MA240: Proability & Statistics

Midterm: 30% (February 23, in lab) Labs: 25% Final exam: 45%

Probability

- Probability is defined as a measure of one's belief in the occurence of a random event
- Probability is also known as "the mathematics of uncertainty"

Assigning Probabilities

- Subjective approach
  - This approach is based on feeling and may not even be scientific
- Relative frequency approach
  - This approach can be used when some random phenomenon is observed repeatedly under identical conditions
- Axiomatic/Model-Based approach (this course)

Definition. *An experiment is any action or prcess whose utcome is subject to uncertainty*

Example. *Tossing coins, throwing dice, observing lifetime of a computer*

Definition. *The sample space, S, of an experiment is the set of all possible outcomes*

Example. *Tossing a coin, $S = \{H, T\}$*

Throw a single die: $S = \{1, 2, 3, 4, 5, 6\}$

Observe lifetime, t, of computer: $S = \{t\colon t \geq 0\}$

Definition. *An event is a set of outcomes of the random phenomenon; the event is a subet of S*

Example. *Tossing a coin: $A = \{H\}$*

Throw a single die: $A = \{2, 4, 6\}$

Observe lifetime, t, of a computer: $A = \{t\colon t \geq 12\,\text{months}\}$

Definition. *A probability model is a mathematical description that shows the sample space, S, and a way of assigning probabilities to events*

Throw a Single Die

- There are six possible outcomes - the sample space is $\{1, 2, 3, 4, 5, 6\}$
- Example event: the face the shows is even - $\{2, 4, 6\}$

- Probability model - assign a number $\frac{1}{6}$ to each one of the outcomes of the sample space (include each face of a die)

Basic Set Theory

- Suppose that A and B are sets (events). A is a subset of B if every outcome in A is also in B, denoted $A \subset B$ or $A \subseteq B$.

  Note. $A \cap B \subseteq A \cup B$

- Disjoint sets are denoted $A \cap B = \emptyset$

- A compliment denoted, A' or A^C

  Homework: represent Distributive Laws and DeMorgan's Laws with venn diagrams

- A countable set A is a set whose elements can be put into a 1-1 correspondence with $\mathbb{N} = \{1,\ 2,\ ...\}$, the set of natrual numbers. A set that is not countable is said to be uncountable. They can be divided into further sets:

  - A countably infinite set has an infinite number of elements

  - A countably finite set has a finite number of elements

- **Homework:** $S = \{1, 2, ..., 9\}$, $A = \{1, 3, 5, 7\}$, $B = \{6, 7, 8, 9\}$, $C = \{2, 4, 8\}$, $D = \{1, 5, 9\}$

  a) $A' \cup B$

  b) $(A' \cap B) \cap C$

  c) $B' \cup C$

  d) $(D' \cup C) \cap D$

  e) $A' \cup C$

  f) $A' \cup C \cap D \cap B$

More on Events:

Let $A_1, ..., A_n$ be a sequence of events of a sample space, S.

We can define $A_1 \cup A_2 \cup ... \cup A_n = \bigcup_{i=1}^{n} A_i$ and $A_1 \cap ... \cap A_n = \bigcap_{i=1}^{n} A_i$

DeMorgan's Law:

1. $\left(\bigcup_{i=1}^{n} A_i\right)' = \bigcap_{i=1}^{n} A_i'$

2. $\left(\bigcap_{i=1}^{n} A_i\right)' = \bigcup_{i=1}^{n} A_i'$

Suppose $A_1 \cup ... \cup A_k = S$, events A_j for $j = 1, ..., k$ are exhaustive of at least one event occurs.

Example. $S = \{1, 2, 3, ..., 10\}$

$A\{1,2,3\}$, $B=\{4,5,6,7,8,9\}$ and $C=\emptyset$

$A\cup B\cup C=S$ $\quad\therefore A$, B, C are exhaustive

3 Discrete Random Variables & Probability Distributions

Theorem. ***Kolmogorov Axioms of Probability***

Given a nonempty sample space, S, he probability of A is a function, $P(A)$, satisfying three axioms:

1. *$P(A)\geq 0$ for every $A\subseteq S$*
2. *$P(S)=1$*
3. *If $A_1, A_2, \ldots$ is an infinite collection of distinct events, such that $A_i\cap A_j=\emptyset$, then:*

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

Remark. Based on *3.*, we have:

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

provided that $A_i\cap A_j=\emptyset$.

Example.

1. Let $S=\mathbb{Z}$

   $A=\{\text{pos. integers}\}$, $B=\{\text{neg. integers}\}$ and $C=\{0\}$

   $A\cap B=\emptyset$, $B\cap C=\emptyset$ and $A\cap C=\emptyset$

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

   $P(A\cup B\cup C)=P(S)=1$

Convention. Rules for $P(A)$

1. **Complement Rule:** $P\left(A'\right)=1-P(A)$

   Proof. We know $A\cap A'=\emptyset$

   $\therefore P(A\cup A')=P(A)+P\left(A'\right)$

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

   $\therefore 1=P(A)+P(A')$ and thus, $P(A')=1-P(A)$ $\square$

2. $P(\emptyset)=0$

   Since $\emptyset\cap S=\emptyset$, $\therefore S=\emptyset$

   By the Complement Rule, we have:

$$P(\emptyset)=1-P(S)=1-1=0$$

3. **Inclusion-Exclusion Rule**

   If A and B are any two events in a sample space, then

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

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

   Example. The probability that train 1 is on time on is 0.95. The probability of train 2 is on time is 0.93. The probability of both trains being on time is 0.90.

   a) What is the probability that at least on train is on time?

Solution. $A_1 = \{\text{train 1 is on time}\}$, $A_2 = \{\text{train 2 is on time}\}$

$P(A_1) = 0.95$ and $P(A_2) = 0.93$ are given.

$A_1 \cap A_2 = \{\text{both trains are on time}\}$ with $P(A_1 \cap A_2) = 0.90$

$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = 0.95 + 0.93 - 0.90 = 0.98$

   b) What is the probability that neither of the trains is on time?

Solution. $A_1' = \{\text{train 1 is not on time}\}$ and $A_2' = \{\text{train 2 is not on time}\}$

$A_1' \cap A_2' = \{\text{neither train is on time}\}$

$P(A_1' \cap A_2') = P(A_1 \cup A_2)' = 1 - P(A_1 \cup A_2) = P(A_1') + P(A_2') - P(A_1' \cup A_2') = 1 - 0.98 = 0.02$

Definition. *[Assigning probability] If a sample space for an experiment contains a finite or countable number of outcomes, then S, is a discrete sample space.*

Definition. ***An equaprobability model:***

Suppose that a discrete sample space, S, contains $N < \infty$ units, each of which are equally likely.

Example. $S = \{H, T\} \rightarrow P(\{H\}) = P(\{T\})$

Let A be an element of the discrete sample space, S, and n_a ($n_a \leq N$) be the number of outcomes (units) in A, then:

$$P(A) = \frac{n_a}{N}$$

Example.

1. Tossing two coins:

$$S = \{HH, \mathrm{HT}, \mathrm{TH}, TT\} \text{ with } N = 4$$

   with $A = \{\text{set of at least one head}\} = \{HH, \mathrm{HT}, \mathrm{TH}\}$, $n_a = 3 \rightarrow P(A) = \frac{3}{4}$

2. Two jurors are needed from pool of 2 men and 2 women.

   What is the probability that the two jurors are chosen, consist of 1 male and 1 female.

$$S = \{(M_1, M_2), (M_1, F_1), (M_1, F_2), (M_2, F_1), (M_2, F_2), (F_1, F_2)\} \text{ with } N = 6$$

$$A = \{(M_1, W_1), (M_1, W_2), (M_2, W_1), (M_2, W_2)\}, \therefore n_a = 4 \rightarrow P(A) = \frac{4}{6} = \frac{2}{3}$$

Counting

Multiplication Rule: Multiplicatively summing up each level of a tree diagram

$$\begin{aligned} n_1 &= \#\,\text{of ways of stage } 1 \\ n_2 &= \#\,\text{of ways of stage } 2 \\ &\vdots \\ n_k &= \#\,\text{of ways of stage } k \end{aligned}$$

Thus, the total number of ways of the operation $= n_1 \times n_2 \times ... \times n_k$

Example. Throwing a die twice, how many possibilities?

$$6 \times 6 = 36$$

Example. Menu: 4 soups, 3 sandwiches, 5 desserts and 4 drinks.

$$\text{Total number of choices} = 4 \times 3 \times 5 \times 4 = 240$$

[by Multiplication Rule]

Definition. *A permutation is an arrangement of distinct objects in a particular order.*

Definition. *If r objects are chosen from a set of n distinct objects, any particular order of these r objects is called a permutation.*

$$\underline{n} \times \underline{n-1} \times \underline{n-2} \times ... \times \underline{n-(r-1)}$$

By multiplication rules: # of permutations $= n \times n-1 \times n-2 \times ... \times (n-(r-1)) = \frac{n!}{(n-r)!}$

Formally:

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

Example. How many permutations are there of all three letters, a,b,c?

Solution. $_nP_n = 3!$

Example. Four names are drawn from 24 members of a club for P, VP, T, and S. In how many ways can this be done?

Solution. $24P4 = 24 \times 23 \times 22 \times 21 = 255024$

Special Types of Permutation:

1. Permutation of n distinct objects

$$_nP_n$$

2. Cirrcular Permutation: permutation that occurs whn objects are arranged in a circle. In order to do this: fix one object, permute the rest of the $n-1$ objects

$$\#\,\text{of permutations} = {}_{n-1}P_{n-1}$$

Example. How many permutations are there of four persons play bridge?

$$3! = 6$$

3. Permuations with repeated objects: $BO_1O_2K \to {}_nP_n = 24$

$$\mathrm{BO_1KO_2} = \mathrm{BO_2KO_1}$$

Thus the total number of ways $= \frac{24}{2!} = 12$

$\therefore$ In general:

$$\begin{gathered}\text{let } n = \#\,\text{of objects}\\ k = \#\,\text{of objects}\\ n_j = \text{number of objects in } j^{\text{th}}\text{ group}\\ \therefore n = n_1 + n_2 + \ldots + n_k = \sum_{j=1}^{k} n_j\\ \therefore \#\,\text{of permutations} = \frac{n!}{n_1!n_2!\ldots n_k!}\end{gathered}$$

Example. Pepper

$$\begin{gathered}n_p = 3, \qquad n_e = 2, \qquad n_r = 1\\ n_p + n_e + n_r = 6\\ \therefore \#\,\text{of permutations} = \frac{6!}{3!2!1!} = 60\end{gathered}$$

Definition. *A* <u>*combinations*</u> *is a selection of r objects from n distinct objects without considering the order in which they are selected.*

Example. Club members, choose 4 members from 24 for the executives.

We have ${}_{24}C_4$.

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

Example. 20 computers. If 5 defect, we will return the order. Suppose that there are 3 defects in the order, what is the probability of them accepting the order. Worst question.

Solution. Total # of outcomes in S:

$$N = \binom{20}{5} = \frac{20!}{(20-5)!5!} = 15504$$

$A = \{\text{order is accepted}\} = \{5 \text{ good computers}\}$

$\#$ of ways A can occur:

$$\underline{\text{Stage 1:}}\,\text{choose 5 good computers:}\binom{12}{5},{}_{12}C_5$$
$$\underline{\text{Stage 2:}}\;\text{choose 0 bad computers:}\binom{8}{0},{}_8C_0$$

$\therefore$ By multiplication rule, $n_n = \binom{12}{5}\binom{8}{0} = 792$, $P(A) = \frac{792}{N} =$

Example.

1. What is the probability of drawing an ace from a deck of 52 cards? $P(\text{ace}) = \frac{4}{52}$
2. What is the probability to have a full house from a desk of 52? [3 of a kind, 2 of a pair]

Solution. There are 13 face values.

$\#$ of ways of getting a 3 of a kind: $\binom{4}{3}$

$\#$ of ways of getting a pair: $\binom{4}{2}$

$\#$ of ways of getting a full house: $13 \cdot \binom{4}{3} \times 12\binom{4}{2}$

$\#$ of ways of getting a full house: $\binom{52}{5}$

$\therefore P(\text{full house}) = \frac{13 \cdot \binom{4}{3} \times 12 \cdot \binom{4}{2}}{\binom{52}{5}}$

Definition. *Two events, A and B, are* <u>*independent*</u> *iff:*

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

provided that $P(A), P(B) > 0$.

[If the probability of event A is not related to the probability of event B, then A is said to be independent to B.]

Note. *A and B are mutually exclusive, $\neq$, A and B are independent.* $P(A \cap B) = \emptyset$

Example. Coins being tossed:

$$\{\text{HT}, \text{HT}, \text{TH}, TT\} = S$$

We have $P(\text{HH}) = \frac{1}{4}$, $P(H) = \frac{1}{2}$, $P(H) = \frac{1}{2}$

From this, $P(HH) = P(H) \cdot P(H) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

Properties of Independent Events:

If A and B are independent events:

a) A' and B are independent $\rightarrow P(A' \cup B) = P(A') \cdot P(B)$

b) A and B' are independent $\rightarrow P(B' \cap A) = P(B') \cdot P(A)$

c) A' and B' are independent $\rightarrow P(A' \cap B') = P(A') \cdot P(B')$

Conditional Probability

Motivation. *Prior knowledge about the likelihood of events may be related to the event of interest*

Definition. *Let A and B be events in a nonempty sample space, S. The conditional probability of A, given B has occurred is given by:*

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

provided that $P(B) > 0$.

Example. A family has two children.

a) What is the probability that both are girls?

Solution. $S = \{(B,B),(B,G),(G,B),(G,G)\}$

$A_1 = \{1^{\text{st}}\text{ is } a \text{ girl}\} = P(A_1) = \frac{1}{2}$

$A_2 = \{2^{\text{nd}}\text{ is } a \text{ girl}\} = P(A_2) = \frac{1}{2}$

$P(A_1 \cap A_2) = \{(G,G)\} = P(A_1) \cdot P(A_2) = \frac{1}{4}$

b) What is the probability of both girls if the elder is a girl?

Solution. $P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{\frac{1}{2}}{\frac{1}{4}} = \frac{1}{2}$

|

Axiom. ***Properties of Conditional Probability***

1. $P(A|B) > 0$
2. $P(B|B) = 1$
3. $P\left(\bigcup_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} P(A_i|B)$

   First 3 imply 4, 5, 6:
4. $P(A'|B) = 1 - P(A|B)$
5. $P(\emptyset|B) = 0$
6. *Inclusion-Exclusion laws:* $P(A \cup C|B) = P(A|B) + P(C|B) - P(A \cap C|B)$
7. *Multiplication Rule:* $P(A \cap B) = P(A|B) \cdot P(B)$
8. *Divide A into two paritions: $A \cap B$ and $A \cap B'$*

   $A = (A \cap B') \cup (A \cap B)$ *with $A \cap B'$ and $A \cap B$ disjoint (mutually exclusive)*

   $\therefore$ $P(A) = P(A \cap B') + P(A \cap B) \rightarrow$ *the* ***Law of Total Probability***

   Using the Multiplication Rule: $P(A) = P(A|B') \cdot P(B') + P(A|B) \cdot P(B)$
9. $P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \rightarrow = \frac{P(A|B) \cdot P(B)}{P(A|B') \cdot P(B') + P(A|B) \cdot P(B)}$

Example. In an apartment building: 36% of the residents own a dog, $P(D)$, 30% of residents own a cat, $P(C)$, and 22% of residents that own a dog, $P(C|D)$, own a cat.

a) What is the probability that a resident owns both a cat and a dog?

Solution. $P(C \cap D) = ?$

$P(C) = 30\%$

b) What is the probability that a resident own a dog, given that it owns a cat?

Solution. $P(D|C) = ?$

Example. An insurance company classifies people as accident prone, $P(B)$, or non-accident prone, $P(B')$.

The probability that an accident-prone person has an accident is 0.4, $P(A|B)$. The probability that a non-accident prone person has an accident is 0.2, $P(A|B')$.

a) What is the probability that a new policy holder will have an accident?

Solution. $A = \{\,\text{policy holder has an accident}\,\}$, $B = \{\,\text{policy holder is accident} - \text{prone}\,\}$

$$P(A) = P(A|B) \cdot P(B) + P(A|B') \cdot P(B')$$

$= 0.4 \times 0.3 + 0.2 \times 0.7$

$= 0.26$

b) Suppose that the policy holder does not have an accident, what is the probability that she is an accident-prone person?

Solution. $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B) \cdot P(B)}{P(A)} = 0.46$

> **Remark.** If A and B are independent, $P(A|B) = P(A)$ and $P(B|A) = P(B)$

c) Are A and B independent?

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

$P(A \cap B) = 0.4 \times 0.3 = 0.12$

$P(A) \cdot P(B) = 0.26 \times 0.3 = 0.078$

Since, $P(A \cap B) \neq P(A) \cdot P(B)$, A and B are not independent events.

Example. Allergic reaction to a drug; 8 subjects

Outcomes in sample space $= 2^8 = 256$

$\rightarrow$ very large

$\rightarrow$ may not be our interest

Random Variable

Definition. *If, S, is a sample space, a random variable, x, is a real-valued function of S, such that:*

$$x: S \rightarrow \mathbb{R}$$

Example.

a) *A fair coin is tossed twice: $S = \{HH, \mathrm{HT}, \mathrm{TH}, TT\}$. We have $x =$ total number of heads*

Each probability is $\frac{1}{4}$, *where* x *is 2, 1, 1, 0.*

X: *upper case, generic function*

x: *lower case, possible value of* X, *realization of* X

Characteristics of Random Variable

1. Probability distribution of $x \Rightarrow$ a function of x, that describes the probability of x

   $P(X=2)=\frac{1}{4}$. Ex. $P(X=x)=\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{1-x}$, $x=0,1,2$

   $F(x)=P(X\leq x)\Rightarrow$ **cummulative distribution function (cdf)** or $P(a\leq b\leq c)$

2. Expectations of a random variable: (values that are derived from the probability functions).

   Definition. *The* ***cdf*** *of a random variable,* X*, is the function* $F:\mathbb{R}\rightarrow[0,1]$ *given by* $F(x)=P(X\leq x)$

   Properties of cdf:

   1. $\underset{x\rightarrow\infty}{\mathcal{L}} F(x)=0$

   2. $\underset{x\rightarrow\infty}{\mathcal{L}} F(x)=1$

3. If $a<b$, then $F(a)<F(b)$.

4. $0\leq F(x)\leq 1$

Types of Random Variables

- Discrete Random Variable
  - Defined over discrete, S
  - x takes value from finite or countably infinite set

    Notation. $P(x)=P(X=x)$ *for* $x\in R$, $R\subseteq\mathbb{R}$ *where* R *is called the support of random variabel*

- Continuous Random Variable
  - Define over continuous , S
  - x takes value from a given interval

Example. Given a discrete random variable, X, and $P(x)=\frac{x+2}{25}$ for $x\in R$, $R=\{1,2,3,4,5\}$ and 0 for $x\notin R$.

1. Is $P(x)$ a pmf. $P(x)>0$

$$\sum_{x\in R} P(x)=\frac{3}{25}+\frac{4}{25}+\frac{5}{25}+\frac{6}{25}+\frac{7}{25}=1$$

2. $P(x \le 3) = P(x < 1) + P(x = 1) + P(x = 2) + P(x = 3) = 0 + \frac{3}{25} + \frac{4}{25} + \frac{5}{25} = \frac{12}{25}$

Example. Determine C in the following function such that $P(x)$ is a pmf.

$P(x) = C(1-p)^{x-1}p$ for $x \in \{1, 2, ..., k\}$, 0 otherwise

Solution.

1. $C > 0$

2. $P(x)$ is a pmg if $\sum_{x \in R} P(x) = 1$ or $\sum_{x \in R} C(1-p)^{x-1}p = 1$, $C = \frac{1}{\sum_{x \in R}(1-p)^{x-1}p} = \frac{1}{1-p}$

   $\sum_{x \in R}(1-p)^{x-1}p = p\sum_{x=1}^{k}(1-p)^{x-1}p$, let $y = x - 1$

   $= \sum_{y=0}^{k-1}(1-p)^y$, $x = 1$, $y = 0$

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

Remark. Suppose X is a d.r.v, the probability of an event $\{X \in B\}$ is computed by adding the $P(x)$ for $x \in B$.

$$P(X \in B) = \sum_{x \in B} P(x), B \subseteq R$$

Example. $P(x \le 3)$, $B = \{1, 2, 3\} = \sum_{x \in B} P(x)$

Example. $P(x) = \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{1-x}$, x

$$F(x) = \sum_{t \le x} \left(\frac{1}{3}\right)^t \left(\frac{2}{3}\right)^{1-t}$$

Example. There are three stop lights from work to school. Stop/cross

$S = \{CCC, C\,\text{CS}, \text{CSC}, \text{SC}\,C, \text{CS}\,S, \text{SCS}, S\,\text{SC}, SSS\}$, $x =$ number of stops ($x = 0, 1, 2, 3$)

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

Table 1.

CDF.

I_1	I_2	I_3	I_4	I_5
	0	1	2	3 x

with $F(x) = \sum_{t \le x} P(t), 0 \le x \le 1$

$$F(x): \begin{cases} P(x=0) = \frac{1}{8} \text{ with } x > 0 = F(1) \\ P(x=0) + P(x=1) = \frac{1}{2} \text{ where } 0 \le x < 1 = F(2) \\ P(x=0) + P(x=1) + P(x=2) = \frac{7}{8} \text{ where } 2 \le x < 3 = F(3) \\ P(x=0) + P(x=1) + P(x=2) + P(x=3) = 1 \text{ where } 3 \le x = F(4) \end{cases}$$

Answer.

1. If the range of a d.r.v consists of values: $x_1 < x_2 < ... < x_n$

   a. $P(x_1) = F(x_1)$

   b. $P(x_i) = F(x_i) - F(x_i - 1)$

2. $P(a \leq x \leq b) = F(b) - F(a-1)$, where $a, b, c \in$ non-integer set

Example.

$$F(x) = \begin{cases} 0 \text{ where } x < 1 \\ \frac{1}{3} \text{where } 1 \leq x < 4 \\ \frac{1}{2} \text{ wnere } 4 \leq x < 6 \\ \frac{5}{6} \text{ where } 6 \leq x < 10 \\ 1 \text{ where } x \geq 10 \end{cases}$$

Find:

1. $P(2 < x \leq 6)$ *

2. $P(x=4) = F(4) - F(3) = \frac{1}{2} - \frac{1}{3}$

3. $P(x \leq 5) = F(5) = \frac{1}{2}$

4. $P(0.5 < x \leq 2) = P(0.5 < x < 1) + P(1 \leq x \leq 2) = F(2) - F(0) = \frac{1}{3}$

Expectations

Let $g(x)$ be a function of a d.r.v then the expectation of $g(x)$ is:

$$E(g(x)) = \sum_x g(x) \cdot P(x)$$

Example. $P(x) = \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{1-x}$, then we have;

$$E(X) = \sum_{x=0,1} \left(\frac{2}{3}\right)^x \left(\frac{1}{3}\right)^{1-x} \cdot x = \left(\frac{1}{3}\right)^{1-0} \cdot 0 + \left(\frac{2}{3}\right)^1 \cdot 1 = \frac{2}{3}$$

Special Expectations.

1. *Mean* of x (expected value of x), $g(x) = X$ (first moment of x). $\mu = E(x) = \sum_x x \cdot P(x)$

2. *Variance* of x, $g(x) = (x - \mu)^2$

$$\sigma^2 = E[(x-\mu)^2] = \sum_x (x-\mu)^2 \cdot P(x)$$

   Standard deviation: $\sqrt{\sigma^2} = \sigma$

3. *Second moment of* x, $g(x) = X^2$

$$E\left(x^2\right) = \sum_x x^2 \cdot P(x)$$
$$\sigma^2 = E(x^2) - E(x)^2$$

Example. Game: [Dice are rolled]

	Payoff
$x \leq 6$	0
$7 \leq x \leq 9$	5
$10 \leq x \leq 11$	15
$x = 12$	20

Table 2.

x is sum readings. What is the expected value of payoff? In other words, the variance of the payoff?

$Y =$ payoff, $y = 0, 5, 15, 20$

y	0	5	15	20
$P(y)$	$P(x \leq 6)$	$P(7 \leq x \leq 9)$	$P(10 \leq x \leq 11)$	$P(x = 12)$
y^2	0	25	225	400

$$E(Y) = \sum_y y \cdot P(y) = 0 \times \frac{15}{36} + 5 \times \frac{15}{36} + 15 \times \frac{5}{36} + \frac{20}{36}$$

$$E(Y^2) = \sum_y y^2 \cdot P(y) = 0 \times \frac{15}{36} + 25 \times \frac{15}{30} + 225 \times \frac{5}{30} + 400 \times \frac{1}{36}$$

Properties of Expectation (for both c.r.v and d.r.v) (section 3.4 omitted)

1. $E(a) = a$; a is a constant

$$E(a) = \sum_{x \in R} a \cdot P(x) = a \cdot \sum_{x \in R} P(x)$$

2. *Linearity property*:

$$E(a \cdot g(x) + b) = a \cdot E(g(x)) + b$$

where a, b are constants.

$$E\left(\frac{g(x)}{q(x)}\right) \neq \frac{E(g(x))}{E(q(x))}$$

Example. $E\left[\frac{x^2}{e^x}\right] = \sum_{x \in R} \frac{x^2}{e^x} \cdot P(x) \neq \frac{\sum x^2 \cdot P(x)}{\sum e^x \cdot P(x)}$

3. $E[g_1(x_1) + g_2(x_2) + \ldots + g_k(x_k)] = E[g_1(x_1)] + E[g_2(x_2)] + \ldots + E[g_k(x_k)]$

$$E\left[\sum_{i=1}^{k} g_i(x_i)\right] = \sum_{i=1}^{k} E(g_i(x_i))$$

More on Special Expectation:

- $E(g(x))$: expected value of $g(x)$

    1. $g(x) = X^k \Rightarrow$ The k^{th} moment of the random variable X

    Notation. *k^{th} moment $\equiv \mu_k$*

    $E(x) = \mu_1,\ E(x^2) = \mu_2$

    - $g(x) = (x - \mu_1)^k$

    k^{th} central moment

    $\sigma^2 = \text{Var}(X) = E[(x - \mu_1)^2] \equiv 2^{\text{nd}}$ *central moments*

    - $g(x) = e^{t \cdot x}$*; t is a real-valued number*

    Moment generating function $= E(e^{t \cdot x})$

Variance:

1. $\text{Var}(x) = \mu_2 - \mu_1^2 = E(x^2) - E(x)^2$

    Proof.

$$\begin{aligned} \text{Var}(X) &= E[(x - \mu_1)^2] \\ &= E[x^2 - \mu_i \cdot x + \mu_i^2] \end{aligned}$$

Note. $E(g(x))$ is a constant

$$\begin{aligned} &= E(x^2) + E(-2\mu_1 \cdot x) + E(\mu_1^2) \\ &= \mu_2 - 2\mu_1 \cdot E(x) + \mu_1^2 \\ &= \mu_2 - 2\mu_1^2 + \mu_1^2 \\ \text{Var}(x) &= \mu_2 - \mu_1^2 \end{aligned}$$

□

2. $\text{Var}(a + bx) = b^2 \cdot \text{Var}(x)$

Discrete Uniform Distribution

Pmf: $P(x) = \begin{cases} \frac{1}{m}, x = 1, 2, \ldots, m \\ 0, \text{otherwise} \end{cases}$

$$\begin{aligned} E(x) &= \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m} \cdot \sum_{x=1}^{m} x = \frac{m+1}{2} \\ E(x^2) &= \frac{1}{6}(m+1)(2m+1) \\ \text{Var}(x) &= E(x^2) - E(x)^2 = \frac{m^2 - 1}{12} \end{aligned}$$

Bernoulli Trial

1. Each trial results in a “success” and “failure.”

2. All the trials are independent.

   $\Rightarrow$ Probability getting a "success" is not affected by the probability of another trial

3. Probability of sucess: p

Bernoulli Distribution

Random variable: $X = \#$ of successes in one trial

Values of r.v: $\quad x = 0, 1$

$$\text{Pmf:} \quad P(X = x) = \begin{cases} p, x = 1 \\ 1 - p, x = 0 \end{cases}$$

$$P(x) = \begin{cases} p^x (1 - p)^{1 - x}, x = 0, 1 \\ 0, \text{otherwise} \end{cases}$$

Derivations:

$$\begin{aligned} E(x) &= \sum_{x = 0, 1} x \cdot P(x) = 0 \cdot p(0) + 1 \cdot p(1) = p \\ E(x^2) &= 0^2 \cdot p(0) + 1^2 \cdot p(1) = p \\ \text{Var}(x) &= p - p^2 = p \cdot (1 - p) \end{aligned}$$

Binomial Distribution

Rule.

1. n Bernoulli trials

$$\begin{aligned} r.v \quad & x = \#\,\text{of success in}\, n\,\text{trials} \\ & x = 0, 1, 2, \ldots, n \\ \text{Notation:} \quad & x \sim b(n, p) \end{aligned}$$

$P(X = x) = ?$

Total number of sequences: $\binom{n}{x}$

Probability of getting x successes: $p^x (1 - p)^{n - x}$

$\therefore P(x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n - x}, x = 0, 1, \ldots, n \\ 0, \text{otherwise} \end{cases}$

Properties of Binomial Distribution

1. $P(x) = b(x; n, p) = \binom{n}{x} p^x (1 - p)^{n - x} = \binom{n}{n - x} (1 - p)^{n - x} p^x$

2. $\sum_{x = 0}^{n} \binom{n}{x} p^x (1 - p)^{n - x} = 1 \Rightarrow$ *binomial sum*

$$\Rightarrow \text{ Note: } \sum_{k = 0}^{m} \binom{m}{k} r^k = (1 + r)^m$$

3. $E(x) = n \cdot p$

4. $\text{Var}(x) = n \cdot p(1 - p)$

5. $F(x) = B(x; n, p) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k} = P(X \le x)$

Example. The probability of recovery from a tropical disease is 0.8. Given that there are 10 patients in the clinic, that have the disease.

a) Find the probability of 7 people recovering.

$P(\text{recovery}) = 0.8 = P,\ n = 10$

$P(X = x)$, where $x = 0, 1, 2, ..., 10$

$P(x) = b(x; 10, 0.8) = \binom{10}{x} 0.8^x 0.2^{10-x};\ P(7) = \binom{10}{7} 0.8^7 0.2^3 \approx 0.2$

b) What is the probability that at least 5 people will recover?

$P(\text{at least 5 people}) = P(x \ge 5) = 1 - P(x < 5) = 1 - P(x \le 4)$

$$\begin{aligned} &= 1 - F(4) \\ &= 1 - B(4; 10, 0.8) \\ &= 1 - 0.006 \\ &= 0.994 \end{aligned}$$

c) What is the probability that 6 to 8 people will recover?

$P(6 \le x \le 8) = P(6) + P(7) + P(8) = b(6; 10, 0.8) + b(7; 10, 0.8) + b(8; 10, 0.8)$ OR

$$\begin{aligned} &= F(8) - F(5) \\ &= 0.6242 - 0.0328 \\ &= 0.59.14 \end{aligned}$$

d) What is the expected number of recovery?

$E(x) = n \times p = 10 \times 0.8 = 8$

e) What is the variance of the number of recovered patients?

$$\begin{aligned} \text{Var}(x) &= n \cdot p(1-p) \\ &= 10 \times 0.8 \times 0.2 \\ &= 1.6 \end{aligned}$$

Negative Binomial Distribution

Bernoulli Trials

x^{th} trial: $\{SSSS...FS...SF\} \leftarrow\ x - 1$ trials, with $k - 1$ successes

$$\begin{aligned} 52^{\text{nd}}\,\text{trial} & \quad 53^{\text{rd}}\,\text{trial} \\ 10\,\text{winning} & \quad 11\,\text{winnings} \\ \text{Random variable} &= x\,\text{trials where}\,k\,\text{success happens} \\ & \quad \text{no total number of trials} \end{aligned}$$

Properties of Negative Binomial Distribution

1. $P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k} = \frac{k}{2} b(k; x, p)$

2. $E(x) = \frac{k}{p}$

3. $\text{Var}(x) = \left(\frac{1}{p} - 1\right) \cdot \frac{k}{p}$

Example. $P(\text{catch the flu}) = 0.40$. What is the probability that the 10^{th} child exposed to the flu is the third person to catch it.

$P = 0.40$, $x = 10$, $k = 3$

$$P(x = 3) = \binom{9}{2} 0.4^3 \cdot 0.6^7 = 0.0645$$

Geometric Distribution

Special case of Negative Binomial Distribution

$k = 1$, Trial: $\{FFFF\ldots FS\} \leftarrow$ tht is the first success

$P(x) = p(1-p)^{x-1}$

Example. Tossing coin until heads appears.

$E(x) = \frac{1}{p}$, $\text{Var}(x) = \frac{1}{p} \cdot \left(\frac{1}{p} - 1\right)$

Hypergeometric Distribution

1. Success/failure

2. Dependent trials

$$\begin{aligned} n &= \text{total \# of successes} \\ x &= \text{\# of success in the selected unit} \rightarrow n - x \end{aligned}$$

Random variable:

$$X = \text{\# of success in the } n \text{ selected unit}$$

Given information:

$$\begin{aligned} N &= \text{\# of finite population} \\ a &= \text{\# of success} \\ n &= \text{\# of selected} \end{aligned}$$

$$\begin{aligned} P(x) = \quad P(X = x) &= \frac{\binom{N-a}{n-x} \cdot \binom{a}{x}}{\binom{N}{n}} \\ E(x) &= \frac{n \cdot a}{N} \\ \text{Var}(x) &= \frac{n \cdot a(N-a) \cdot (N-n)}{N^2(N-1)} \end{aligned}$$

Example. Computer defects

$\#$ of defects $= 2 = a$

\# of computers in the order $= 12 = N$ and we select 3 computers to test, n

1. Probability mass function for the defects

$$P(x) = \frac{\binom{2}{x} \cdot \binom{10}{2-x}}{\binom{12}{3}}$$

2. Expected # of defects

$$E(x) = \frac{n \cdot a}{N} = \frac{3 \times 2}{12} = \frac{1}{2}$$

$$\mathrm{Var}(x) = \frac{15}{44}$$

3. $P(\text{at least two defects}) = P(x \leq 2)$

$$= \sum_{t \leq 2} P(t) = P(0) + P(1) + P(2)$$

Example. If the probability of suffering from heat exhaustion is 0.005, $p = 0.005$;

a) What is the probability that 18 of the 3000 people attending the parade with suffer from heat exhaustion?

   $n = 3000$

$$P(x = 18) = b \cdot (18; n, p) = \binom{3000}{18} \cdot (0.005)^{18} \cdot (0.995)^{2985}$$

b) What is the probability that more than 10 people will suffer heat exhaustion? [Completed after **Poisson Distribution**]

$$P(x > 10) = 1 - P(x \leq 10) = 1 - F(10) = 1 - B \cdot (10; 3000, 0.005)$$

$$\text{Need to solve:} \qquad F(x; \lambda) = \sum_{y=0}^{x} \frac{\lambda^y \cdot e^{-\lambda}}{y!}$$

$$1 - (10; 15) = 1 - 0.118 = 0.882$$

Poisson Distribution

For a binomial distribution, when n is large and p is small , we have:

$$\lim_{\substack{n \to 0 \\ p \to 0}} n \cdot p = \lambda$$

where $n \cdot p = \text{constant}$.

Random variable:

$x =$ total number of success, $\lambda =$ parameter, average number of success

$$P(x; \lambda) = \begin{cases} \frac{\lambda^x \cdot e^{-\lambda}}{x!} \text{ for } x = 0, 1, 2, ... \\ 0 \text{ otherwise} \end{cases}$$

$$\lim_{\substack{n \to 0 \\ p \to 0}} \binom{n}{x} p^x \cdot (1-p)^x = \qquad \frac{n \cdot p^x \cdot e^{-(n \cdot p)}}{x!} \qquad = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

$$E(X) = \lambda$$

$$\mathrm{Var}(x) = \lambda$$

Example. The average number of trucks arriving on any given day at a truck depot is 12. What is the probability that on a given day, fewer than 9 trucks arrive at this depot?

Solution. $\lambda =$ average of number of successes = a new arrival; $x =$ number of successes = number of arrivals on a given day

$$\begin{aligned} P(x<9) &= P(x \leq 8) = \sum_{y=0}^{8} \frac{12^y \cdot e^{-y}}{y!} = F(8;12) \\ &\approx \frac{F(8;10) \cdot F(8;15)}{2} \\ &\approx \frac{0.37}{2} \\ &\approx 0.1535 \end{aligned}$$

Poisson Process

A collection of Poisson random variables that are indexed by non-negative integers, say t.

Example. $$x(0), x(1), x(2), ..., x(t), ..., x(100), ...$$

Notation. *X(t) where* $t = 0, 1, 2, ...$

1. *X(0)=*
2. *The process has independent increments*
3. *The number of successes in any interval of length t is Poisson distribution with mean,* $\alpha \cdot t$
   - α *is called the rate, per unit time or, per unit region*

$$P(X(t) = x) = \frac{e^{-\alpha \cdot t} \cdot (\alpha \cdot t)^x}{x!}$$

Example. A certain type of fabric has 2 defects per 10 square yards. If one assumes the number of defects follows a Poisson distribution, what is the probability that 30 square yards of the fabric will have 4 or more defects?

Solution. $X(t) =$ number of defects in a 30 yard square bolt, where $t = 0, 1, 2, 3$

Unit area: 10 yd^2, $\alpha = 2$, $t = 3$

$$\begin{aligned} &P(X(t);3) = \frac{e^{-6} \cdot 6^x}{x!} \\ &P(x>4) = 1 - P(x \leq 4) = 1 - F(4;6) \end{aligned}$$

4 Continuous Random Variables and Probability Distributions

Probability Density Function

Definition. *A function with value $f(x)$, defined over the set of all real number is called a probability density function of a continuous random variable, x, if and only if:*

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

for any real constants a and b, where $a \leq b$.

Note. $P(a \leq x < b) = P(a < x < b) = P(a < x \leq b) = P(a \leq x \leq b)$

$P(x = a) = P(a \leq x \leq a) = \int_a^a f(x)\,\mathrm{dx} = 0$

Properties of *PDF*

1. $f(x) \geq 0 \ \forall \ -\infty < x < \infty$
2. $\int_{-\infty}^{\infty} f(x)\,\mathrm{dx} = 1$

Example. Consider a function

$$f(x) = \begin{cases} k \cdot e^{-3x} \text{ for } x > 0 \\ 0 \text{ otherwise} \end{cases}$$

a) Find R such that $f(x)$ is a valid pdf.

1. $f(x) = k \cdot e^{-3x} > 0$ iff $k > 0$
2. $\int_{-\infty}^{\infty} f(x)\,\mathrm{dx} = \int_{-\infty}^{\infty} k \cdot e^{-3x}\,\mathrm{dx} = 1$ for a valid pdf.

$$\therefore k = \frac{1}{\int_{-\infty}^{\infty} e^{-3x}\,\mathrm{dx}}$$

$$\int_{-\infty}^{\infty} e^{-3x}\,\mathrm{dx} = \int_{-\infty}^{0} 0\,\mathrm{dx} + \int_{0}^{\infty} e^{-3x}\,\mathrm{dx} = \frac{-e^{-3x}}{3} = \lim_{t \to \infty} \frac{-e^{-3t}}{3} + \frac{e^{-3 \times 0}}{3} = \frac{1}{3}$$

$$\therefore f(x) = \begin{cases} 3e^{-3x} \text{ if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

b)

$$P(0.5 \leq x \leq 1)$$
$$= \int_{0.5}^{1} f(x)\,\mathrm{dx}$$
$$= \int_{0.5}^{1} 3e^{-3x}\,\mathrm{dx}$$

Continuous Random Variables

Interested in $P(a \leq x \leq b)$. Define pdf: $P(a \leq x \leq b) = \int_b^a f(x)\,\mathrm{dx}$

Properties of cdf: $\to F(x) = P(X \leq x)$

1. $P(a \leq x \leq b) = F(b) - F(a) = \int_b^a f(x)\,\mathrm{dx}$
2. $F(x) = \int_{-\infty}^{\infty} f(t)\,\mathrm{dt}$, $f(x) = \frac{\delta F(x)}{\delta x}$

Derivative of cdf from pdf

1. Divide the real line according to the support of x (domain of x)

2. Derive the cdf for eah regment of the real line

Example.

$$f(x) = \begin{cases} 3e^{-3x}, 0 \leq x, \text{find cdf} \\ 0\,\text{otherwise} \end{cases}$$

$$\begin{aligned}
x < 0 \quad F(x) &= \int_{-\infty}^{x} f(t)\,\mathrm{dt} = \int_{-\infty}^{x} 0\,\mathrm{dt} = 0 \\
x \geq 0 \quad F(x) &= \int_{-\infty}^{x} f(t)\,\mathrm{dt} = \int_{-\infty}^{0} f(t)\,\mathrm{dt} + \int_{0}^{x} f(t)\,\mathrm{dt} \\
&= \int_{-\infty}^{0} 0\,\mathrm{dt} + \int_{0}^{x} 3e^{-3t}\,\mathrm{dt} \\
&= 1 - e^{-3x}
\end{aligned}$$

$$F(x) = \begin{cases} 3e^{-3x}, x \geq 0 \\ 0, x < 0 \end{cases}$$

Percentile of Distribution

Median$=m$ s.t $P(X \leq m) = \frac{1}{2}$

Definition. $\alpha = F(\delta(\alpha)) = \int_{-\infty}^{\delta(\alpha)} f(t)\,\mathrm{dt}$

where $\delta(\alpha)$ is the $(100\alpha)^{\text{th}}$ percentile.

Example.

a) $\alpha = \frac{1}{2}$ *such that* $\delta\left(\frac{1}{2}\right) = \text{median}$

b) $\frac{1}{2} = F\left(\delta\left(\frac{1}{2}\right)\right)$, $\delta\left(\frac{1}{2}\right) = ? = m$

$$\begin{aligned}
F(x) &= 1 - e^{-3x}\,\text{where}\,x \geq 0 \\
\therefore\; F(m) &= 1 - e^{-3m} = \frac{1}{2} \\
\Rightarrow\; e^{-3m} &= \frac{1}{2} \Rightarrow \text{solve for}\,m \\
\ln(e^{-3m}) &= \ln\left(\frac{1}{2}\right) \\
-3m &= -\ln(2) \\
\therefore\; m &= \frac{\ln(2)}{3}
\end{aligned}$$

Expectation: $E(g(x)) = \int_{-\infty}^{\infty} g(x) \cdot f(x)\,\mathrm{dx}$

1.

$$\begin{aligned}
\mu &= E(X) &&= \int x \cdot f(x)\,\mathrm{dx} \\
\mu_2 &= E(X^2) &&= \int x^2 \cdot f(x)\,\mathrm{dx} \\
\sigma^2 &= \mathrm{Var}(x) &&= \int_{\infty}^{\infty} (x - \mu)^2 \cdot f(x)\,\mathrm{dx} \\
& &&= E(X^2) - E(X)^2 \\
& &&= \mu_2 - \mu^2 \\
\sigma &= &&\sqrt{\sigma^2}
\end{aligned}$$

Example. $f(x) = \frac{4}{\pi \cdot (1 + x^2)}$ for $0 < x < 1$ and $\int f(x)\,\mathrm{dx} = 1$

We need to find $E(X)$ and $\mathrm{Var}(x)$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{dx} \\ &= \int_0^1 x \cdot f(x)\,\mathrm{dx} \\ &= \int_0^1 \frac{4x}{\pi(1+x^2)}\,\mathrm{dx} \end{aligned}$$

Integration by substitution, $\frac{\ln(4)}{\pi}$

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \cdot f(x)\,\mathrm{dx} \\ &= \int_0^1 \frac{4x^2}{\pi(1+x^2)}\,\mathrm{dx} \\ &= \frac{4}{\pi} \cdot \int_0^1 \frac{1+x^2-1}{1+x^2}\,\mathrm{dx} \\ &= \frac{4}{\pi} \cdot \int_0^1 \left(1 - \frac{1}{1+x^2}\right)\mathrm{dx} \\ &= \frac{4}{\pi} \cdot \int_0^1 1\,\mathrm{dx} - \int_0^1 \frac{4}{\pi(1+x^2)}\,\mathrm{dx} \\ &= \frac{4}{\pi} - 1 \end{aligned}$$

Uniform Distribution

Definition. *x follows a uniform distribution between θ_1 and θ_2*

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

CDF:

1. $x \leq \theta_1$:

$$F(x) = \int_{-x}^{x} f(t)\,\mathrm{dt} = \int_{-x}^{x} 0\,\mathrm{dt} = 0$$

2. $\theta_1 < x < \theta_2$:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{dt} = \int_{-\infty}^{\theta_1} 0\,\mathrm{dt} + \int_{\theta_1}^{x} \frac{1}{\theta_2 - \theta_1}\,\mathrm{dt} = \frac{x - \theta_1}{\theta_2 - \theta_1}$$

3. $\theta_2 < x$:

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{dt} = \int_{-\infty}^{\theta_1} 0\,\mathrm{dt} + \int_{\theta_1}^{\theta_2} \frac{1}{\theta_2 - \theta_1}\,\mathrm{dt} + \int_{\theta_2}^{x} 0\,\mathrm{dt} = 1$$

$$F(x) = \begin{cases} 0, \theta_1 \leq x \\ \frac{x - \theta_1}{\theta_2 - \theta_1}, \theta_1 < x < \theta_2 \\ 1, x \geq \theta_2 \end{cases}$$

With $E(X) = \frac{\theta_2 + \theta_1}{2}$ and $\mathrm{Var}(x) = \frac{1}{12}(\theta_2 - \theta_1)^2$

Normal Distribution: ($X \sim$ Normal(μ, σ^2))

Probably the most common form of dist.

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

Where $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$

$f(\mu - a) = f(\mu + a)$ where $a > 0$

Normal dist. also has symmetric dist.

$$\begin{aligned} f(x) &= f(-x) \\ A = B \Rightarrow \ P(X \le -x) &= P(X \ge x) \end{aligned}$$

CDF:

$$\begin{aligned} F(x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2x\sigma^2}} e^{-\frac{1}{2\sigma^2}(t-\mu)^2} \mathrm{dt} \\ &= \Phi(x; \mu, \sigma^2) \end{aligned}$$

Standard Normal: ($\mu = 0, \sigma^2 = 1$)

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2x\sigma}} e^{\frac{-x^2}{2}} \\ F(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2x\sigma}} e^{\frac{-t^2}{2}} \mathrm{dt} \Rightarrow \text{Table } A.3 \end{aligned}$$

Notation. $X \sim N(0, 1)$

Example. Find the probability that z is between 0.87 and 1.28 given $z \sim N(10, 1)$.

$$\begin{aligned} P(0.87 < z < 1.28) &= F(1.28) - F(0.87) \\ &= 0.8997 - 0.8087 \\ &= 0.0919 \end{aligned}$$

$N(\mu, \sigma^2)$ vs. $N(0, 1)$:

1. If $x \sim N(\mu, \sigma^2) \longrightarrow z = \frac{x-\mu}{\sigma}$, then $z \sim N(0, 1) \Longrightarrow$ standardization

Note. $X \sim N(\mu, \sigma^2)$

$$\begin{aligned} P(a \le x \le b) &= P\left(\frac{a-\mu}{\sigma} < z < \frac{b-\mu}{\sigma}\right) \\ F_x(b) - F_x(a) &= F_z\left(\frac{b-\mu}{\sigma}\right) - F_z\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

Example. Suppose that the amount of cosmic radiation to which a person is exposed is $X \sim N(4.35, 0.59^2)$. What is the probability that a person will be exposed to more than 5.20?

$X \sim N(4.35, 0.59^2),\ \mu = 4.35$ and $\sigma^2 = 0.59^2$

$$\begin{aligned}
P(X \geq 5.20) &= P\left(\frac{X-\mu}{\sigma} \geq \frac{5.20-\mu}{\sigma}\right) \\
&= P\left(z \geq \frac{5.20-4.35}{0.59}\right) \\
&= P(z \geq 144) \\
&= 1 - P(z < 144) \\
&= 1 - F(1.44) \\
&= 1 - 0.9251 \\
&= 0.0749
\end{aligned}$$

Example. Fish were studied to exaime the level of mercury contamination, Y, which varies according to a normal distribution with mean 18 and variance 16. $Y \sim N(18, 16)$, $\mu = 18$, $\sigma^2 = 4^2$

a) What proportion of contamination levels are between 11 and 21?

$$\begin{aligned}
P(11 < Y < 21) &= P\left(\frac{11-18}{4} < \frac{Y-18}{4} < \frac{21-18}{4}\right) \\
&= P(-1.75 < z < 0.75) \\
&= F_z(0.75) - F_z(-1.75) \\
&= 0.7734 - 0.0401 \\
&= 0.7333
\end{aligned}$$

b) 90% of all contamination levels are above what mercury level?

$P(Y \geq q) = 0.9$ and $q = ?$. Note that if $P(Y \geq q) = 0.9$, then $P(Y < q) = 0.1$

$\therefore q$ is the 10^{th} percentile of Y.

$$\begin{aligned}
P(Y < q) &= P\left(\frac{Y-18}{4} < \frac{q-18}{4}\right) \\
&= P\left(z < \frac{q-18}{4}\right)
\end{aligned}$$

$\therefore\ P(Y < q) = 0.1$ implies $P\left(z < \frac{q-18}{4}\right) = 0.1$ where $z \sim N(0,1)$

$\therefore\ \frac{q-18}{4}$ is the 10^{th} of z or $F_z\left(\frac{q-18}{4}\right) = 0.1$

From the table A3, we see that the value that corresponds to $F(x) = 0.1$ is -1.28

OR: $\frac{q-18}{4} = -1.28$ OR $q = 12.88$

$$\therefore P(Y \geq 12.88) = 0.9$$

Binomial Distribution vs. Normal Distribution

As $n \to \infty$, the Binomial distribution is approximately "bell" shape.

Theorem. *If* $X \sim \text{Binomial}(n, p)$*, then* $X \sim N(np, np(1-p))$ *as* $n \to \infty$.

Remark. *In Poisson,* $n \to \infty$ *and* $p \to 0$, $\therefore\ n \cdot p \to$ *constant*

Example. Finding the probability of getting 6 heads and 10 tails in 16 tosses of a coin.

1. pmf $\quad x=6,\ n=16,\ p=\frac{1}{2}$

$$P(x=6)=b\left(6;16,\frac{1}{2}\right)=\binom{16}{6}\left(\frac{1}{2}\right)^6\left(1-\frac{1}{2}\right)^{10}=0.1222$$

2. Finding $P(x=6)$ using the normal approximation.

$$\begin{aligned}
P(x=6)\approx P(5.5<x<6.5) &= P(x\leq 6.5)-P(x\leq 5.5)\\
x\sim N(np,np(1-p)) &= N(8,4)\\
P(x\geq 6.5) &= P\left(\frac{x-np}{\sqrt{np(1-p)}}\leq\frac{6-np}{\sqrt{np(1-p)}}\right)\\
&= P\left(z\leq\frac{6.5-8}{\sqrt{4}}\right)\\
&= F_z(-0.75)\\
&= 0.3944\\
P(x\leq 5.5) &= F_z(-0.75)=0.2734
\end{aligned}$$

$$\therefore P(x=6)\doteq F_z\,(-1.25)-F_z(-0.75)=0.121$$

Gamma Family Distributions

Notation.
$$f(x;\alpha,\beta)=\begin{cases}\frac{1}{\beta^\alpha P(\alpha)}x^{\alpha-1}e^{\frac{-x}{\beta}} & \text{for } x>0\\ 0\,\text{otherwise}\end{cases}$$

where $\alpha>0,\ \beta>0$

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

If $\alpha=$ *an integer,* $P(\alpha)=(\alpha-1)!$

Special Cases:

1. $\beta=1$, *standard gamma*

2. $\alpha=1$, *exponential*

3. $\alpha=\frac{v}{2}$, $\beta=2$, *chi-squared distribution where* $v\in\mathbb{Z}$

5 Joint Probability Distributions

Random vector: $(x_1,\ldots,x_n)$ where x_i's are random variables.

Divariate case: (X,Y)

Discrete Case: X and Y are d.r.v's

$$P(X=x,Y=y)=P(\{X=x\}\cap\{Y=y\})$$

Joint pmf: $P(X,Y)$

Properties of Joint pmf:

1. $P(x,y) > 0$

2. $\sum_{x \in R_x} \sum_{y \in R_y} P(x,y) = 1$ where R_x and R_y are supports for x and y, respectively

Result. $(X,Y) \in B$, where B is in a subset of $(R_X \cap R_Y)$

$$P((X,Y) \in B) = \sum_{(x,y) \in B} P(x,y)$$

Example. An insurance company determines the annual number of tomatoes in Waterloo and Oxford counties. $X =$ the number of tomatoes in Waterloo and $Y =$ the number of tomatoes in Oxford counties. [A bunch of numbers]

$P(X,Y)$	$y=0$	$y=1$	$y=2$	$y=3$	$P_x(x)$
$x=0$	0.12	0.06	0.05	0.02	$P_x(0) = 0.25$
$x=1$	0.13	0.15	0.12	0.03	$P_x(1) = 0.43$
$x=2$					$P_x(2) = 0.32$
$P_Y(y)$	0.3	0.36	0.27	0.07	1

Table 3.

a) What is the probability that there is no more than tomato in both counties

$$\begin{aligned} P(X+Y \leq 1) = P((X,Y) \in B) \text{ where } B &= \{(X,Y): X+Y \leq 1\} \\ &= \sum_{(x,y) \in B} P(x,y) \\ &= P(0,0) + P(0,1) + P(1,0) \\ &= 0.12 + 0.13 + 0.06 \\ &= 0.31 \end{aligned}$$

Marginal Distribution

$$\begin{aligned} P(X=x) = \ P_X(x) \ &= \sum_y P(x,y) \\ P(Y=y) = \ P_Y(y) \ &= \sum_x P(x,y) \end{aligned}$$

Marginal pmf is obtained by summing $P(x,y)$ over the other variable.

Conditional Expectation (DRV)

$$\begin{aligned} E(X|Y=y) &= \sum_{x \in R_X} x \cdot P(X|Y) \\ E(g(X)|Y=y) &= \sum_{x \in R_X} g(x) \cdot P(x|y) = h(y) \\ E(g(Y)|X=x) &= \sum_{y \in R_Y} g(y) \cdot P(y|x) = q(x) \\ \mathrm{Var}(X|Y=y) &= E(X^2|Y=y) - E(X|Y=y)^2 \end{aligned}$$

Joint Distribution of CRV

Univariate function:

$$y = f(x) \longrightarrow \int_{x \in D} f(x)\,\mathrm{dx}$$

Bivariate function:

$$z = f(x, y) \longrightarrow \int_{y \in D_Y} \int_{x \in D_X} f(x, y)\,\mathrm{dx}\,\mathrm{dy} = \int_{x \in D_X} \int_{y \in D_Y} f(x, y)\,\mathrm{dy}\,\mathrm{dx}$$

Example. $g(x, y) = x^2 y$ where $0 < x < 1$ and $-3 < y < 3$

$$\int_{y=-3}^{y=3} \int_{x=0}^{x=1} x^2 y\,\mathrm{dx}\,\mathrm{dy} = \int_{-3}^{3} y\frac{x^3}{3}|_0^1\,\mathrm{dy} = \int_{-3}^{3} \frac{y}{3}\,\mathrm{dy} = \frac{1}{2x3} x y^2|_{-3}^3 = 0$$

Consider continuous r.v X and Y

$$\begin{aligned} P(a \leq X \leq b, c \leq Y \leq d) &= P((X, Y \in B^2) \\ &= \int_c^d \int_a^b f(x, y)\,\mathrm{dx}\,\mathrm{dy} \\ \text{where } B^2 &= \{(x, y)\colon a \leq x \leq b, c \leq y \leq d\} \end{aligned}$$

Properties of $f(x, y)$:

1. $f(x, y) \geq 0$
2. $\int\limits_{y \in R_Y} \int\limits_{x \in R_X} f(x, y)\,\mathrm{dx}\,\mathrm{dy} = 1$

Definition. ***Joint cdf***: $F(X, Y) = P(X \leq x, Y \leq y) = \int\limits_{-\infty}^{y} \int\limits_{-\infty}^{x} f(t, s)\,\mathrm{dt}\,\mathrm{ds}$

Example. A fast good restaurant has dine-in and takeout. On a randomly selected day, let:$X =$ the proportion of drive-thru service and $Y =$ proportion of dine-in service.

Suppose the pdf is:

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) \text{ if } 0 < x < 1, 0 < y < 1 \\ 0 \text{ otherwise} \end{cases}$$

What is the probability that neither serveices are busy more than $\frac{1}{4}$ of the day?

Solution. $P\left(X \leq \frac{1}{4}, Y \leq \frac{1}{4}\right) = ?$

$$\begin{aligned} R^2 &= \{(x, y)\colon 0 < x < 1, 0 < y < 1\} \\ P\left(X \leq \frac{1}{4}, Y \leq \frac{1}{4}\right) &= P\left(0 \leq X \leq \frac{1}{4}, 0 \leq Y \leq \frac{1}{4}\right) \\ &= \int_0^{\frac{1}{4}} \int_0^{\frac{1}{4}} f(x, y)\,\mathrm{dx}\,\mathrm{dy} \\ &= \int_0^{\frac{1}{4}} \int_0^{\frac{1}{4}} \frac{6}{5}(x + y^2)\,\mathrm{dy}\,\mathrm{dx} \\ &= \int_0^{\frac{1}{4}} \frac{6}{5}\left(\frac{x}{4} \ldots\right) \end{aligned}$$

Marginal Distribution

$$\begin{aligned} f_X(x) &= \int\limits_{y \in R_Y} f(x, y)\,\mathrm{dy} \\ f_Y(y) &= \int\limits_{x \in R_X} f(x, y)\,\mathrm{dx} \end{aligned}$$

Example.

$$\begin{aligned} f_Y(y) &= \int\limits_0^1 f(x, y)\,\mathrm{dx} = \int\limits_0^1 \frac{6}{5}(x + y^2)\,\mathrm{dx} = \frac{6}{5}\left(\frac{1}{2} + y^2\right) \\ f_X(x) &= \int\limits_0^1 f(x, y)\,\mathrm{dy} = \int\limits_0^1 \frac{6}{5}(x + y^2)\,\mathrm{dy} = \frac{6}{5}x + \frac{2}{5} \end{aligned}$$

Conditional PDF

$$P(a < x < b, \text{given}\, Y = y) = \int\limits_a^b f(x|y)\,\mathrm{dx}$$

$f(x|y)$ is called the conditional pdf of x given $Y = y$.

$$\begin{aligned} f(x|y) &= \frac{f(x, y)}{f_Y(y)} \\ f(y|x) &= \frac{f(x, y)}{f_X(x)} \end{aligned}$$

Independence:

X and Y are independent iff:

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

Remark. If $x \leftarrow dY$ are independent:

$$f_X(x) = f(x|y)$$

$\Rightarrow$ distribution of y does not connect to distribution of $f_X(x)$

Proof.

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Since X and Y are independent

$$\therefore f(x|y) = \frac{f_X(x) \cdot f_Y(y)}{f_Y(y)} = f_X(x)$$ □

Similarly, if X and Y are independent, $f_Y(y) = f(y|x)$

Example. $f(x,y) = \frac{6}{5}(x+y^2)$

$$\begin{aligned} f_X(x) \cdot f_Y(y) &= \frac{6}{25}(3x+1+6xy^2+2y^2) \\ f(x|y) &= \frac{f(x,y)}{f_Y(y)} = \frac{x+y^2}{\frac{1}{2}+y^2} \\ f_X(x) &= \frac{6}{5}\left(\frac{1}{2}+y^2\right) \end{aligned}$$

$\therefore X$ and Y are not independent.

$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{\frac{6}{5}(x+y^2)}{\frac{6}{5}\left(x+\frac{1}{3}\right)} = f\left(y\middle|\frac{1}{2}\right) = \frac{\left(\frac{1}{2}+y^2\right)}{\frac{5}{8}}$

Expectations

Let $g(X,Y)$ be a function of X and Y.

$$E[g(X,Y)] = \int_{y \in R_Y} \int_{x \in R_X} g(x,y) \cdot f(x,y)\, \mathrm{dx}\, \mathrm{dy}$$

1. $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$, $\mu_X = E(X)$ and $\mu_Y = E(Y)$

   $\Rightarrow$ covariance of X and Y

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

2. $g(X,Y) = XY$

Example. $$f(x,y) = \begin{cases} 4xy \text{ if } 0<x<1, 0<y<1 \\ 0 \text{ otherwise} \end{cases}, \mathrm{Cov}(X,Y) = ?$$

$$\begin{aligned} \mathrm{Cov}(X,Y) &= E(XY) - E(X)E(Y) \\ f_X(x) &= \int_0^1 4xy\, \mathrm{dy} = 2x \\ f_Y(y) &= \int_0^1 4xy\, \mathrm{dx} = 2y \\ E(X) = \mu_X &= \int_0^1 x \cdot f_X(x)\, \mathrm{dx} = \frac{2}{3} \\ E(Y) = \mu_Y &= \int_0^1 y \cdot f_Y(y)\, \mathrm{dy} = \frac{2}{3} \\ E(XY) &= \int_0^1 \int_0^1 xy f(x,y)\, \mathrm{dx}\, \mathrm{dy} = \int_0^1 \int_0^1 4x^2y^2\, \mathrm{dx}\, \mathrm{dy} = \frac{4}{9} \end{aligned}$$

Thus, $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0$

Condition Expectations

$$
\begin{aligned}
E(X|Y=y) &= \int_{x\in R_X} x\cdot f(x|y)\,\mathrm{dx} = h(x) \\
E(Y|X=x) &= \int_{y\in R_Y} y\cdot f(y|x)\,\mathrm{dy} = q(x)
\end{aligned}
$$

$E(X|Y=y)\,\mathrm{vs}\,E(X)$ if X and Y are independent?

Example. $X\sim N(\mu,\sigma^2)$, $X\sim\mathrm{Bin}(n,p)$

$$f(x)=\frac{1}{\sqrt{2\lambda\sigma^2}}e^{\left(-\frac{1}{2\sigma^2}\right)(x-\mu)^2}\,\text{with}\,-\infty<x<\infty$$

$E(X)=\mu$ and $\mathrm{Var}(X)=\sigma^2$

$$p(x)=\binom{n}{x}p^x(1-p)^{n-x}\,\text{for}\,x=1,2,...,n$$

$E(X)=n\cdot p$ and $\mathrm{Var}(X)=n\cdot p(1-p)$

Statistics

- Variable of interest (value of random variable
- Measure of a variable $\Rightarrow$ observations (data)
- Population: Collection of objectives of interest (text book)
- The totality of elements which are under discussion and in formations are desired

  $\Longrightarrow$ distribution function

  $\Longrightarrow$ a set of parameters

Sample (data)

- A <u>subset</u> of populations
- We dserve all the values
- We may not know the value of parameters

Distributive Statistics

- location $\rightarrow$ parameter $\Longleftarrow\begin{cases}1,\text{sample mea}\\ 2\text{sample mean}\end{cases}$
- scale $\rightarrow$ parameter $\Longleftarrow\begin{cases}1,\text{sample variance},\text{sample standard deviation}\\ 2,Q_1,Q_3,\text{min},\text{max}\end{cases}$

Notation. *Suppose $x_1,...,x_n$ are observations and n is called the sample size.*

Sample mean:

$$\bar{x}=\frac{x_1+...+x_n}{n}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

Suppose average:

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

Sample median: $x_1, \ldots n$

- *Step 1: Arrange the observations:* $x_1, \ldots, x_n$ *is small to largest;* $x_{11,} \ldots, x_n$
- *Step 2:*

  Case A: If $n =$ *an odd number*

$$\begin{aligned}\text{Median} &= \frac{(n+1)}{2}^{\text{th}} \\ &= x_{\left(\frac{n+1}{2}\right)}\end{aligned}$$

  Case B: If $n =$ *an even number*

$$\begin{aligned}\text{Median} &= \text{average of the two center observations} \\ &= \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}\end{aligned}$$

Mean and Median Comparison

If mean $=$ median, symmetric distribution

If mean $\neq$ median, anti-symmetric distribution

Data

- Sample mean $= \frac{1}{n} \cdot \sum x_i = \bar{x}$
- Sample variance $= \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 = S^2$
- Standard deviation $= \sqrt{S^2}$
- Median: $\begin{cases} x_{\frac{n}{2}} + x_{\frac{n}{2}x} \text{ if } n \text{ is even} \\ x_{\frac{n+1}{2}} \text{ if } n \text{ is odd} \end{cases}$

Graphic Technique for Numerical Variable

Create a histogram.

Boxplot.

Using sample distribution $(\mu, \sigma^2, \sigma, \text{proportion})$ to calculuate $\bar{X}, S^2, S, \hat{P}$.

Definition. *A random sample,* $X_1, \ldots, X_n$*, is constituted by a set of independent and identically distributed random variables.*

$$\begin{aligned} f(x_1, x_2) &= f_1(x_1) \cdot f_2(x_2) \Leftrightarrow x_1 \text{ and } x_2 \text{ are independent} \\ f(x_1, \ldots, x_n) &= f_1(x_1) \cdot \ldots \cdot f_n(x_n) \Leftrightarrow x_1, \ldots, x_n \text{ are independent} \\ &= \prod_{i=1}^{n} f_i(x_i) \\ & f_1 = f_2 = \ldots = f_n \Rightarrow x_1, \ldots, x_n \text{ are independently distributed} \end{aligned}$$

Notation. *IID: Independent and Identically Distributed*

$$f(x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i)$$

Random samlple $\Leftrightarrow$ *iid random variables*

Definition. *A* <u>*statistic*</u> *is a function of a random sample.*

- *A statistic is a random variable*
- *The distribution of a statistic is called a sampling distribution*

Sampling Distribution of Sample Mean

[Linear functions of random variables] Suppose x_1, ..., x_n are random variables and $E(x_i) = \mu_i$, $\mathrm{Var}(x_i) = \sigma_j^2$ and $\mathrm{Cov}(x_i, x_j) = \sigma_{ij}$.

$$U = \sum_{i=1}^{n} a_i x_i$$

is a linear combination of x_i's.

Properties of *U*:

1. $E(U) = E\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} E(a_i x_i) = \sum_{i=1}^{n} a_i \cdot E(x_i) = \sum_{i=1}^{n} a_i \mu_i$
2. $\mathrm{Var}(U) = \mathrm{Var}\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} \mathrm{Var}(a_i x_i) + \sum_{i \neq j}\sum \mathrm{Cov}(a_i x_i,\ a_j x_j) = \sum_{i=1}^{n} a_i \mathrm{Var}(x_i) + \sum_{i \neq j}\sum a_i a_j \mathrm{Cov}(x_i, x_j)$

Statistic:

- *Random variable* - function of a st of a random variables
- *Sample* - $\subseteq$ population $(F, \theta_1, ..., \theta_j)$
- $\hat{\theta} =$ statistic $= T(X_1, ..., X_n)$, $X_1, ..., X_n$ is a set of sample$\sim F$

Sampling Distribution of Sample Mean

- $\hat{\theta} = \bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$, where $X_1, ..., X_n$, $E(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$, $\mathrm{Cov}(X_i, Y_j) = 0$
- $E(\bar{X}) = \frac{1}{n} \cdot \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot \sum_{i=1}^{n} \mu = \frac{1}{n} \cdot n \cdot \mu$
- $\mathrm{Var}(\bar{X}) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{n^2}\left[\mathrm{Var}\left(\sum_{i=1}^{n} x_i\right) + \sum_{i \neq j}\sum \mathrm{Cov}(X_i,\ Y_j)\right] = \frac{1}{n^2}\left[\sum_{i=1}^{n} \sigma^2 + \sum\sum 0\right] = \frac{1}{n^2} \times n \cdot \sigma^2 = \frac{1}{n}\sigma^2$
- $\bar{X} \sim \left(\mu, \frac{\sigma^2}{n}\right) = (E(\bar{X}), \mathrm{Var}(\bar{X}))$

Results:

1. If $X_1, ..., X_n \sim \text{Normal}(\mu, \sigma^2)$, iid $\longrightarrow \bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$

2. If $X_1, ..., X_n \sim ?(\mu, \sigma^2)$, iid

   $\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$ as $n \to \infty$

Theorem. ***Central Limit Theorem***

Let $X_{1,}..., X_n$ be a random sample from a population with mean, μ, and finite variance, σ^2. Then:

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

is a random variable whose distribution approaches to standard normal, $N(0, 1)$ as $n \to \infty$.

Example. A soft-drink vending machine is set so that the amount of drink dispensed is a random variable with $\mu = 200$mL and $\sigma = 15$mL. What is the probability that the average amount dispensed in a random variable of size 36 is at least 200mL.

Solution.

Given a random variable, $X_i =$ amount of drink

We have $\mu = 200$, $\sigma = 15$, $n = 36$. $P(\bar{X} \geq 204) = ?$

Using CLT, we know

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

as n is large.

$$\begin{aligned} P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{204 - 200}{\frac{15}{\sqrt{36}}}\right) &= P(z \geq 1.6) \\ &= 1 - P(z \geq 1.6) \end{aligned}$$

Remark.

1. We will not be able to answer this question if n is small, say 5.

2. If $X_i \sim \text{Normal}(\mu, \sigma^2)$, then we can calculate the probability even if n is small

Sampling Distribution of S^2

Theorem. *If $(X_1, ..., X_n) \sim N(\mu, \sigma^2)$ and iid, then*

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

Recall. $\text{Gamma}(\alpha, \beta)$

$$f_Y(y) = \begin{cases} \frac{1}{\beta^\alpha P(\alpha)} y^{\alpha - 1} e^{\frac{-\alpha}{\beta}}, \text{for } y > 0 \\ 0 \text{ otherwise} \end{cases}$$

When $\beta = 2$, $\alpha = \frac{v}{2} \longrightarrow Y \sim \chi(v)$, v is called the degree of freedom

$E(Y) = v and \ \mathrm{Var}(Y) = 2v$

$$\therefore E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1 \Longrightarrow \frac{(n-1)}{\sigma^2} E(S^2) = n-1$$

$$\therefore E(S^2) = n-1$$

Point Estimation

We use the value of a statistic to estimate a population parameter (μ, σ^2, proportion, ...).

Definition. *An <u>estimator</u> is a statistic, who value is used to estimate a parameter $\Rightarrow$ random variable.*

Definition. *An <u>estimate</u> is the value of an estimator.*

Notation.

- *Denote an estimator as, $\hat{\theta}$ (a random variable)*

Evaluation of an Estimator:

1. Consistency: $\mathcal{L}_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| < \epsilon\right) = 1$

2. Unbiasness: $E\left(\hat{\theta}\right) = \theta$

3. Small variance (efficiency)

Example. $E(\bar{X}) = \mu$, $E(S^2) = \sigma^2$ (necessary to be iid $N(\mu, \sigma^2)$)

Example. Let T_1 and T_2 be to unbiased estimator, then T_1 is more efficient if:

$$\mathrm{Var}(T_1) < \mathrm{Var}(T_2)$$

$X_1, ..., X_n \sim N(\mu, \sigma^2)$

An estimator of $\mu = \bar{X} = T_1$. Another estimator of $\mu = X_5 = T_2$

$$\begin{array}{rcl} E(\bar{X}) & \rightarrow & \mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n} \\ E(X_5) & \rightarrow & \mathrm{Var}(X_5) = \sigma^2 \\ & \therefore & \dfrac{\sigma^2}{n} \leq \sigma^2 \\ & \therefore & \bar{X}\,\text{is more efficient} \end{array}$$

Point Estimation

- Parameters of interest: $E(X) = \mu$, $\mathrm{Var}(X) = \sigma^2$, proportion of a group $X = \left\{ \begin{smallmatrix} 1 w.p \pi \\ 0 w.p 1-\pi \end{smallmatrix} \right.$

  Notation: θ

- Point estimators: Notation: $\hat{\theta}$, $\hat{\theta} = \bar{X} = E(X)$, $\hat{\theta} = S^2 = \mathrm{Var}(X) \Longrightarrow \hat{\theta}$ is a random variable, sample proportion $\hat{\theta} = p$

Evaluation of the Quality of $\hat{\theta}$

1. Unbiasness: $E(\hat{\theta}) = \theta$

2. Efficiency: Given $E(\hat{\theta}) = \theta$, variance of $\hat{\theta}$ is the smallest

$$E[(\hat{\theta} - \theta)^2] = \text{mean square error of } \hat{\theta} = \text{MSE}(\hat{\theta})$$

- If (mean square error) $\text{MSE}(\hat{\theta})$ is small, then the estimator is good

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2] = E[(A + B)^2]$$

- $\text{Var}(\hat{\theta})$ measures the precision of $\hat{\theta}$ (spread from centre of distribution)
- $\text{bias}(\hat{\theta})$ measures the accuracy of $\hat{\theta}$ (distance from centre of true value)
- Using $\text{MSE}(\hat{\theta})$, we observe that there is a trade off between accuracy and precision
- We impose the assumption (bias), to solve the trade off problem assumptio: $E(\hat{\theta}) = \theta$, $\text{bias}(\hat{\theta}) = 0$

Estimation of Population Mean

- Random sample: $X_1, ..., X_n \sim F(x, \mu, \sigma^2)$ or $\sim(\mu, \sigma^2)$; $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$ where $X_1, ...,$ X_n is independent
- Parameter of interest: $\mu = E(X)$
- Estimator: $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

  Estimates: $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

Theorem. ***Law of Large Numbers*** *(consistency of $\bar{X}$)*

$$\mathcal{L}_{n \to \infty} P(|\bar{X} - \mu| < \epsilon) = 1; \bar{X} \to \mu \text{ in probability}$$

- Unbiasness: $E(\bar{X}) = \mu$
- Efficiency: $\bar{X}$ is efficient, $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$

Distribution of $\bar{X}$: If <u>σ^2 is known</u>

1. $X_1, ... X_n \sim N(\mu, \sigma^2)$, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ [exact]

2. $X_1, ..., X_n \sim (\mu, \sigma^2)$, Central Limit Theorem: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ [approximate] only if n is large

Distribution of $\bar{X}$: If <u>σ^2 is unknown</u> $(X_1, ..., X_n \sim (\mu, \sigma^2))$

$\Rightarrow$ Assume $X_1, ..., X_n \sim N(\mu, \sigma^2)$ $\qquad$ $\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \longrightarrow = \widehat{\text{Var}(\bar{X})} = \frac{S^2}{\sqrt{n}}$

Then:

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

estimates σ^2.

$$\begin{aligned} &\Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \\ t &= \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{\bar{X}-\mu}{\sqrt{\widehat{\mathrm{Var}(\bar{X})}}} \end{aligned}$$

We have :

$$t = \frac{\bar{X}-\mu}{\sqrt{\frac{S^2}{n}}} \sim t(n-1)$$

- $t(n-1)$ denotes a $t-$distribution with degree of free $n-1$ with degree of freedom $n-1$
- $t-$distribtion: continuous, symmetric (bell-shaped), longer tail than normal distribution

Hypothesis Testing:

1. Hypothesis: $H_0 \cup H_1$
2. Calculate test statistics: $\hat{\theta}$, under $H_\circ$
3. Critical value
4. Rejection rules
5. Answer the question

Hypothesis Testing for μ (σ is known):

- $X_1, ..., X_n \sim N(\mu, \sigma^2)$ or n is large
- $H_\circ$: $\mu = \mu_\circ$

Alternative Hypothesis	*Test Statistic*	*Critical Value*
H_1^1: $\mu > \mu_\circ$		z_d (right end)
	$z = \frac{\bar{X}-\mu_\circ}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$	
H_1^2: $\mu < \mu_\circ$		$-z_d$ (left end)
	$z = \frac{\bar{x}-\mu_\circ}{\frac{\sigma}{\sqrt{n}}}$	
H_1^3: $\mu \pm \mu_\circ$		$z_{\frac{\alpha}{2}}$ (both ends)

$P-$ Value	*Rules of Rejection*
$P(z > z_\circ)$	reject $H_\circ$ if $z_\circ > z_\alpha$
$P(z < z_\circ)$	reject $H_\circ$ if $z_\circ < z_\alpha$
$2 \times P(z < \lvert z_\circ \rvert)$	reject $H_\circ$ if $\left\lvert z_\circ < z_{\frac{\alpha}{2}} \right\rvert$
$z \sim N(0,1)$	For al the cases, reject $H_\circ$ if $P-$ value $< \alpha$

Inference For Two Populations (means, proportions)

- estimation $\begin{cases} \text{point estimation} \\ \text{interval estimation} \end{cases}$

- hypothesis tests

Properties of $\bar{X}-\bar{Y}$:

1. $E(\bar{X}-\bar{Y})=E(\bar{X})-E(\bar{Y})=\mu_1-\mu_2$ $\qquad \therefore \bar{X}-\bar{Y}$ is unbiased

2. $\mathrm{Var}(\bar{X}-\bar{Y})=\mathrm{Var}(\bar{X})+\mathrm{Var}(\bar{Y})-2\cdot\mathrm{Cov}(\bar{X},\bar{Y})$

Assumptions:

- $X_1,...,X_n$ and $Y_1,...,Y_n$ are two independent samples

  $\bar{X}$ and $\bar{Y}$ are independent

  $\therefore \mathrm{Cov}(\bar{X},\bar{Y})=0$ $\qquad \therefore \mathrm{Var}(\bar{X},\bar{Y})=\mathrm{Var}(\bar{X})+\mathrm{Var}(\bar{Y})$

Hypothesis Testing for: $\mu_1-\mu_2$

- $(X_1,...,X_n)$ and $(Y_1,...,Y_n)$ are two sets of independent samples;

$$x_i\sim(\mu_1,\sigma^2) \qquad y_i\sim(\mu_2,\sigma^2)$$

- Hypotheses: H_0: $\mu_1-\mu_2=\delta_0$

Alternative H_1	*Assumptions*	*Test Statistic (z-test)*
H_1: $\mu_1-\mu_2>\delta_0$	$n_1>30, n_2>30$	$z_\circ=\dfrac{(\bar{X}-\bar{Y})-\delta_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$
H_1^2: $\mu_1-\mu_2<\delta_0$	σ_1^2,σ_2^2 are known	
H_1^3: $\mu_1-\mu_2\neq\delta_0$	σ_1^2,σ_2^2 are known	$z_\circ=\dfrac{(\bar{X}-\bar{Y})-\delta_0}{\sqrt{\frac{S_1^2}{n_1}}+\sqrt{\frac{S_2^2}{n_2}}}$

Critical Value	*P-Value*	*Rules of Rejection*
z_α	$P(z\geq z_0)$	reject H_0 if $z_0>z_\alpha$
$-z_\alpha$	$P(z<z_0)$	reject H_0 if $z_0<z_\alpha$
$z_{\frac{\alpha}{2}}$	$2P(z\geq\lvert z_0\rvert)$	reject H_0 if $\lvert z_0\rvert>z_{\frac{\alpha}{2}}$

Parameters of Interest: $\mu_1-\mu_2$

$$\bar{X}-\bar{Y}=\frac{1}{n}\sum x_i-\frac{1}{n}\sum y_i=\frac{1}{n}\sum_{i=1}^{n}(x_i-y_i)$$

where $D_i=x_i-y_i\Longrightarrow\bar{D}=\frac{1}{n}\sum_{i=1}^{n}D_i$

Confidence Interval for $\mu_1-\mu_2$ in a match design:

$\bar{d}\pm t_{n-1,\frac{\alpha}{2}}\frac{S_d}{\sqrt{n}}$ where $S_d^2=\frac{1}{n-1}\sum_{i=1}^{n}(D_i-\bar{D})^2$