)dx = \int_{-\infty}^{\infty} F(x)dx = 1$

(52) Let X be a random variable with the probability density function:

$$f(x) = \begin{cases} kx^2 \text{ if } 0 < x < 1 \\ 0 \text{ if } elsewhere \end{cases}$$

i) Find k

Since $1 = \int_{-\infty}^{\infty} f(x)dx = \int_0^1 kx^2 dx = 1$

`$k \frac{x^3}{3} |_0^1 = k(1/3 - 0) = k/3 = 1 \rightarrow k=3$`

ii) Find $P(1/2 < x < 2/3)$
$= \int_{1/2}^{2/3} 3x^2 dx = x^3|_{1/2}^{2/3} = (\frac{2}{3})$
Professor was moving too fast here, and therefore ii) is incomplete

iii) Find $E(x)$ and σ^2
$E(x-\mu)^2 = E(x^2) - \mu^2$
$E(x^2) = \int_0^1 x^2(3x^2)dx = \frac{3}{5}$
$\sigma^2 = \frac{3}{5} - \frac{9}{16} = \frac{48-45}{80} = \frac{3}{80}$

iv) Find cumulative distribution function.

$$F(x) = P(X \leq x) = \begin{cases} 0 \text{ if } x < 0 \\ \int_0^x 3t^2 dt \text{ if } 0 < x < 1 \\ 1 \text{ if } x \geq 1 \end{cases}$$

$$F(x) = \begin{cases} 0 \text{ if } x < 0 \\ x^3 \text{ if } 0 \leq x \leq 1 \\ 1 \text{ if } x \geq 1 \end{cases}$$

$\int_0^x 3t^2 dt = t^3|_0^x = x^3$

*Few notes and remarks for continuous random variables.

i) $P(X = a) = \int_a^a f(x)dx = 0$
$P(2 < X < 3) = P(2 \leq X < 3) = P(2 \leq X \leq 3)$
In continuous, the chance of getting the EXACTLY same result twice is 0. Someone will not die at EXACTLY 71, because a millisecond before or after is also possible. Therefore you can drop or add "or equal to" on the ranges.

ii) $P(a \leq X \leq b) = F(b) - F(a)$

iii) f(x) (probability density functions) can also take values larger than one, unlike continuous distribution functions.

(53) Uniform distribution example.

If x has a constant probability density functions on [a,b], we denote:

$X\tilde{}Unif[a,b]$ (x has a uniform distribution on a to b, meaning it is constant.)

$$f(x) = \begin{cases} c & \text{if } a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$$

We can calculate c by noticing that:

$1 = \int_a^b cdx = cx|_a^b = cb - ca = 1 \rightarrow C = \frac{1}{b-a}$

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

$E(x) = \int_a^b \frac{1}{b-a} x dex = \frac{x^2}{2(b-a)}|_a^b = \frac{a+b}{2}$

$E(x^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{1}{b-a}(b^3/3 - a^3/3) = \frac{a^2+b^2+ab}{3}$

$\sigma^2 = E(x^2) - (\frac{a+b}{2})^2 = \frac{(b-a)^2}{12}$

(54) Pick a point from [a,b] at random. Call the point you picked X. With uniform density,

$P(X \le x) = \frac{x-a}{b-a} = F(x)$ where $a \le x \le b$

$F(x) = \int_a^x \frac{dt}{b-a} \longrightarrow f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{elsewhere} \end{cases}$

$E(x) = \int_a^b x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{x^2}{2(b-a)}|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}$

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

$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{x^3}{3(b-a)}|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+b^2+ab}{3}$

$\sigma^2 = \frac{a^2+b^2+ab}{3} - (\frac{a+b}{2})^2 = \frac{4a^2+4b^2+4ab-3(a+b)^2}{12} = \frac{a^2+b^2-2ab}{12}$

(55) A student leaves home between 8 and 8:30am with uniform distribution. It takes between 30 and 40 minutes, uniformly, to arrive at school.

X = The departure time ~u[8,8.5]

$f(x) = \begin{cases} \frac{1}{0.5} & \text{if } 8 \le x \le 8.5 \\ 0 & \text{elsewhere} \end{cases}$

Y = the amount of time it takes to arrive at school.

$g(y) = \begin{cases} \frac{1}{2/3-1/2} & \text{if } 1/2 < y < 2/3 \\ 0 & \text{elsewhere} \end{cases}$

$g(y) = \begin{cases} 6 & \text{if } 1/2 < y < 2/3 \\ 0 & \text{elsewhere} \end{cases}$

i) **Find the mean and variance of the arrival time. X and Y are independent.**

$$\begin{aligned} E(X+Y) = E(X)+E(Y) &= 8.25 + \frac{1/2+2/3}{2} \\ &= 8.25 + \frac{7}{12} \\ &= 8.25 + 0.583 \\ &= 8.833 \\ &\approx 8.50am \end{aligned}$$

$Var(X+Y) = Var(X) + Var(Y)$

$Var(X) = \frac{(b-a)^2}{12} = \frac{0.5^2}{12} = \frac{1}{48}$

$Var(Y) = \frac{(2/3-1/2)^2}{12} = \frac{1}{12\times 36}$

$Var(X+Y) = \frac{1}{48} + \frac{1}{12\times 36}$

ii) **If a student leaves home between 8-8:30am in 5 days in a row, what is the probability that he leaves home after 8:15am at least 3 of those days?**

P(leaving home after 8:15am)

$q = 0.5$

$n = 5$

$\int_8 .25^8 .5 \frac{1}{0.5} dx = 0.5$

T = # days he leaves after 8:15am in 5 different days.

$P(T \ge 3) = \binom{5}{3} \frac{1}{2}^3 \frac{1}{2}^2 + \binom{5}{4} \frac{1}{2}^4 \frac{1/2}{+} \binom{5}{5} \frac{1}{2}^5$

$= \frac{10}{32} + \frac{5}{32} + + \frac{1}{32} = \frac{16}{32} = \frac{1}{2}$

Gaussian (Normal) Distribution

Flip 3 coins:

x	0	1	2	3
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Flip 4 coins:

x	0	1	2	3	4
$P(X=x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Flip 5 coins:

x	0	1	2	3	4	5
$P(X=x)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Notice the numbers from pascal's triangle.

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

- f is symmetrical around zero.
- $f(z) \to 0$ as $z \to \pm\infty$
- f(0) is the largest value of f.
- $\int_{-\infty}^{\infty} f(z)dz = 1$

So, f is a probability density function. This is called the standard normal distribution.

See pages 742-743.

$P(Z \leq 2)$

$P(-2 \leq Z \leq 1.4)$

$P(-1.645 \leq Z \leq 1.645$

Expected Value and Variance of Gaussian Distribution

$E(Z) = 0$

$E(Z^2) = 1$

$Var(Z) = 1$

$\phi(z) = \int_{-\infty}^{\infty} f(u)du$

$1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$

$z = \frac{x-\mu}{\sigma}$

$dz = 1/\sigma dx$

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

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is a PDF.

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

$E(Z) = \frac{E(x-\mu)}{\sigma} = \frac{E(X)-\mu}{\sigma} = 0 \longrightarrow E(X) = \mu$

$1 = Var(Z) = E(\frac{(x-\mu)^2}{\sigma^2} = \frac{1}{\sigma^2} E(x-\mu)^2 \longrightarrow E(X-\mu)^2 = \sigma^2$

(56) Given $X\tilde{}N(\mu, \sigma^2)$

Find:

i) $P(\mu - \sigma < X < \mu + \sigma)$
$P(\mu - \sigma < X < \mu + \sigma) = P(\frac{\mu-\sigma-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{\mu+\sigma-\mu}{\sigma})$
$P(-1 < Z < 1) = \phi(1) - \phi(-1)$
$= 0.841345 - 0.158655$
$= 0.682690$

ii) $P(\mu - 2\sigma < X < \mu + 2\sigma)$
$= P(-2 < Z < 2) = \phi(2) - \phi(-2) = 0.977250 - 0.022750$
≈ 0.954

iii) $P(\mu - 3\sigma < X < \mu + 3\sigma)$
$= P(-3 < Z < 3) = \phi(3) - \phi(-3) = 0.998650 - 0.001350$
≈ 0.99

iv) $P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95$

(57) Scores of students in a test follow a normal distribution.

Find μ, σ^2 if:

$P(X > 18) = 0.05$

$P(X < 5) = 0.05$

Solution:

μ, σ^2 are the mean and variance of scores (X)

$0.05 = P(X > 18) = P(Z > \frac{18-\mu}{\sigma})$

To equate to 0.05, $\frac{\mu-18}{\sigma} = -1.645$

$0.05 = P(X < 5) = P(Z < \frac{5-\mu}{\sigma}) = 0.05$

$\frac{5-\mu}{\sigma} = -1.645$. Solve for μ and σ to get final answer.

(58) Let X be a random variable with a noraml distribution with mean 2, $\sigma = 1$

Find $P(1.8 < X < 2.5)$

Solution:

$P(\frac{1.8-2}{1} < \frac{X-\mu}{\sigma} < \frac{2.5-2}{1}) = P(-0.2 < Z < 0.5)$

$= \phi(0.5) - \phi(-0.2)$

$= 0.691462 - 0.420740$

Exponential and Gamma Distribution

Exponential and Gamma Distribution If λ is the traffic rate for a possoin random variable, in a time interval [0,t] in average there are λt customers.

$N(t)$ = number of customers in a time interval [0,t].

$P[N(t) = k] = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$ where $k = 0, 1, 2, 3, \ldots$

Define T = waiting time to see the 1st customer arrive.

$P(T > t) = P(N(t) = 0) = e^{-\lambda t}(\lambda t)^0/0! = e^{-\lambda t}$

Cumulative distribution function for T is $F(t) = P(T \leq t) = 1 - e^{-\lambda t}$

The probability density for T is $f(t) = F'(t) = \begin{cases} \lambda e^{-\lambda t} & t > 0 \\ 0 & \text{elsewhere} \end{cases}$

(59) There are, on average, 2 accidents per day in a city.

i) What is the chance that there are no accidents in the next 5 days?

$P(T > 5) = P(\text{no accident in the next 5 days})$
$= e^{-\lambda t} = e^{-2\times 5} = e^{-10}$

ii) What is the chance that the first accident occurs on day 2?

$P(2 < T < 3) = \int_2^3 \lambda e^{-\lambda t} dt = \int_2^3 2e^{-2t} dt = e^{-4} - e^{-6}$

Properties of Exponential Distribution

- $P(T > s + t | T > s)$
  $= \frac{P(T > s + t and T > s)}{P(T > s)} = \frac{P(T > s + t)}{P(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t)$
  In other words, exponential distribution is memoryless.

Mean and Variance for Exponential Distribution

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

$\Gamma(1) = 1$

Normal Approximation to Binomial Distribution to Poisson Distribution

$If X \sim Bin(n, p)$, i.e.

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

p = P(success) q = 1-p = P(Failure) k = 0,1,2,3,. . . ,n

If n is large there is no table to calculate P(X=k). (In the textbook, n $\leq$ 20.)

If p is small, $\lambda = np$, $P(X = k) = \frac{e^{-\lambda}\lambda}{k!}$ [n is large, p is small.]

If p is not small and n is large, we use central limit theorem.

Central Limit Theorem:

If X~Bin(n,p)

$P(a \leq \frac{x-np}{\sqrt{npq}} \leq b) \longrightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^Z/Z} dz$

In other words, if n is large, X~N(np, σ^2 = npq)

Guideline for when to use central limit theorem:

- $np \geq 5$
- $nq \geq 5$

(60) Let X = # heads in flips of m coins where:

i) m=3

x	0	1	2	3
$f(x)$	1/8	3/8	3/8	1/8

ii) m=4

x	0	1	2	3	4
$f(x)$	1/16	4/16	6/16	4/16	1/16

(61) Let Y = #heads in flips of an unbiased coin when n = 10. What is the chance that the number of heads is $P(3 \leq Y \leq 8)$

i) Accurately

$$\sum_{k=3}^{7} \binom{10}{k} (1/2)^k (1/2)^{10-k} = 0.5683 = P(Y \leq 7) - P(Y \leq 2)$$

ii) Approximately (normal approximation)

$\mu = np = 5$

$\sigma^2 = npq = 10 * .5 * .5 = 2.5$

$P(3 \leq Y \leq 7) = P(\frac{3-5}{\sqrt{2.5}} \leq 2 \leq \frac{7-5}{\sqrt{2.5}})$

But this needs continuity correction, by going to the mid points:

$P(2.5 \leq Y \leq 7.5)$

$= P(-1.518 \leq z \leq 0.316)$

$= P(z \leq 0.316) - P(z \leq -1.581) = 0.6240 - 0.0570 = 0.567$

(62) Find the probability that more than 30 but less than 35 of the next 50 births are boys.

Solution:

n = 5,

p = 1/2 = q,

X = # boys ~ Bin(50, 1/2)

$\mu = np = 25$

$\sigma^2 = npq = 12.5$

$P(31 \leq X \leq 34) = ?$

Continuity correction.

$P(30.5 \leq X \leq 34.5)$

$= P(\frac{30.5-25}{\sqrt{12.5}} \leq Z \leq \frac{34.5-25}{\sqrt{12.5}})$

$= P(1.55 \leq Z \leq 2.68) = 0.0569$

Normal Approximation to Poisson Distribution

Theorem: If X ~ Poisson (λ)

$P(X = k) = \frac{e^{-lambda}\lambda^k}{k!}$

k = 0, 1, 2, ...

$E(X) = \lambda$

$Var(X) = \lambda = \sigma^2$

$P(a \leq X \leq b) \to P(\frac{a-\lambda}{\sqrt{\lambda}} \leq Z \leq \frac{b-\lambda}{\sqrt{\lambda}}) if \lambda \to \infty$

(63) Life insurance claims follow a poisson distribution. In average, a life insurance company gets 200 claims a year. Find the probability that, in the next 5 years, there are between 800 and 1200 claims.

Solution: $\lambda = 1000$ claims/5 years

$P(800 \leq X \leq 1200)$

$\lambda = \mu = 1000$

$\sigma = \sqrt{1000} = \sqrt{\lambda}$

$P(799.5 \leq X \leq 1200.5)$

$P(\frac{799.5-1000}{\sqrt{1000}} \leq Z \leq \frac{1200.5-1000}{\sqrt{1000}}$

Descriptive Statistics

If $x_1, \ldots, x_n$ are samples from a population.

Sample mean $= \dfrac{\sum x_i}{n} = \bar{x}$

Sample variances $= S^2$

$= \frac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - x)^2$

$\bar{x}$ is a measure for center.

S^2 is a measure for spread (variability)

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 + \bar{x}^2 - 2\bar{x}x_i) = \frac{1}{n-1}(\sum x_i^2 + n\bar{x}^2 - 2\bar{x}\sum_{i=1}^{n} x_i)$

$= \frac{1}{n-1}(\sum x_i^2 + n\bar{x}^2 - 2n\bar{x}^2) = \frac{1}{n-1}(\sum x_i^2 - n\bar{x}^2)$

$S^2 = \frac{1}{n-1}(\sum x_i^2 - \frac{(\sum x_i)^2}{n})$

$n\bar{x}^2 = n(\frac{\sum x_i}{n}^2)$

Range = Max$(x_1 \ldots x_n)-$ Min$(x_1 \ldots x_n) = 5$

Median$(x_1 \ldots x_n) = 3$

q q Plot: We can calculate quantile of a normal distribution.

$\alpha = P(Z < q_\alpha)$

q_α is called $100x\alpha$ percentile.

You plot quantiles for the data versus quantiles of normal distribution. Noticing a linear pattern confirms normality of the data values.

Central Limit Theorem

Let $x_1, \ldots, x_n$ be an i.i.d. (Independent, identically distributed)

Let $\bar{x} = \frac{\sum x_i}{n}$

If the population average is μ, $(E(x_i) = \mu)$

$-\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$

Roughly speaking $\bar{x} \sim N(\mu, \sigma^2/n)$, n is large.

Exam Examples

(64) Let $x_1, \ldots, x_n$ a sample from a population with mean $\mu = 5$ and variance $\sigma^2 = 1.5$

Find c such that:

$P[\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq c] = 0.9$

Solution:

$P[Z \leq c] = 0.90$

$c = 1.28$

(65) Suppose $x_1, \ldots, x_n$ (iid)

$f(x) \begin{cases} 1/2e^{-\frac{x}{2}} & x > 0 \\ 0 & textelsewhere \end{cases}$

E(x) = 2

Var(x) = 4

Answer = 1-P(Z < 4/9)

(66) Let X be a r.v.

$f(x) = \begin{cases} cx(x+1) & 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$

Find E(X)

We first find c > 0

$c \int_0^1 x(x+1)dx = 1$

$c \int_0^1 (x^2 + x)dx = c(x^3/3 + x^2/2)|_0^1$

$= c(1/3 + 1/2) = c(5/6)$

$c = 6/5$

$E(X) = \int_0^1 xf(x)dx = 6/5 \int_0^1 x^2(x+1)dx = 6/5(x^4/4 + x^3/3)|_0^1 = 6/5(1/4 + 1/3)$

(67) A circle is drawn by choosing a radius from the uniform distribution on (0,1). Find the expected value of the area of the circle.

Solution R ~ U(0,1)

$$f(r) = \begin{cases} \frac{1}{1-0} = 1 & 0 < r < 1 \\ 0 & elsewhere \end{cases}$$

Area = πR^2

$E(\pi R^2) = \pi \int_0^1 r^2 dr = \frac{\pi}{3}$

(68) Find E(X)

$$f(x) = \begin{cases} c|x-1| & 0 < x < 2 \\ 0 & elsewhere \end{cases}$$

Solution

$c\int_0^2 |x-1|dx = c(\int_0^1 |x-1|dx + \int_1^2 |x-1|dx)$

$= c(\int_0^1 (1-x)dx + \int_1^2 (x-1)dx)$

$= c(((x - x^2/2)|_0^1) + (x^2/2 - x)|_1^2)$

$= c(1/2) + c((4/2 - 2) - (1/2 - 1))$

$= c/2 + c(1/2) = c = 1$

$c = 1$

What is the average of x?

$E(X) = \int_0^2 x|x-1|dx = \int_0^1 x(1-x)dx + \int_1^2 x(x-1)dx$

Properties of Normal Distribution

If x and y are two independent random variables, (a,b,c constant)

$E(ax + by + c) = a\int xf(x)dx + b\int yg(y)dy + c = aE(x) + bE(y) + c$

$Var(ax + by + c) = a^2Var(x) + b^2Var(y)$

$Var(ax + b) = E[(ax + b) - E(ax + b))^2]$

$= E[(ax + b - aE(x) - b)^2]$

$= E[a^2(x - E(x))^2]$

$= a^2E(x - E(x))^2$

$= a^2Var(x)$

(69) Let x, y be two independent random variables such that:

E(x) = E(y) = 2

Var(x) = Var(y) = 1

i) Find E(x-y)^2

Since $(x-y)^2 = x^2 + y^2 - 2xy$

$E((x-y)^2) = E(x^2) + E(y^2) - 2E(x)E(y)$

$= 10 - 2 * 4 = 2$

$= Var(x) = 1 = E(x^2) - E(E(x))^2$

$= E(x^2) - 4 \longrightarrow E(x^2) = 5$

Similarly,

$E(y^2) = 5$

ii) Find Var(2x-3y+5)
$Var(2x-3y+5) = 4Var(x) + 9Var(y) = 13$

(70) Let x_1, x_2, x_3 be three independent random variables such that:
$E(x_1) = E(x_2) = E(x_3) = \mu$
$Var(x) = Var(x_2) = Var(x_3) = \sigma^2$
Find $E(\frac{x_1+x_2+x_3}{3}$
Find $Var(\frac{x_1+x_2+x_3}{3}$
Solution
$E(\frac{x_1+x_2+x_3}{3} = \frac{1}{3}E(x_1) + \frac{1}{3}E(x_2) + \frac{1}{3}E(x_3) = \mu$
$Var(\frac{x_1+x_2+x_3}{3} = \frac{1}{3}^2\sigma^2 + \frac{1}{3}^2\sigma^2 + \frac{1}{3}^2 sigma^2 = sigma^2/3$

Property of x bar:

Let $x_1, \ldots, x_n$ iid sample $E(x_i) = \mu$, $Var(x_1) = \sigma^2$
$E(\bar{x}) = E(\frac{x_1+\cdots+x_n}{n}) = \mu$
$Var(\bar{x}) = \sigma^2/n$

(71) A random sample of size $n_1 = 16$ is selected from a population with mean 175 and standard deviation 288.

A second random sample of size $n_2 = 9$ is selected from another population with mean 80 and standard deviation 162.

Let $\bar{x_1}$ and $\bar{x_2}$ be the sample average of these two samples. Find $Var(\bar{x_1} + \bar{x_2})$ and $E(\bar{x_1} + \bar{x_2}) = E(\bar{x_1}) + E(\bar{x_2}) = 175 + 80 = 255$
$Var(\bar{x_1} + \bar{x_2}) = Var(\bar{x_1}) + Var(\bar{x_2}) = \frac{288}{16} + \frac{162}{9}$

Properties of Normal Distribution

If x and y are two independent random variables such that
x~$N(\mu_1, \sigma_1^2)$
y~$N(\mu_2, \sigma_2^2)$
$ax + by + c \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$
y~$N(\mu_2, \sigma_2^2)$

(72) Practice Exam Question: Suppose that $x_1, \ldots, x_n = 81$ iid.

$$f_x(x) = \begin{cases} 1/2e^{-x/2} & x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

$P[\sum_{i=1}^{8} 1x_i > 170]$
Solution
We can use the central limit theorem.
$\sum_{i=1}^{n} x_i \sim N(n\mu, n\sigma^2)$
$\mu = E(x_1) = \frac{1}{1/2} = 2$

$\sigma^2 = Var(x) = \frac{1}{(1/2)^2} = 4$

Special case of Erlang distribution

$r = 1$

$\lambda = 1/2$

$E(x) = \frac{r}{\lambda}$

$Var(x) = \frac{r}{\lambda^2}$

(Page 141 textbook)

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

(73) The effective life of a certain manufactured product is a random variable with mean 5000 hrs. A new company manufactures a similar component, but claims that the mean life is increased to 5050 hours and decreases the standard deviation to 30 hours. A random sample of size m = 16 and n = 25 are selected from these companies, respectively. The data follows a normal distribution.

$P(\bar{y} - \bar{x} > 25)$

$\bar{y} \sim 5050, \frac{30^2}{25}$

$\bar{x} \sim 5000, \frac{40^2}{16}$

$\bar{y} - \bar{x} \sim (50, \frac{30^2}{25} + \frac{40^2}{16}$

$P(\bar{y} - \bar{x} > 25) = P(Z \geq \frac{25-50}{\sqrt{\frac{30^2}{25} + \frac{40^2}{16}}}) = P(2 \geq -2.14) = 0.9838$

Estimate of the Parameters

Let θ be an unknown parameter in a population. Our goal is to estimate θ using a sample from the population.

We use a function of observations to estimate θ using a sample from the population.

$\hat{\theta} = t(x_1, \ldots, x_n)$

$\hat{\theta}$ is a random variable and depends on $x_1, \ldots, x_n$ (set of observations)

Definition: $\hat{\theta}$ is called an unbiased estimate for θ

$E(\hat{\theta}) = 0$

(74) Let $x_1, \ldots, x_n \sim$ N(μ, σ^2)

We need to estimate both μ and σ^2

One estimate for μ is x_i

$E(x_1) = \mu$

$E(\frac{x_1+x_2}{2} = \frac{E(x_1)}{2} + \frac{E(x_2)}{2} = \mu/2 + \mu/2 + \mu$

$E(\bar{x}) = E(\frac{x_1+\cdots+x_n}{n} = \mu$

For m=3

$E(\frac{2x_1-x_2+x_3}{2} = (2E(x_1) - E(x_2) + E(x_3))/2 = \frac{2\mu-\mu+\mu}{2} = \mu$

Theorem

$E[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2] = \sigma^2$

Proof:

$E(\sum_{i=1}^{n}(x_i - \bar{x})^2) = E(\sum_{i=1}^{2}(x_i - \mu - (\bar{x} - \mu))^2)$

$= E[\sum_{i=1}^{n}(x_i - \mu)^2 + \sum_{i=1}^{2}(\bar{x} - u)^2 - 2\sum(x_i - \mu)(\bar{x} - \mu)]$

$= \sum_{i=1}^{n} E(x_i - \mu)^2 + nE(\bar{x} - \mu)^2 - 2E(n\bar{x} - n\mu)(\bar{x} - \mu)]$

$= n\sigma^2 + n\sigma^2/n - 2nE((\bar{x} - \mu)^2)$

$= (n-1)\sigma^2$

$E[\frac{\sum(x_i - \bar{x})^2}{n-1}] = \sigma^2$

$\bar{\mu} = \bar{x}$

$\bar{\sigma^2} = S^2$

Confidence interval for Unknown Parameters

Let θ be an unknown parameter in a population. We draw a sample from this population. Let $x_1, \ldots, x_n$ be our iid sample.

n = sample size.

We will find $L(x_1, \ldots, x_n)$ and $U(x_1, \ldots, x_n)$ such that:

$P[L(x_1, \ldots, x_n) < \theta < U(x_1, \ldots, x_n)] = 1 - \alpha$

$1 - \alpha$ is the confidence level and usually is close to 1, say 95%, 99%, or 90%

Note: choosing $1 - \alpha$ very close to 1 is not appropriate because in most cases, $U(x_1, \ldots, x_n) - L(x_1, \ldots, x_n)$ will be very large and confidence intervals will become trivial.

The most important parameter in statistics is the mean of the population.

Let $E(x_i) = \mu$ be unknown. The point estimate for μ is $\bar{x}$.

$\hat{\mu} = \bar{x}$

Using C.L.T., $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \longrightarrow Z \sim N(0,1)$

$P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$

The relevant table is in the textbook on pages 742-743

1) For a given confidence level $1 - \alpha$, find z_α from the table of the normal distribution.
   $P(-z_{\alpha/2} < \frac{\bar{x}-\mu}{\gamma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha$
   $P(-z_{\alpha/2} * (\gamma/\sqrt{n}) < \bar{x} - \mu < z_{\alpha/2} * (\gamma/\sqrt{n})) = 1 - \alpha$
   $P(\bar{x} - z_{\alpha/2} * (\gamma/\sqrt{n}) < \mu < \bar{x} + z_{\alpha/2} * (\gamma/\sqrt{n})) = 1 - \alpha$
2) Calculate the set
   $[\bar{x} - z_{\alpha/2} * (\gamma/\sqrt{n}), \bar{x} + z_{\alpha/2} * (\gamma/\sqrt{n})], \bar{x} \pm z_{\alpha/2} * (\gamma/\sqrt{n})$

Sample Size

For a given margin of error E and the confidence level $1-\alpha$,

$E = \frac{z_{\alpha/2}\gamma}{\sqrt{n}} \longrightarrow \sqrt{n} = fracz_{\alpha/2}\gamma E \longrightarrow n = \frac{z_{\alpha/2}^2\gamma^2}{E^2}$

Confidence interval for mean when sigma is unknown. We can estimate sigma sqared by S squared.

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Student:

$\bar{x} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$

$n-1 =$ degrees of freedom.

$df \longrightarrow \infty \longrightarrow t_{n-1} \longrightarrow Z$

$n > 30$

The t table is located on pages 744-745.

$df = \infty$ gives you z values.

$z_{.25} = 0.253$, $z_{.1} = 0.674$, $z_{.05} = 1.645$, $z_{.025} = 1.96$, $z_{.01} = 2.326$

(75) In a sample size of n=20 from a normal population with the variance $\sigma^2 = 225$, the sample mean is $\bar{x} = 64.3$

Contruct a 95% confidence interval for the population mean.

Solution $\alpha = 0.05 \longrightarrow \alpha/2 = 0.025 \longrightarrow z_{\alpha/2} = 1.96$

$\bar{x} \pm z_{\alpha/2}\gamma/\sqrt{n} \longrightarrow 64.3 \pm 1.96\frac{\sqrt{225}}{sqrt20}$

$[57.7, 70.9]$

(76) We would like to estimate the mean thermal conductivity of a certain iron with error less than .1 with 95% confidence. (19/20)

From the previous investigation, we know σ is 0.3. Find the sample size.

$n = \frac{\sigma^2 z_{\alpha/2}^2}{E^2} = \frac{.3^2(1.96)^2}{.1^2}$

$= 34.57 \longrightarrow 35$

$E = \frac{z_{\alpha/2}\gamma}{\sqrt{n}}$

$z_{\alpha/2} = 1.96$

$P[x \in (\mu - 2\sigma, \mu + 2\sigma)] = 0.95$

$R = 4\sigma \longrightarrow \sigma = R/4$

(77) A seed distributor claims 80% of its best seeds will grow.

How many seeds must be tested in order to estimate p (the proportion that will germinate) so that the maximum error of the estimate is 3% with 95% confidence?

$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

error $= z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$

This means that z on each side must be 0.025.

$z_{0.025} = 1.96$

$\hat{p} = 0.80$

$\hat{q} = 0.20$

$0.03 = 1.96\sqrt{0.8 * 0.2/n}$

Solve for m:

$m = z^2_{\alpha/2}\hat{q}\hat{p}/E^2$

$m = \frac{1.96^2 * 0.8 * 0.2}{0.03^2} = 682.95 \longrightarrow 683$

This problem is solved, but in the case that no prior information is provided about p:

$\text{error}^2 = E^2 = z^2_{\alpha/2}\frac{\hat{p}\hat{q}}{m} \longrightarrow n = \frac{z_{\alpha/2}\hat{p}\hat{q}}{E^2}$

$\hat{p}\hat{q} = \hat{p}(1-\hat{p}) = g(\hat{p})$

(78) If pucks used in NHL must have a thickness between 0.9 and 1.1 inches, what percentage of the pucks manufactured by this company can be used in NHL?

This company makes pucks at the size of 1 inch, with a variance of 0.05. Note:

X ~ N(μ, σ^2)

$P(0.9 \leq X \leq 1.1) = P(\frac{0.9-1}{0.05} \leq Z \leq \frac{1.1-1}{0.05}) = P(Z \leq 2) = P(Z \leq -2)$

(79) Let x, y be independent variables.

$E(x) = E(y) = 4$

$Var(x) = Var(y) = z$

$u = 3x - 2y$

$E(u), Var(u)$

$E(3x - 2y) = 3E(x) - 2E(y)$

$= 12 - 8 = 4$

$Var(3x - 2y) = 9 * 2 + 4 * 2 = 26$

$Var(ax \pm by) = a^2Var(x) + b^2Var(y)$

If x is independent from y.

Testing Statistical Hypotheses

Based on a sample of size n we would like to reject or accept a claim (null hypothesis) about a parameter in the population.

(80) A light-bulb company (ACME) has a problem. They have 10,000 light-bulbs (unlabeled), but they do not know if they are long life or regular. The lifetime of both types assumed to have normal distribution with standard deviation of 500 hours.

$E[\text{lifetime}|\text{regular light bulb}] = 1000hours.$

$E[\text{lifetime}|\text{regular light bulb}] = 1500hours.$

$H_0 : \mu 1000$ (null hypothesis)

$H_1 : \mu 1500$ (the alternative)

To solve, draw a sample of n=10 light-bulbs is taken.

$\bar{x} =$ the sample mean

$\bar{x} \geq C \longrightarrow$ reject H_0

How large is C?

C = critical value.

$P($ Type I Error$) = \alpha$

$= P[$ Reject $H_0 | H_0$ is true$]$

$P($ Type II Error$) = P[$Accept $H_0 | H_1$ is true$]$

$= \beta$

$\bar{x}\ N(\mu, \sigma^3/n)$

Let $\alpha = 0.05$

$P[\bar{x} \geq C | \mu = 1000, \sigma = 500] = 0.05$

$P[Z \geq \frac{C-1000}{\frac{500}{\sqrt{10}}}] = 0.05$

$\frac{C-1000}{500/\sqrt{10}} = 1.645$

For $\alpha = 0.05$, find the probability of type II error

$P(\text{Type II Error}) = P[AH_0 | H_1 \text{ is true}]$

$= P[\bar{x} \leq 1260.0973 | \mu = 1500, \sigma = 500]$

$= P[Z \leq \frac{1260.0973-1500}{500/\sqrt{10}}] = 0.0647$

Let $\alpha = 0.01$

$P[\bar{x} \geq C | \mu = 1000, \sigma = 500] = 0.01$

$P[Z \geq \frac{C-1000}{\frac{500}{\sqrt{10}}}] = 0.01$

$\frac{C-1000}{500/\sqrt{10}} = 1.645 \longrightarrow C = 1000 + 2.33 * 500/\sqrt{10} = 1368.4053$

For $\alpha = 0.01$, find the probability of type II error

$P(\text{Type II Error}) = P[AH_0 | H_1 \text{ is true}]$

$= P[\bar{x} \leq 1365.4053 | \mu = 1500, \sigma = 500]$

$= P[Z \leq \frac{1260.0973-1500}{500/\sqrt{10}}] = 0.203$

Since β is much larger, we increase n = 25.

$\alpha = 0.01$

In this case, $C = 1000 + 2.33\frac{500}{\sqrt{25}} = 1233$

$\bar{x} \geq 1233 \longrightarrow$ reject H_0

$\beta = P[\bar{x} \leq 1233 | \mu = 1500, \sigma = 500]$

$P[Z \leq \frac{1233-1500}{500/\sqrt{25}}] = 0.038$

(81) A sample of size n = 10 is taken. All 10 lightbulbs were tested. The mean life time is calculated to be 1300 hours. What is probability we could get a value of $\bar{x}$ at least this extreme value if in fact H_0 is true.

$P[\bar{x} \geq 1300 | \mu = 1000, \sigma = 500](p - value)$

$P(Z \geq \frac{1300-1000}{500/\sqrt{10}} = 0.0288$

Testing on Proportions

In Bernoulli trials, let p = P(success) and q = P(failure).

$x_1, \ldots, x_n$

x_i can be 0 for failure or 1 for success.

$Y = \sum_{i=1}^{n} x_i = \#$ successes $\ Bin(n, p)$

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

$k = 0, 1, 2, \ldots, n$

We would like to test:

1 way: $H_0 : p = p_0$

$H_1 : p > p_0$

1 way:

$H_0 : p = p_0$

$H_1 : p < p_0$

2 way test:

$H_0 : p = p_0$

$H_1 : p \neq p_0$

Algorithm:

First, calculate:

(Z_{ob} is observed Z.)

$Z_{ob} = \frac{\hat{p} - \hat{p_0}}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$ if n is large.

$[np_0, n(1 - p_0) > 5]$

$\hat{p}$ = the sample proportion of the successes.

n = sample size.

1. p - value $= P(Z > Z_{ob})$
2. p - value $= P(Z < Z_{ob})$
3. p - value $= 2 * P(Z > |Z_{ob}|)$
   p-value $< \alpha \longrightarrow RH_0$

(82) The lightbulb company does not want the proportion of defective lightbulbs to exceed 5%.
$H_0 : p = 0.05$
$H_1 : p > 0.05$
A sample of n=100 lightbulbs is taken and we notice that X = # defective lightbulbs = 8
Test H_0 vs H_1, find the value if $\alpha = 0.10$ make a decision.
Solution:
$\hat{p} = \frac{8}{100} = 0.08$
$Z_{ob} = \frac{\hat{p} - 0.05}{\sqrt{0.05 * 0.95/100}}$
$= \frac{0.08 - 0.05}{\sqrt{\frac{0.05 * 0.95}{100}}} = 1.3765$
p - value $= P(Z > 1.3765) = 0.0843 < \alpha \longrightarrow RH_0$
So you reject H_0 and you tell the lightbulb company that they need to step up their game!

Linear Regression

Linear regression is trying to predict something based on something else. For example, if you know the price of gold for the last 200 days, you may be able to predict what will happen to the price of gold tomorrow. It is a very valuable question.

$(x_1, y_1), \ldots, (x_n, y_n)$ are n pairs of observations. If you plot these on a plane, you will see that the data closely follows a linear line.

$y = \beta_0 + \beta_1 x$

β_0 is the intercept, β_1 is the slope.

We try to minimize D, or D^2, where D is the distance between each point on the plane and the line we just made.

$D^2 = (y_1 - \beta_0 - \beta_1 x_1)^2 + (y_2 - \beta_0 - \beta_1 x_2)^2 + \cdots + (y_n - \beta_1 x_n)^2$

Minimize D^2 with respect to β_0 and β_1.

$\frac{\sigma D^2}{\sigma \beta_0} = -2(y_1 - \beta_0 - \beta_1 x_1) - 2(y_2 - \beta_0 - \beta_1 x_2) + \cdots - 2(y_n - \beta_1 x_n) = 0$

$\frac{\sigma D^2}{\sigma \sigma \beta_1} = -2x_1(y_1 - \beta_0 - \beta_1 x_1) - 2x_2(y_2 - \beta_0 - \beta_1 x_2) + \cdots - 2x_n(y_n - \beta_1 x_n) = 0$

$$\begin{cases} \sum_{i=1}^n y_i = n\beta_0 + \beta_1 \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i = \beta_0 \sum_{i=1}^n x_i + \beta_1 \sum x_i^2 \end{cases}$$

$\bar{y} = \beta_0 + \beta_1 \bar{x}$

$y = \beta_0 + \beta_1 x$

$A|_{\bar{y}}^{\bar{x}}$

$\bar{y} = \beta_0 + \beta_1 x$

$$\begin{cases} \sum x_i * (\sum_{i=1}^n y_i = n\beta_0 + \beta_1 \sum_{i=1}^n x_i) & (\sum x_i)(\sum y_i) = n\beta_0 \sum x_i + \beta_1 (\sum x_i)^2 \\ \sum_{i=1}^n x_i y_i = \beta_0 \sum_{i=1}^n x_i + \beta_1 - n * (\sum x_i^2) & -n \sum x_i y_i = -n\beta_0 \sum x_i - n\beta_1 \sum x_i^2 \end{cases}$$

$-n \sum x_i y_i + (\sum x_i)(\sum y_i) = \beta_1 (-n \sum x_i^2 - (\sum x_i)^2)$

$\hat{\beta_1} = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$

We will manipulate the formula later to make it prettier. It is a useful one, so useful that most calculators now have “a and b”, which uses this formula when you enter data.

(83) Earthquakes were recorded all over the world. The x=latitude and y=magnitude of the earthquakes were found to be:

x	y
60	4.1
77.5	4
50.7	2.6
...	...
48.3	0.9

In R, enter the data and use plot(x,y), then a = lm(y~x), finally, use summary(a).

What we need to find manually…

x	y	x^2	xy
60	4.1	3600	246
77.5	4	6006.25	310
50.7	2.6	2570.49	131.82
…	…	…	…
48.3	0.9	2332.89	43.47

Note: there are 13 entries.

The sum of the data above:

x	y	x^2	xy
742.3	35.1	43344.79	2100.84

$\hat{\beta}_1 = 0.1007$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 x_1 = 3.04997$

$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i * \sum y_i}{n}}{\sum x_i^2 - \frac{\sum x_i}{n}^2} = \frac{S_{xy}}{S_{xx}}$

$S_{xx} = \sum (x_i - \bar{x})^2$

$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}$