Introduction to Probability and Statistics

Subhayan De

Fall 2017

Contents

Chapter 1

Descriptive Statistics

Statistical analysis learns from data.

Consider the following dataset obtained after testing 10 beams at the lab:

Load in kN	
First crack load	Failure load
26.75	42.25
41.35	41.35
37.75	41.35
28.90	42.25
47.15	47.15
26.75	44.95
42.25	42.25
28.90	45.85
46.00	46.00
28.90	42.25

1.1 Numerical Summaries

n observed values are $x_1, x_2, \ldots, x_n$.

- The sample mean

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

  - The sample mean of the first crack load is $\frac{1}{10} \times 354.7 = 35.47$ kN.
  - The sample mean of the failure load is $\frac{1}{10} \times 435.65 = 43.565$ kN.

- The sample median
  Order the observed values x_i. If n is odd then the median is $(n+1)/2$th value. If n is even the median is the average of values at $n/2$ and $n/2+1$th places.

  - The sample median of the first crack load is 33.325 kN.

- The sample median of the failure load is 42.25 kN.

- The sample mode
  most frequently occurring value(s)
  - The sample mode of the first crack load is 28.90 kN.
  - The sample modal value of the failure load is 43.565 kN.

- The sample variance

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

- Unbiased estimator of variance

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

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

- The sample standard deviation

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

  - The sample variance of the failure load is 4.265 kN^2.
  - The sample standard deviation of the failure load is 2.0652 kN.

- The sample coefficient of variation (COV)

$$v = \frac{s}{\bar{x}}$$

  - The sample coefficient of variation (COV) of failure load is 0.0474.

Data observed in pairs

Two sets of data $\{x_i\}_{i=1}^{n}$ and $\{y_i\}_{i=1}^{n}$.

- The sample covariance

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

- The sample correlation coefficient

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_X}\right)\left(\frac{y_i - \bar{y}}{s_Y}\right)$$

  $-1 \leq r_{XY} \leq 1$

  - The sample correlation coefficient between first crack and failure loads is 0.2605.

1.2 Sample Percentile

Sample $100p$ percentile:

The data value such that $100p\%$ of the data are less than or equal to it.
25 percentile = first quantile
50 percentile = second quantile
75 percentile = third quantile

1.3 Chebyshev's Inequality

Data set: $x_1, x_2, \dots, x_n$

Sample mean $\bar{x}$
Sample standard deviation $s > 0$
Define: $S_k = \{i, 1 \le i \le n : |x_i - \bar{x}| < ks\}$
$N(S_k)$ = Number of elements in the set S_k (i.e., No. of i such that $|x_i - \bar{X}| < ks$)
For $k \ge 1$

$$\frac{N(S_k)}{n} \ge 1 - \frac{n-1}{nk^2} > 1 - \frac{1}{k^2}$$

One sided version: $N(k)$ = No. of i such that $x_i - \bar{x} \ge ks$
Then for $k \ge 1$:

$$\frac{N(k)}{n} \le \frac{1}{1+k^2}$$

1.4 Graphical Displays

- Histograms
- Cumulative frequency plot
- Box plots

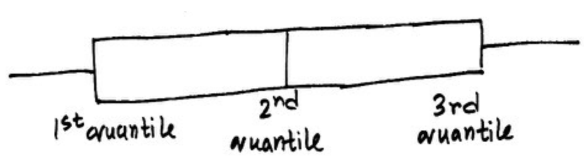

Figure 1.1: Box plot.

Chapter 2

Elements of Probability

Probability — two interpretations —

- frequency interpretation
- subjective interpretation

Sample space

The set of all possible outcomes of an experiment (denoted by S or Ω)

Any subset E of the sample space is known as an event.

Example 1.

A coin is to be tossed until a head appears twice in a row.

Sample space, $S = \{(H, H), (T, H, H), (H, T, H, H), (T, T, H, H), \ldots\}$.

We can also write this in a different way: $S = \{(e_1, e_2, \ldots, e_n, e_{n-1}), n \geq 2\}$ where e_i is either H or T and $e_{n-1} = e_n = H$, $e_{n-2} = T$.

2.1 Axiomatic Definition of Probability

Sample space: S
Event: E
Probability of event E, $P(E)$ satisfies:

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

Axiom 2: $P(S) = 1$

Axiom 3: Mutually exclusive events $E_1, E_2, \ldots$ (i.e., $E_i \cap E_j = \phi$, when $i \neq j$)
$P\left(\cup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$, $n = 1, 2, \ldots, \infty$.

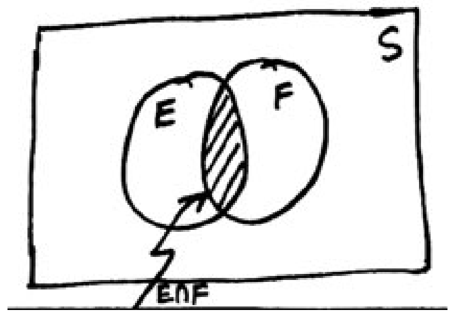

Figure 2.1: Venn diagram showing two events E and F.

Corollary:

(i) E and E^c are mutually exclusive, i.e.,

$$E \cup E^c = S$$
$$P(E \cup E^c) = P(S) = 1$$
$$P(E^c) = 1 - P(E)$$

(ii) Two events E and F, $P(E \cup F) = P(E) + P(F) - P(E \cap F)$. (Note: $E \cap F$ is also written as EF) see Figure 2.2.

Inclusion-Exclusion Identity

$$P(E_1 \cap E_2 \cap \cdots \cap E_n)$$
$$= \sum_i P(E_i) - \sum_{i<j} P(E_i E_j) + \sum_{i<j<k} P(E_i E_j E_k) - \cdots + (-1)^{n+1} P(E_1 E_2 \ldots E_n)$$

2.2 Conditional Probabilities

Probability that E occurs given that F has occurred, denoted by

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

2.3 Independent Events

If $P(EF) = P(E)P(F)$ then E F are independent. We also have $P(E|F) = P(E)$

Example 2.

A fair coin is to be tossed until a head appears twice in a row. What is the probability that it will be tossed exactly three times?

$$\begin{aligned} P\{3 \text{ tosses}\} &= P\{(T, H, H)\} \\ &= \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) \\ &= \frac{1}{8} \end{aligned}$$

What is the probability that it will be tossed exactly four times?

$$\begin{aligned} P\{4 \text{ tosses}\} &= P\{(T, T, H, H) \cup (H, T, H, H)\} \\ &= P\{(T, T, H, H)\} + P\{(H, T, H, H)\} \\ &= \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) + \left(\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}\right) \\ &= \frac{1}{16} + \frac{1}{16} = \frac{1}{8} \end{aligned}$$

Example 3.

A, B, C three events.

(a) only C occurs: $C \cap A^c \cap B^c$

(b) at least two events occur: $(A \cap B) \cup (A \cap C) \cup (B \cap C)$

(c) at least one event occurs: $A \cup B \cup C$

(d) all three events occur: $A \cap B \cap C$

(e) at most two occur: $(A \cap B \cap C)^c$

(f) none occurs: $(A \cup B \cup C)^c = A^c B^c C^c$

Example 4.

Boole's inequality: $P\left(\cup_{i=1}^n E_i\right) \leq \sum_{i=1}^n P(E_i)$

Proof: $\cup_{i=1}^n E_i = E_1 \cup E_1^c E_2 \cup E_1^c E_2^c E_3 \cup \cdots \cup E_1^c \ldots E_{n-1}^c E_n^c$

Denote $F_1 = E_1$, $F_2 = E_1^c \cap E_2, \ldots, F_n = E_1^c \ldots E_{n-1}^c E_n$.

Hence, $\cup_{i=1}^n E_i = \cup_{i=1}^n F_i$.

But F_i are mutually exclusive.

$P\left(\cup_{i=1}^n E_i\right) = P\left(\cup_{i=1}^n E_i\right) = \sum_{i=1}^n P(F_i) = \sum_{i=1}^n P(E_1^c \ldots E_{i-1}^c E_i) \leq \sum_{i=1}^n P(E_i)$.

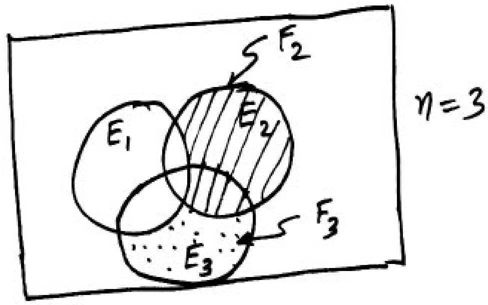

Figure 2.2: Venn diagram showing three events E_1, E_2, and E_3. Here, $F_1 = E_1$, $F_2 = E_2 \cap E_1^c$ (hatched), and $F_3 = E_3 \cap E_1^c \cap E_2^c$ (dotted).

Example 5.

A deck of 52 playing cards, containing all 4 Jacks, is randomly divided into 4 piles of 13 cards each.

(a)

$$\begin{aligned} P(E_1) &= P(\{\text{the first pile has exactly 1 Jack}\}) \\ &= \frac{\binom{4}{1} \times \binom{48}{12}}{\binom{52}{13}} \approx 0.4388 \end{aligned}$$

(b) Similarly,

$$\begin{aligned} P(E_2) &= P(\{\text{the second pile has exactly 1 Jack}\}) \\ &= \frac{\binom{4}{1} \times \binom{48}{12}}{\binom{52}{13}} \approx 0.4388 \end{aligned}$$

(c)

$$\begin{aligned} P(E_2|E_1) &= P(\{\text{the second pile has exactly 1 Jack given that first pile has exactly 1 Jack}\}) \\ &= \frac{\binom{3}{1} \times \binom{36}{12}}{\binom{39}{13}} \approx 0.4623 \end{aligned}$$

(d)

$$\begin{aligned} P(E_3|E_1E_2) &= P(\{\text{the third pile has exactly 1 Jack} \mid \text{first and second pile have exactly 1 Jack each}\}) \\ &= \frac{\binom{2}{1} \times \binom{24}{12}}{\binom{26}{13}} \approx 0.52 \end{aligned}$$

(e)

$$P(E_4|E_1E_2E_3) = 1$$

(f)

$$\begin{aligned} P(E_1E_2E_3E_4) &= P(\{\text{all the piles have exactly 1 Jack each}\}) \\ &= P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3) \\ &\approx 0.1055 \end{aligned}$$

Example 6.

N graduating students throw their graduate caps and then each student randomly selects one.

Define the events E_i = ith student gets his/her own cap

The probability that none of the N students gets his/her own cap is

$$\begin{aligned}
P(\text{no one selects own cap}) &= 1 - P(E_1 \cup E_2 \cup \cdots \cup E_N) \\
&= 1 - \left[\sum_{i=1}^{N} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \cdots + (-1)^{N+1}P(E_1E_2\ldots E_N)\right] \\
&= 1 - \sum_{i=1}^{N} P(E_i) + \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \cdots - (-1)^{N+1}P(E_1E_2\ldots E_N)
\end{aligned}$$

Now, $P(E_{i_1}E_{i_2}\ldots E_{i_k}) = \frac{(N-k)!}{N!}$ and

$$\begin{aligned}
\sum_{i_1<i_2<\cdots<i_k} P(E_{i_1}E_{i_2}\ldots E_{i_k}) &= \sum_{i_1<i_2<\cdots<i_k} \frac{(N-k)!}{N!} \\
&= \binom{N}{k}\frac{(N-k)!}{N!} \\
&= \frac{N!}{(N-k)!k!}\frac{(N-k)!}{N!} \\
&= \frac{1}{k!}
\end{aligned}$$

Hence,

$$\begin{aligned}
P(\text{no one selects own cap}) &= 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots - (-1)^{N+1}\frac{1}{N!} \\
&= 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + (-1)^{N}\frac{1}{N!}
\end{aligned}$$

Example 7.

You want to invest in a computer hardware tool. The probability that in any year that hardware will get damaged $= p$. The probability that the hardware will become obsolete in year $i = q_i$ (given the tool is not obsolete in the prior years).

Define events:

D_i = the hardware gets damaged in year i

O_i = the hardware becomes obsolete in year i

$P(D_i) = p$ and $P(O_i|O_{i-1}) = q_i$

The probability that the bridge's life does not end in the first year

$$\begin{aligned}
&= P(D_1^c \cap O_1^c) = P(D_1^c)P(O_1^c) \\
&= [1 - P(D_1)][1 - P(O_1)] \\
&= (1-p)(1-q_1)
\end{aligned}$$

For 2nd year,

$$\begin{aligned}= P[(D_1^c \cap O_1^c) \cap (D_2^c \cap O_2^c)] &= P(D_1^c \cap O_1^c) P(D_2^c \cap O_2^c | D_1^c \cap O_1^c) \\ &= P(D_1^c \cap O_1^c) P(D_2^c | O_2^c \cap D_1^c \cap O_1^c) P(O_2^c | D_1^c \cap O_1^c)\end{aligned}$$

Since the events D_i and O_i are independent

$$P(D_2^c | O_2^c \cap D_1^c \cap O_1^c) = P(D_2^c) = 1 - p$$

$$P(O_2^c | D_1^c \cap O_1^c) = P(D_2^c) = 1 - q_2$$

Hence,

$$\begin{aligned}&P[(D_1^c \cap O_1^c) \cap (D_2^c \cap O_2^c)] \\ &= (1-p)(1-q_1)(1-p)(1-q_2) \\ &= (1-p)^2(1-q_1)(1-q_2)\end{aligned}$$

For nth year,

$$\begin{aligned}&P[\text{survival through } n \text{ years}] \\ &= P(D_1^c \cap O_1^c \cap \cdots \cap D_n^c \cap O_n^c) \\ &= (1-p)^n \prod_{i=1}^{n} (1-q_i)\end{aligned}$$

The life of the hardware ends in year $n =$ the hardware has survived $n-1$ years

$$\begin{aligned}&P[\text{survival through } n-1 \text{ years}] \\ &= (1-p)^{n-1} \prod_{i=1}^{n-1} (1-q_i)\end{aligned}$$

Also,

$$\begin{aligned}&P[\text{the hardware's life ends in year } n] \\ &= P(D_n \cup O_n | \text{survival through } n-1 \text{ years}) \\ &= P(D_n | \text{previous survival}) + P(O_n | \text{previous survival}) - P(D_n \cap O_n | \text{previous survival}) \\ &= P(D_n) + P(O_n) - P(D_n) P(O_n | O_{n-1}^c) \\ &= p + q_n - pq_n\end{aligned}$$

Hence,

$$P(\text{life ends in year } n) = (p + q_n - pq_n)(1-p)^{n-1} \prod_{i=1}^{n-1} (1-q_i)$$

2.4 Total Probability

- An event A
- N mutually exclusive events $B_n, n = 1, 2, \ldots, N$ where $\cup_{i=1}^{N} B_i = S$

Then $P(A) = \sum_{n=1}^{N} P(A|B_n)P(B_n)$

Proof: $A = A \cap S = A \cap \left(\cup_{i=1}^{N} B_n\right) = \cup_{i=1}^{N}(A \cap B_n)$. Also, $(A \cap B_n)$ are mutually exclusive events. Hence,

$$\begin{aligned} P(A) = P(A \cap S) &= P\left[\cup_{i=1}^{N}(A \cap B_n)\right] \\ &= \sum_{i=1}^{N} P(A \cap B_n) \\ &= \sum_{i=1}^{N} P(A|B_n)P(B_n) \end{aligned}$$

2.5 Bayes' Theorem

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{\sum_{j=1}^{N} P(A|B_j)P(B_j)}$$

Proof:

$$\begin{aligned} P(B_n|A) &= \frac{P(B_n \cap A)}{P(A)} \\ &= \frac{P(A|B_n)P(B_n)}{P(A)} \\ &= \frac{P(A|B_n)P(B_n)}{\sum_{j=1}^{N} P(A|B_j)P(B_j)} \qquad \text{[using theorem of total probability]} \end{aligned}$$

Example 8.

Basket 1: 7 Red balls & 5 Blue balls

Basket 2: 4 Red balls & 12 Blue balls

A ball is selected randomly from one of the baskets. If the selected ball is Red what is the probability that it has been selected from Basket 2?

Define:

R = event of selecting a Red ball,

B_1 = selecting Basket 1,

B_2 = selecting Basket 2.

Hence, $P(B_2|R) =?$

$P(R|B_1) = 7/12$, $P(R|B_2) = 4/16 = 1/4$, $P(B_1) = P(B_2) = 1/2$.

Using Bayes' theorem,

$$\begin{aligned}
P(B_2|R) &= \frac{P(R|B_2)P(B_2)}{P(R|B_1)P(B_1) + P(R|B_2)P(B_2)} \\
&= \frac{\frac{1}{4} \times \frac{1}{2}}{\frac{7}{12} \times \frac{1}{2} + \frac{1}{4} \times \frac{1}{2}} \\
&= \frac{3}{10}
\end{aligned}$$

Example 9.

40% of the students from a class are good in a subject. Class tests are performed but the tests are only 90% reliable, *i.e.*, tests can identify good students only 90% of the time. Define the events:

G = good student
T = the student scores well in the test
$P(G) = 0.4$, $P(T|G) = 0.9$, $P(T|G^c) = 0.1$.
What is the probability that the student is good if he/she passes the test, i.e., $P(G|T) =?$
Using Bayes' theorem,

$$\begin{aligned}
P(G|T) &= \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|G^c)P(G^c)} \\
&= \frac{0.9 \times 0.4}{0.9 \times 0.4 + 0.1 \times 0.6} \\
&\approx 0.8571
\end{aligned}$$

Example 10.

Two routes from Los Angeles to Santa Barbara.

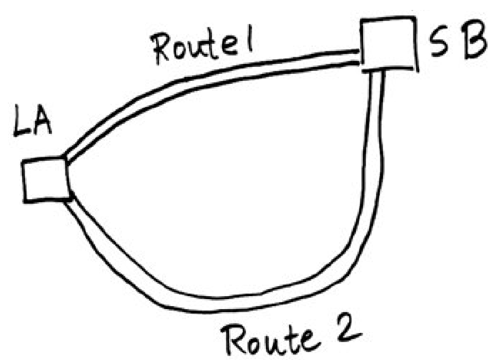

Figure 2.3: Two routes from LA to SB.

Define events:
R_1 = Route 1 is open
R_2 = Route 2 is open
During a wildfire, $P(R_1) = 0.8$, $P(R_2) = 0.4$, $P(R_1 \cap R_2) = 0.25$.
What is the probability that Route 1 is open given that Route 2 is open?

$$P(R_1|R_2) = \frac{P(R_1 \cap R_2)}{P(R_2)} = \frac{0.25}{0.4} = 0.625$$

What is the probability that Route 1 is closed given that Route 2 is closed?

$$P(R_1^c|R_2^c) = \frac{P(R_1^c \cap R_2^c)}{P(R_2^c)} = ?$$

$P(R_1^c) = 1 - P(R_1) = 0.2$
$P(R_2^c) = 1 - P(R_2) = 0.6$

$$\begin{aligned}
P(R_1^c \cap R_2^c) &= 1 - P\big([R_1^c \cap R_2^c]^c\big) \\
&= 1 - P(R_1 \cup R_2) \qquad \text{[see figure below]} \\
&= 1 - [P(R_1) + P(R_2) - P(R_1 \cap R_2)] \\
&= 1 - [0.8 + 0.4 - 0.25] \\
&= 1 - 0.95 \\
&= 0.95
\end{aligned}$$

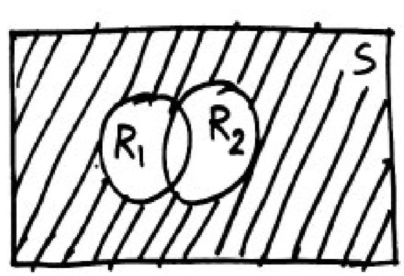

Figure 2.4: Venn diagram showing two events R_1 and R_2.

Hence,

$$P(R_1^c|R_2^c) = \frac{P(R_1^c \cap R_2^c)}{P(R_2^c)} = \frac{0.05}{0.6} = 0.0833$$

The probability of Route 1 being open given that Route 2 is closed is
$P(R_1|R_2^c) = 1 - P(R_1^c|R_2^c)$

Chapter 3

Random Variables

A mapping that transforms the events to the real line.

Example 1.

Toss a fair coin.
Define a random variable X where
X is 1 if head appears and
X is 0 if tail appears.
Hence,
$P(X = 0) = 1/2$
$P(X = 1) = 1/2$

Example 2.

Cast two dice.
Define the random variable as sum of the outcomes.
Hence,

$$
\begin{aligned}
&P(X = 2) = P\{(1,1)\} = 1/36 \\
&P(X = 3) = P\{(1,2),(2,1)\} = 2/36 \\
&P(X = 4) = P\{(1,3),(2,2),(3,1)\} = 3/36 \\
&P(X = 5) = P\{(1,4),(2,3),(3,2),(4,1)\} = 4/36 \\
&P(X = 6) = P\{(1,5),(2,4),(3,3),(4,2),(5,1)\} = 5/36 \\
&P(X = 7) = P\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} = 6/36 \\
&P(X = 8) = P\{(2,6),(3,5),(4,4),(5,3),(6,2)\} = 5/36 \\
&P(X = 9) = P\{(3,6),(4,5),(5,4),(6,3)\} = 4/36 \\
&P(X = 10) = P\{(4,6),(5,5),(6,4)\} = 3/36 \\
&P(X = 11) = P\{(5,6),(6,5)\} = 2/36 \\
&P(X = 12) = P\{(6,6)\} = 3/36
\end{aligned}
$$

3.1 Cumulative Distribution Function (CDF)

For any real number x

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

i.e., the probability that the random variable X takes on a value less than or equal to x.

Note:

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \end{aligned}$$

3.2 Types of RV

3.2.1 Discrete Random Variable

X takes discrete values

Probability Mass Function (pmf)

$$p(a) = P(X = a)$$

Hence,

- CDF: $F(a) = \sum_{\text{all } x \leq a} p(x)$,
- $F(\infty) = \sum_{i=1}^{\infty} p(x_i) = 1$,
- $F(-\infty) = 0$.

Example 3.

Cast a die.

X = outcome

Hence, the probability mass function

$$p_X(i) = \frac{1}{6}, \qquad i = 1, 2, \ldots, 6$$

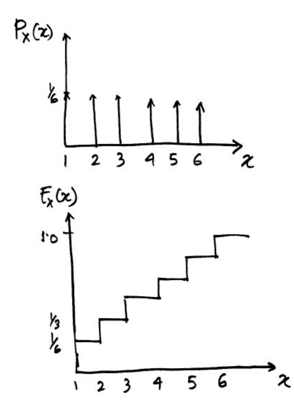

Figure 3.1: pmf and CDF of die cast experiment.

3.2.2 Continuous Random Variable

possible values of X is an interval.

$$P(X \in B) = \int_B \underbrace{f(x)}_{\text{pdf}} dx$$

Note:

- CDF: $F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)dx$,
- $f(x)$ is called probability density function (pdf) of X,
- $F(\infty) = \int_{-\infty}^{\infty} f(x)dx = P[X \in (-\infty, \infty)] = 1$,
- $P(a \leq X \leq b) = \int_a^b f(x)dx$ but $P(X = a) = \int_a^a f(x)dx = 0$,
- $\frac{d}{da}F(a) = f(a)$,
- $F(-\infty) = 0$.

Example 4.

Let the random variable X has a probability density function (pdf)

$$f(x) = \begin{cases} c & 0 < x \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

where c is a constant.

To estimate c use $F(\infty) = 1$

$$\begin{aligned} &\int_{-\infty}^{\infty} f(x)dx = 1 \\ \Rightarrow &\int_0^{10} cdx = 1 \\ \Rightarrow &c \cdot x\Big|_0^{10} = 1 \\ \Rightarrow &c = 0.1 \end{aligned}$$

The cumulative distribution function

$$F(x) = \begin{cases} \int_0^x cdx = cx = 0.1x, & 0 < x \leq 10 \\ 1, & x > 10 \\ 0, & x < 0 \end{cases}$$

What is the probability that X is between 2 and 5?

$$P(2 < X \leq 5) = \int_2^5 cdx = 0.3$$

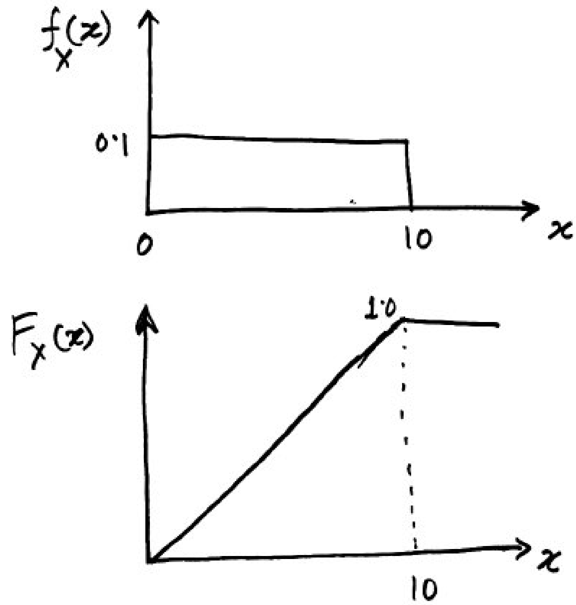

Figure 3.2: pdf and CDF of X.

3.3 Expectation

$$\mathbb{E}[X] = \sum_i x_i P(X = x_i) = \sum_i x_i p(x_i)$$

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

Example 5.

Cast a die and denote the outcome by a random variable X. Hence,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_i x_i P(X = x_i) \\
&= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) \\
&\quad + 4 \cdot P(X = 4) + 5 \cdot P(X = 5) + 6 \cdot P(X = 6) \\
&= 21 \cdot \frac{1}{6} = 3.5
\end{aligned}$$

Example 2 contd.

From Example 2, the expected sum of two dice

$$\begin{aligned}
\mathbb{E}[X] &= 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + \cdots + 12 \cdot P(X = 12) \\
&= \sum_{i=2}^{12} i P(X = i) \\
&= 252/36 = 7
\end{aligned}$$

Example 4 contd.

From Example 4, the expected value of X is

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^{10} cxdx \\ &= 0.1 \int_0^{10} xdx \\ &= 0.1 \cdot \frac{x^2}{2}\Big|_0^{10} \\ &= 5 \end{aligned}$$

3.4 Variance

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}[(X-\mu)^2] \qquad \text{where } \mu = \mathbb{E}[X] \\ &= \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \end{aligned}$$

- $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$
- $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
- $\mathrm{Var}(b) = 0$
- $\mathrm{Var}(X+X) = 4\mathrm{Var}(X)$

Standard deviation $= \sqrt{\mathrm{Var}(X)}$
Covariance of two random variables X, Y

$$Cov(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,X) = \mathrm{Var}(X)$
- $Cov(aX,Y) = a\ Cov(X,Y)$
- $Cov(X+Z,Y) = Cov(X,Y) + Cov(Z,Y)$
- $Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m Cov(X_i, Y_j)$
- $\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i=1}^n \sum_{j=1, j\neq i}^n Cov(X_i, X_j)$

If X and Y are independent random variables, then

$$Cov(X,Y) = 0$$
$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i)$$

The correlation coefficient

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

3.5 Properties of the Expected Value

- Discrete RV: $\mathbb{E}[g(X)] = \sum_x g(x)p(x)$
- Continuous RV: $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$
- $\mathbb{E}[aX+b] = a\mathbb{E}[X] + b$
- $$\mathbb{E}[X^n] = \begin{cases} \sum_x x^n p(x) & \text{Discrete RV} \\ \int_{-\infty}^{\infty} x^n f(x)dx & \text{Continuous RV} \end{cases}$$
- Expected value of a function of two RVs
$$\mathbb{E}[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y) & \text{Discrete RV} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy & \text{Continuous RV} \end{cases}$$
- $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Example 2 contd.

Let us denote the outcome of the first die by X_1 and the second die by X_2. Hence, $X = X_1 + X_2$.

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_1 + X_2] \\ &= \mathbb{E}[X_1] + \mathbb{E}[X_2] \\ &= 3.5 + 3.5 = 7 \qquad \text{(see Example 6)} \end{aligned}$$

3.6 Moment Generating Function, $\phi(t)$

$$\phi(t) = \mathbb{E}\left[e^{tX}\right] = \begin{cases} \sum_x e^{tx}p(x) & \text{Discrete RV} \\ \int_{-\infty}^{\infty} e^{tx}f(x)dx & \text{Continuous RV} \end{cases}$$

nth moment of the random variable is $\mathbb{E}[X^n]$ and it can be computed from $\phi(t)$ using

$$\mathbb{E}[X^n] = \frac{d^n}{dt^n}\phi(t)\bigg|_{t=0}$$

i.e., $\mathbb{E}[X] = \phi'(0)$ and $\mathbb{E}[X^2] = \phi''(0)$.

Example 6.

The moment generating function of X with pmf $p(i) = \frac{\lambda^i e^{-\lambda}}{i!}$ is

$$\begin{aligned}\phi(t) = E[e^{tX}] &= \sum_{i=0}^{\infty} e^{ti} \frac{\lambda^i e^{-\lambda}}{i!} \\ &= e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= e^{\lambda(e^t - 1)}\end{aligned}$$

$\phi'(t) = \lambda e^t e^{\lambda(e^t-1)}$,
$\phi''(t) = (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}$.
This gives $\mathbb{E}[X] = \phi'(0) = \lambda$ and $\mathbb{E}[X^2] = \phi''(0) = \lambda^2 + \lambda$

3.7 Markov's Inequality

For any value $a > 0$,

$$P(X \geq 0) \leq \frac{\mathbb{E}[X]}{a}$$

3.8 Chebyshev's Inequality

If the mean of X is μ and variance is σ^2, for any $k > 0$,

$$P\Big(|X - \mu| \geq k\Big) \leq \frac{\sigma^2}{k^2}$$

3.9 The Weak Law of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. (independent and identically distributed) random variables with $\mathbb{E}[X_i] = \mu$.

For any $\epsilon > 0$

$$P\Big(\Big|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\Big| > \epsilon\Big) \to 0 \qquad \text{as } n \to \infty$$

Proof:

$$\begin{aligned}P\Big(\Big|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\Big| > \epsilon\Big) &= P\Big(\Big|X_1 + X_2 + \cdots + X_n - n\mu\Big| > n\epsilon\Big) \\ &\leq \frac{\mathrm{Var}(X_1 + X_2 + \cdot + X_n)}{n^2\epsilon^2} \qquad \text{[using Chebyshev's inequality]} \\ &= \frac{n\sigma^2}{n^2\epsilon^2} \\ &= \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty\end{aligned}$$

3.10 Jointly Distributed RVs

Joint CDF of X and Y

$$F(x,y) = P(X \leq x, Y \leq y)$$

Hence,

- $F_X(x) = P(X \leq x) = P(X \leq x, Y \leq \infty) = F(x, \infty)$
- $F_Y(y) = P(Y \leq y) = P(X \leq \infty, Y \leq y) = F(\infty, y)$
- Joint pmf $p(x_i, y_i) = P(X = x_i, Y = y_i)$
- Marginal pmf: $p_X(x_i) = P(X = x_i) = P\Big(\cup_j \{X = x_i, Y = y_j\}\Big) = \sum_j p(x_i, y_j)$
- $p_Y(y_i) = P(Y = y_i) = P\Big(\cup_i \{X = x_i, Y = y_j\}\Big) = \sum_i p(x_i, y_j)$
- Joint pdf $f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b)$
- Marginal densities: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$
- $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$
- If X and Y are independent
  - $F(a,b) = F(a)F(b)$
  - Discrete RV: $p(x,y) = p_X(x)p_Y(y)$
  - Continuous RV: $f(x,y) = f_X(x)f_Y(y)$

Example 7.

Let X be a random variable with probability density

$$f(x) = \begin{cases} c(1-x^2), & -1 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Determine the value of c and $F(x)$.

$$\begin{aligned} F(\infty) &= \int_{-\infty}^{\infty} f(x)dx = 1 \\ \Rightarrow \quad & \int_{-1}^{1} c(1-x^2)dx = 1 \\ \Rightarrow \quad & c\big[x - x^3/3\big]_{-1}^{1} = 1 \\ \Rightarrow \quad & c\big[(1-1/3) - (-1+1/3)\big] = 1 \\ \Rightarrow \quad & c\big[2/3 + 2/3\big] = 1 \\ \Rightarrow \quad & c = 3/4 \end{aligned}$$

The cumulative distribution function

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(a)da \\
&= \int_{-1}^{x} c(1-a^2)da \\
&= c\Big[a - \frac{a^3}{3}\Big]_{-1}^{x} \\
&= \frac{3}{4}\Big[x - \frac{x^3}{3} + \frac{2}{3}\Big]
\end{aligned}
$$

Example 8.

The longevity T of light bulbs is described by the following probability density function

$$
f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}
$$

where λ is a constant.

The cumulative distribution function

$$
\begin{aligned}
F(t) &= \int_{-\infty}^{t} f(\tau)d\tau \\
&= \int_{0}^{t} \lambda e^{-\lambda\tau}d\tau \\
&= \lambda\Big[-\frac{1}{\lambda}e^{-\lambda\tau}\Big]_{0}^{t} \\
&= [-e^{-\lambda t} - (-1)] \\
&= 1 - e^{-\lambda t}
\end{aligned}
$$

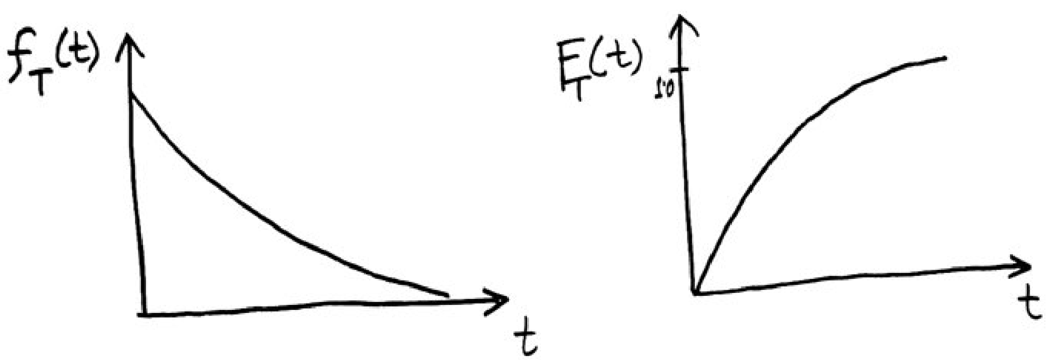

Figure 3.3: pdf and CDF of T.

The mean life is

$$\begin{aligned}\mathbb{E}[T] &= \int_{-\infty}^{\infty} tf(t)dt \\ &= \int_{0}^{\infty} t\lambda e^{-\lambda t}dt \\ &= -te^{-\lambda t}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t}dt \\ &= 0 - \frac{e^{-\lambda t}}{\lambda}\Big|_0^{\infty} \\ &= \frac{1}{\lambda}\end{aligned}$$

Median: $F(t_m) = \int_0^{t_m} \lambda e^{-\lambda t}dt = 0.5$. This gives $t_m = \frac{1}{\lambda}[-\log(0.5)] = 0.693/\lambda = 0.693\mathbb{E}[T]$.

$$\mathrm{Var}(T) = \int_0^{\infty} (t - 1/\lambda)^2 \lambda e^{-\lambda t}dt = 1/\lambda^2$$

Example 9.

X and Y are independent random variables with means μ_X and μ_Y, respectively, and variances σ_X^2 and σ_Y^2, respectively.

$\mathbb{E}[XY] =?$ and $\mathrm{Var}(XY) =?$

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = \mu_X\mu_Y$$

$$\begin{aligned}\mathbb{E}[(XY)^2] &= \mathbb{E}[X^2]\mathbb{E}[Y^2] \\ &= \{\mathrm{Var}(X) + (\mathbb{E}[X])^2\}\{\mathrm{Var}(Y) + (\mathbb{E}[Y])^2\} \\ &= (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2)\end{aligned}$$

Hence,

$$\begin{aligned}\mathrm{Var}(XY) &= \mathbb{E}[(XY)^2] - (\mathbb{E}[XY])^2 \\ &= (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) - \mu_x^2\mu_Y^2 \\ &= \sigma_X^2\sigma_Y^2 + \sigma_X^2\mu_Y^2 + \sigma_Y^2\mu_X^2\end{aligned}$$

Example 10.

At the graduation ceremony N students throw their caps and the select one at random. What is the expected number of students who will get their own cap back?

Let X denote the no. of students who select their own cap

$$X_i = \begin{cases} 1, & \text{if } i\text{th student selects own cap} \\ 0, & \text{otherwise} \end{cases}$$

Hence, $X = \sum_{i=1}^{N} X_i$

Also, $P(X_i = 1)$ = probability that ith student select own hat $= 1/N$

$$\mathbb{E}[X_i] = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = \frac{1}{N}$$

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{N} X_i\right] \\
&= \sum_{i=1}^{N} \mathbb{E}[X_i] \\
&= \sum_{i=1}^{N} \frac{1}{N} \\
&= N \cdot \frac{1}{N} \\
&= 1
\end{aligned}$$

Example 11.

A basket has n Red balls and m Blue balls. k balls are selected at random from the basket.
Let X denote the number of Red balls selected.
$P(X = i) =$? and $\mathbb{E}[X] =$?

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

Let us denote

$$X_j = \begin{cases} 1, & \text{if } j\text{th ball selected is Red} \\ 0, & \text{otherwise} \qquad j = 1, 2, \ldots, k \end{cases}$$

Hence, $X = \sum_{j=1}^{k} X_j$
$\mathbb{E}[X] = \mathbb{E}\left[\sum_{j=1}^{k} X_j\right] = \sum_{j=1}^{k} \mathbb{E}[X_j]$
Now,

$$\begin{aligned}
\mathbb{E}[X_j] &= 1 \cdot P(X_j = 1) + 0 \cdot P(X_j = 0) \\
&= 1 \cdot \frac{n}{n+m} + 0 \\
&= \frac{n}{n+m}
\end{aligned}$$

This gives

$$\mathbb{E}[X] = \sum_{j=1}^{k} \mathbb{E}[X_j] = \sum_{j=1}^{k} \frac{n}{n+m} = \frac{nk}{n+m}$$

Example 12.

Joint probability mass function of X and Y is given by

$$P(X=i, Y=j) = \binom{j}{i}\frac{e^{-2\lambda}\lambda^j}{j!}, \quad 0 \le i \le j$$

$P(X=i) = ?$

$$\begin{aligned}
P(X=i) &= \sum_{j=i}^{\infty}\binom{j}{i}\frac{e^{-2\lambda}\lambda^j}{j!} \\
&= \frac{1}{i!}e^{-2\lambda}\sum_{j=i}^{\infty}\frac{1}{(j-i)!}\lambda^j \\
&= \frac{\lambda^i}{i!}e^{-2\lambda}\sum_{j=i}^{\infty}\frac{\lambda^{(j-i)}}{(j-i)!} \\
&= \frac{\lambda^i}{i!}e^{-2\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \\
&= \frac{\lambda^i}{i!}e^{-2\lambda}e^{\lambda} \\
&= \frac{\lambda^i}{i!}e^{-\lambda}
\end{aligned}$$

Example 13.

Assume an arrow hitting any point inside a disk is equally likely, *i.e.*, the hitting point is uniformly distributed within the disk of radius R. Hence, $f(x,y) = k, 0 \le x^2+y^2 \le R^2$ Determine $k = ?$ Determine $P(D \le d) = ?$ where D denotes distance between the hitting point and center of the disk.

We know

$$\begin{aligned}
& F(\infty,\infty) = 1 \\
\Rightarrow\ & \iint_{0\le x^2+y^2\le R^2} k\,dx\,dy = 1 \\
\Rightarrow\ & k \underbrace{\left[\iint_{0\le x^2+y^2\le R^2} dx\,dy\right]}_{\text{area of the circle with radius } R} = 1 \\
\Rightarrow\ & k\cdot\pi R^2 = 1 \\
\Rightarrow\ & k = \frac{1}{\pi R^2}
\end{aligned}$$

Hence, $f(x,y) = 1/\pi R^2$

$D =$ distance between the hitting point and center of the disk.
Hence, $D \leq d \quad \Rightarrow \quad x^2 + y^2 \leq d^2$.
So,

$$\begin{aligned}
P(D \leq d) &= \iint\limits_{0 \leq x^2+y^2 \leq d^2} \frac{1}{\pi R^2} dxdy \\
&= \frac{1}{\pi R^2} \underbrace{\left[\iint\limits_{0 \leq x^2+y^2 \leq d^2} dxdy \right]}_{\text{area of the circle with radius } d} \\
&= \frac{1}{\pi R^2} \cdot \pi d^2 \\
&= \frac{d^2}{R^2}
\end{aligned}$$

Example 14.

Let X has density

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

$\mathbb{E}[X^3] = ?$

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} g(x) f(x) dx \\
\mathbb{E}[X^3] &= \int_0^1 x^3 dx \\
&= \left. \frac{x^4}{4} \right|_0^1 \\
&= \frac{1}{4}
\end{aligned}$$

OR
Let $Y = X^3$

$$F_X(x) = \int_0^x 1 \cdot dx = x, \qquad 0 < x < 1$$

Hence,

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X^3 \leq y) \\
&= P(X \leq y^{1/3}) \\
&= y^{1/3}
\end{aligned}$$

So, we can write

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{3}y^{-2/3}, \qquad 0 < y < 1$$

Hence,

$$\begin{aligned}
\mathbb{E}[X^3] &= \mathbb{E}[Y] \\
&= \int_{-\infty}^{\infty} y f_Y(y) dy \\
&= \int_0^1 y \frac{1}{3} y^{-2/3} dy \\
&= \int_0^1 \frac{1}{3} y^{1/3} dy \\
&= \frac{1}{3}\left[\frac{3}{4}a^{4/3}\right]_0^1 \\
&= \frac{1}{4}
\end{aligned}$$

Example 15.

A random variable has a triangular distribution

$$f(x) = \begin{cases} \frac{x-3}{25} & 3 \le x < 8 \\ 0.2 - \frac{x-8}{25} & 8 \le x < 13 \\ 0 & \text{otherwise} \end{cases}$$

Determine $P(X \le 4) =?$

$$\begin{aligned}
P(X \le 4) = F(4) &= \int_{-\infty}^{4} f(x)dx \\
&= \int_3^4 \frac{x-3}{25} dx \\
&= 0.02
\end{aligned}$$

Determine $P(X > 4) =?$

$$P(X > 4) = 1 - P(X \le 4) = 1 - F(4) = 1 - 0.02 = 0.98$$

Determine $P(4 < X \le 9) =?$

$$
\begin{aligned}
P(4 < X \leq 9) = F(9) - F(4) &= \int_{-\infty}^{9} f(x)dx - \int_{-\infty}^{4} f(x)dx \\
&= \int_{4}^{9} f(x)dx \\
&= \int_{4}^{8} \frac{x-3}{25}dx + \int_{8}^{9} \left[0.2 - \frac{x-8}{25}\right] dx \\
&= 0.48 + 0.18 \\
&= 0.66
\end{aligned}
$$

Estimate $\mathbb{E}[X]$

$$
\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} xf(x)dx \\
&= \int_{3}^{8} x\frac{x-3}{25}dx + \int_{8}^{13} x \left[0.2 - \frac{x-8}{25}\right] \\
&= 3.1667 + 4.8333 \\
&= 8
\end{aligned}
$$

Example 16.

Define the indicator random variable as

$$
I = \begin{cases} 1 & \text{if an event } A \text{ happens} \\ 0 & \text{otherwise} \end{cases}
$$

Show that the expected value of the random variable I is same as the probability of event A.

The probability mass function of I is

$$
\begin{aligned}
p(1) &= P(I = 1) = P(A) \\
p(0) &= P(I = 0) = P(A^c) = 1 - P(A)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
E[I] &= \sum_{i=0}^{1} ip(i) \\
&= 1 \cdot p(1) + 0 \cdot p(0) \\
&= 1 \cdot P(I = 1) + 0 \cdot P(I = 0) \\
&= P(I = 1) \\
&= P(A)
\end{aligned}
$$

Example 17.

Prove that $\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$

$$\begin{aligned} 0 \leq \mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ \mathbb{E}[X^2] \geq (\mathbb{E}[X])^2 \end{aligned}$$

$\mathbb{E}[X^2] = (\mathbb{E}[X])^2$ when $\mathrm{Var}(X) = 0$, i.e., X is deterministic.

Example 18.

Prove that $P(A) = P(A|X \leq x)F(x) + P(A|X > x)[1 - F(x)]$

Let us define an event $B = \{X \leq x\}$.
Hence, $B^c = \{X > x\}$.
Further, $P(B) = P\{X \leq x\} = F(x)$ and $P(B^c) = P\{X > x\} = [1 - F(x)]$
Therefore,

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|B^c)P(B^c) \\ &= P(A|X \leq x)F(x) + P(A|X > x)[1 - F(x)] \end{aligned}$$

Chapter 4

Conditional Probability Distributions

Any two events A and B with $P(B) > 0$

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

where $P(B) > 0$.

4.1 Discrete Random Variables

If X and Y are discrete random variables then the conditional pmf of X given $Y = y$

$$\begin{aligned} p_{X|Y}(x|y) &= P(X = x|Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{p(x, y)}{p_Y(y)} \qquad \forall y \text{ such that } p_Y(y) > 0 \end{aligned}$$

Conditional probability distribution function of X given $Y = y$

$$\begin{aligned} F_{X|Y}(x|y) &= P(X \leq x|Y = y) \\ &= \sum_{a \leq x} p_{X|Y}(a|y) \end{aligned}$$

Conditional expectation

$$\mathbb{E}[X|Y = y] = \sum_{x} x p_{X|Y}(x|y)$$

4.2 Continuous Random Variables

Conditional probability density function of X given $Y = y$

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} \qquad \forall y \text{ such that } f_Y(y) > 0$$

Conditional probability distribution function of X given $Y = y$

$$\begin{aligned} F_{X|Y}(x|y) &= P(X \le x | Y = y) \\ &= \int_{-\infty}^{x} f_{X|Y}(a|y) da \end{aligned}$$

Conditional expectation

$$\mathbb{E}[X|Y = y] = \int_{\infty}^{\infty} x f_{X|Y}(x|y) dx$$

Example 1.

Joint density of X and Y

$$f(x, y) = \begin{cases} 6xy(2 - x - y), & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

$\mathbb{E}[X|Y = y] = ?$
The marginal density

$$\begin{aligned} f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx &= \int_{0}^{1} 6xy(2 - x - y) dx \\ &= 6y \left[x^2 - \frac{x^3}{3} - \frac{x^2 y}{2} \right]_0^1 \\ &= y(4 - 3y) \end{aligned}$$

The conditional density

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{6xy(2 - x - y)}{y(4 - 3y)} = \frac{6x(2 - x - y)}{(4 - 3y)}$$

$$\begin{aligned} \mathbb{E}[X|Y = y] = \int_{\infty}^{\infty} x f_{X|Y}(x|y) dx &= \int_{0}^{1} \frac{6x^2(2 - x - y)}{(4 - 3y)} dx \\ &= \frac{(2 - y) \cdot 2 - 6/4}{4 - 3y} \\ &= \frac{5 - 4y}{8 - 6y} \end{aligned}$$

Example 2.

Joint density of X and Y

$$f(x, y) = \begin{cases} \frac{1}{2} y e^{-xy}, & 0 < x < \infty, 0 < y < 2 \\ 0, & \text{otherwise} \end{cases}$$

$\mathbb{E}[e^{X/2}|Y = 1] = ?$

$$
\begin{aligned}
f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx &= \int_0^{\infty} \frac{1}{2} y e^{-xy} dx \\
&= \frac{1}{2} y \int_0^{\infty} e^{-xy} dx \\
&= \frac{1}{2} y \cdot (-1/y) \left[e^{-xy}\right]_0^{\infty} \\
&= \frac{1}{2} y \cdot (-1/y)[0-(-1)] \\
&= 1/2
\end{aligned}
$$

Hence,

$$
f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{1}{2} y e^{-xy}}{1/2} = y e^{-xy}
$$

So, $f_{X|Y}(x|1) = e^{-x}$

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

Hence,

$$
\begin{aligned}
\mathbb{E}[e^{X/2}|Y=1] &= \int_{-\infty}^{\infty} e^{x/2} f_{X|Y}(x|1) dx \\
&= \int_0^{\infty} e^{x/2} e^{-x} dx \\
&= \int_0^{\infty} e^{-x/2} dx \\
&= 2
\end{aligned}
$$

Example 3.

Joint density function of two continuous random variables X and Y is given by

$$
\begin{aligned}
f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\Bigg[\frac{-1}{2(1-\rho^2)}\Bigg\{ \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho \left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) \\
+ \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \Bigg\}\Bigg], \qquad -\infty < x < \infty, -\infty < y < \infty
\end{aligned}
$$

where $\rho = Corr(X,Y)$, $\sigma_X = \sqrt{\text{Var}(X)}$, $\sigma_Y = \sqrt{\text{Var}(Y)}$, $\mu_X = \mathbb{E}[X]$, $\mu_Y = \mathbb{E}[Y]$

$$
f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left[-\frac{1}{2}\Big(\frac{y-\mu_Y}{\sigma_Y}\Big)^2\right]
$$

$$
f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{1}{2}\Big(\frac{x-\mu_X}{\sigma_X}\Big)^2\right]
$$

We can write the joint density as

$$f(x,y) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{1}{2}\Big(\frac{x-\mu_X}{\sigma_X}\Big)^2\right]$$
$$\times \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_Y+\rho(\sigma_X/\sigma_Y)(x-\mu_X)}{\sigma_Y\sqrt{1-\rho^2}}\right)^2\right]$$

Hence,

$$\begin{aligned}
f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\
&= \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_Y+\rho(\sigma_X/\sigma_Y)(x-\mu_X)}{\sigma_Y\sqrt{1-\rho^2}}\right)^2\right] \\
\mathbb{E}[Y|X=x] &= \mu_Y - \rho(\sigma_Y/\sigma_X)(x-\mu_X) \\
\mathrm{Var}(Y|X=x) &= \sigma_Y^2(1-\rho^2)
\end{aligned}$$

Example 4.

X is uniformly distributed in $(0,1)$. $\mathbb{E}[X|X\leq 0.25] =?$

$$f(x) = \begin{cases} 1, & 0<x<1 \\ 0, & \text{otherwise} \end{cases}$$

$$P(X\leq 0.25) = F(0.25) = \int_0^{0.25} 1\cdot dx = 0.25$$

Hence,

$$f_{X|X\leq 0.25}(x|x\leq 0.25) = \frac{f(x)}{P(x\leq 0.25)} = \frac{1}{0.25} = 4$$

$$\begin{aligned}
\mathbb{E}[X|X\leq 0.25] &= \int_0^{0.25} x f_{X|X\leq 0.25}(x|x\leq 0.25)dx \\
&= \int_0^{0.25} x\cdot 4dx \\
&= 4\cdot\left[\frac{x^2}{2}\right]_0^{0.25} \\
&= \frac{1}{8}
\end{aligned}$$

Example 5.

Joint density of X and Y

$$f(x,y) = \begin{cases} \frac{e^{-y}}{y}, & 0 < x < y, 0 < y < \infty \\ 0, & \text{otherwise} \end{cases}$$

$\mathbb{E}[X^2|Y = y] =?$

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)dx \\ &= \int_0^y \frac{e^{-y}}{y} dx \\ &= \frac{e^{-y}}{y} \int_0^y dx \\ &= \frac{e^{-y}}{y} \cdot y \\ &= e^{-y} \end{aligned}$$

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{\frac{e^{-y}}{y}}{e^{-y}} \\ &= \frac{1}{y} \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}[X^2|Y = y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x|y) \\ &= \int_0^y x^2 \cdot \frac{1}{y} dx \\ &= \frac{1}{y} \left[\frac{x^3}{3}\right]_0^y \\ &= \frac{y^2}{3} \end{aligned}$$

Chapter 5

Common Probability Distributions: Part 1

5.1 Discrete Random Variables

5.1.1 Bernoulli Random Variable (with parameter p)

The random variable x denotes the success from a trial. The probability mass function of the random variable X is given by

$$\begin{aligned} p_X(0) &= 1-p \\ p_X(1) &= p \end{aligned}$$

Hence,

$$\begin{aligned} &\mathbb{E}[X] = 0 \cdot (1-p) + 1 \cdot p = p \\ &\mathbb{E}[X^2] = p \\ &\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p) \end{aligned}$$

The moment generating function is

$$\phi(t) = \mathbb{E}[e^{tX}] = e^{t \cdot 0}(1-p) + e^{t \cdot 1} \cdot p = 1 - p + pe^t$$

Check: $\phi'(t) = pe^t$, $\phi''(t) = pe^t$. Hence, $\mathbb{E}[X] = \phi'(0) = p$, $\mathbb{E}[X^2] = \phi''(0) = p$.

5.1.2 Binomial Random Variable (with parameters n and p)

The probability mass function of the random variable X is given by

$$p_X(i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \ldots, n$$

Hence,

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{n} i p_X(i) \\
&= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{in!}{(n-i)!i!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^{n} \frac{n!}{(n-i)!(i-1)!} p^i (1-p)^{n-i} \\
&= np \sum_{i=1}^{n} \frac{(n-1)!}{(n-i)!(i-1)!} p^{i-1} (1-p)^{n-i} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{(n-1-k)!k!} p^{k} (1-p)^{n-1-k} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k} (1-p)^{n-1-k} \\
&= np[p + (1-p)]^{n-1} \\
&= np
\end{aligned}
$$

Alternately, Binomial random variable is number of successes in n trials. Hence, $X = \sum_{i=1}^{n} X_i$ where X_i are independent and identically distributed Bernoulli random variable.

$$
\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np
$$

$$
\mathrm{Var}(X) = \mathrm{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p)
$$

The moment generating function

$$
\begin{aligned}
\phi(t) = \mathbb{E}[e^{tX}] &= \mathbb{E}\left[exp\left(t \sum_{i=1}^{n} X_i\right)\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[exp(tX_i)\right] \\
&= (1 - p + pe^t)^n
\end{aligned}
$$

Check: $\phi'(t) = n(pe^t + 1 - p)^{n-1} pe^t$, $\phi''(t) = n(n-1)(pe^t + 1 - p)^{n-2}(pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$. This gives $\mathbb{E}[X] = \phi'(0) = np$ and $\mathbb{E}[X^2] = \phi''(0) = n(n-1)p^2 + np$

5.1.3 Geometric Random Variable (with parameter p)

Let X denote the number of trials until a success occurs. The probability mass function of X is given by

$$p_X(i) = p(1-p)^{i-1}$$

Hence,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{\infty} ip(1-p)^{i-1} \\
&= p\sum_{i=1}^{\infty} iq^{i-1} \qquad \text{where } q = 1-p \\
&= p\frac{d}{dq}\left(\sum_{i=1}^{\infty} q^i\right) \\
&= p\frac{d}{dq}\left(\frac{q}{1-q}\right) \\
&= \frac{p}{(1-q)^2} \\
&= 1/p
\end{aligned}$$

Now, consider the random variable $Y = X - 1$, i.e., $\mathbb{E}[Y] = \mathbb{E}[X] - 1 = 1/p - 1$.

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{i=1}^{\infty} i(i-1)p(1-p)^{i-1} \\
&= p\sum_{i=1}^{\infty} i(i-1)q^{i-1} \qquad \text{where } q = 1-p \\
&= p\frac{d}{dq}\left(\sum_{i=1}^{\infty}(i-1)q^{i}\right) \\
&= p\frac{d}{dq}\left(q^2\sum_{k=2}^{\infty}(k-1)q^{k-2}\right) \\
&= p\frac{d}{dq}\left(q^2\frac{d}{dq}\left(\sum_{k=2}^{\infty}q^{k-1}\right)\right) \\
&= p\frac{d}{dq}\left(q^2\frac{d}{dq}\left(\sum_{k=1}^{\infty}q^{k}\right)\right) \\
&= p\frac{d}{dq}\left(q^2\frac{d}{dq}\left(\frac{1}{1-q} - 1\right)\right) \\
&= p\frac{d}{dq}\left(\frac{q^2}{(1-q)^2}\right) \\
&= p\left(\frac{-2q}{(q-1)^3}\right) \qquad \text{[Backsubstitute } q = (1-p)\text{]} \\
&= p\left(\frac{-2(1-p)}{((1-p)-1)^3}\right) \\
&= p\left(\frac{-2+2p}{-p^3}\right) \\
&= \frac{2-2p}{p^2}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[XY] = \mathbb{E}[X(X-1)] &= \mathbb{E}[X^2 - X] = \mathbb{E}[X^2] - \mathbb{E}[X] \\
&= \mathrm{Var}(X) + (\mathbb{E}[X])^2 - \mathbb{E}[X] \\
\Rightarrow \frac{2-2p}{p^2} &= \mathrm{Var}(X) + 1/p^2 - 1/p \\
\mathrm{Var}(X) &= \frac{1-p}{p^2}
\end{aligned}$$

The moment generating function

$$\phi(t) = \frac{pe^t}{1-(1-p)e^t}$$

5.1.4 Poisson Random Variable (with parameter λ)

The probability mass function of X is given by

$$p_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}$$

Hence,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\infty} \frac{ie^{-\lambda}\lambda^i}{i!} \\
&= \sum_{i=1}^{\infty} \frac{e^{-\lambda}\lambda^i}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{(k)!}}_{e^{\lambda}} \qquad \text{[where } k = i-1\text{]} \\
&= \lambda e^{-\lambda} \cdot e^{\lambda} \\
&= \lambda
\end{aligned}$$

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)+X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \mathbb{E}[X(X-1)] + \lambda.$$

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{i=0}^{\infty} \frac{i(i-1)e^{-\lambda}\lambda^i}{i!} \\
&= \sum_{i=2}^{\infty} \frac{e^{-\lambda}\lambda^i}{(i-2)!} \\
&= \lambda^2 e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} \\
&= \lambda^2 e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{e^{\lambda}} \qquad \text{[where } k = i-2\text{]} \\
&= \lambda^2 e^{-\lambda} \cdot e^{\lambda} \\
&= \lambda^2
\end{aligned}$$

Hence, $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \lambda - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

The moment generating function

$$\phi(t) = E[e^{tX}] = \sum_{i=0}^{\infty} e^{ti} \frac{\lambda^i e^{-\lambda}}{i!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda e^t)^i}{i!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}$$

Check: $\phi'(t) = \lambda e^t e^{\lambda(e^t-1)}$, $\phi''(t) = (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)}$. This gives $\mathbb{E}[X] = \phi'(0) = \lambda$ and $\mathbb{E}[X^2] = \phi''(0) = \lambda^2 + \lambda$

Poisson theorem: If $n \to \infty$ and $p \to 0$ such that $np \to \lambda$ then

$$\binom{n}{i} p^i q^{n-i} \xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^i}{i!}$$

This shows that for large n and small p we can approximate the binomial distribution with Poisson distribution.

5.2 Continuous Random Variable

5.2.1 Uniform Random Variable

Let X is uniformly distributed over (a,b). The probability density function is given by

$$\begin{aligned} f_X(x) &= \frac{1}{b-a}, \text{ for } a < x < b \\ &= 0 \qquad \text{, otherwise} \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{1}{b-a}\int_a^b x dx \\ &= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\ &= \frac{a+b}{2} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx &= \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{1}{b-a}\int_a^b x^2 dx \\ &= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b \\ &= \frac{1}{3}(a^2+ab+b^2) \end{aligned}$$

Hence, $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{12}(b-a)^2$.
The moment generating function

$$\phi(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

5.2.2 Exponential Random Variable with parameter λ

The probability density function is given by

$$f_X(x) = \lambda e^{\lambda x}, \text{ for } x > 0$$
$$= 0 \text{ , for } x < 0$$

Hence,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x} dx$$
$$= -xe^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x} dx \qquad \text{[using integration by parts]}$$
$$= 0 - \frac{e^{-\lambda x}}{\lambda}|_0^\infty$$
$$= \frac{1}{\lambda}$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \qquad \text{[use integration by parts twice]}$$

Hence, $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{\lambda^2}$

The moment generating function

$$\phi(t) = \mathbb{E}[e^{tX}]$$
$$= \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx$$
$$= \lambda \int_0^\infty e^{(t-\lambda)x} dx$$
$$= \frac{\lambda}{\lambda - t}$$

Check: $\phi'(t) = \lambda/(\lambda - t)^2$ and $\phi''(t) = 2\lambda/(\lambda - t)^3$. Hence, $\mathbb{E}[X] = \phi'(0) = 1/\lambda$ and $\mathbb{E}[X^2] = \phi''(0) = 2/\lambda^2$.

Properties of the Exponential Distribution

- The exponential random variable X is memoryless, i.e.,

$$P(X > t + \tau | X > t) = P(X > \tau) \qquad \forall t, \tau \geq 0$$

Proof:

$$P(X > t + \tau | X > t) = \frac{P(X > t + \tau, X > t)}{P(X > t)}$$
$$= \frac{P(X > t + \tau)}{P(X > t)}$$
$$= \frac{e^{-\lambda(t+\tau)}}{e^{-\lambda t}}$$
$$= e^{-\lambda \tau} = P(X > \tau)$$

- X_1 and X_2 are independent exponential random variables with parameters λ_1 and λ_2, respectively. Then

$$\begin{aligned}
P(X_1 < X_2) &= \int_0^\infty P(X_1 < X_2 | X_1 = x)\lambda_1 e^{-\lambda_1 x} dx \\
&= \int_0^\infty P(X_2 > x)\lambda_1 e^{-\lambda_1 x} dx \\
&= \int_0^\infty [1 - P(X_2 \leq x)]\, \lambda_1 e^{-\lambda_1 x} dx \\
&= \int_0^\infty [1 - F_{X_2}(x)]\, \lambda_1 e^{-\lambda_1 x} dx \\
&= \int_0^\infty e^{-\lambda_2 x}\lambda_1 e^{-\lambda_1 x} dx \\
&= \int_0^\infty \lambda_1 e^{-(\lambda_1+\lambda_2)x} dx \\
&= \frac{\lambda_1}{\lambda_1 + \lambda_2}
\end{aligned}$$

- $X_1, X_2, \ldots, X_n$ are independent exponential distributed random variables with parameters $\lambda_i, i = 1, 2, \ldots, n$.

$$\begin{aligned}
P[\min(X_1, X_2, \ldots, X_n) > x] &= P(X_1 > x, X_2 > x, \ldots, X_n > x) \\
&= \prod_{i=1}^n P(X_i > x) \\
&= \prod_{i=1}^n [1 - P(X_i \leq x)] \\
&= \prod_{i=1}^n [1 - (1 - e^{-\lambda_i x})] \\
&= \prod_{i=1}^n e^{-\lambda_i x} \\
&= \exp\left[-\left(\sum_{i=1}^n \lambda_i\right) x\right]
\end{aligned}$$

5.2.3 Gaussian Random Variable with parameters (μ, σ^2)

The probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \text{ for } -\infty < x < \infty$$

Assume $z = \frac{x-\mu}{\sigma}$. Hence,

$$\begin{aligned}\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-z^2/2} dz \\ &= \underbrace{\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz}_{=0} + \mu \cdot \underbrace{\left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz\right]}_{=1} \\ &= \mu\end{aligned}$$

$\mathrm{Var}(X) = \sigma^2$.

The moment generating function of $Z = (X-\mu)/\sigma$

$$\begin{aligned}\phi_Z(t) = \mathbb{E}[e^{tZ}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z^2-2tz)/2} dz \\ &= e^{t^2/2} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2} dz}_{=1} \\ &= e^{t^2/2}\end{aligned}$$

This gives

$$\phi_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[e^{t(\sigma Z+\mu)}\right] = e^{t\mu}\mathbb{E}[e^{t\sigma Z}] = \exp\left[\frac{\sigma^2 t^2}{2} + \mu t\right]$$

Hence,

$$\phi'(t) = (\mu + t\sigma^2) \exp\left[\frac{\sigma^2 t^2}{2} + \mu t\right]$$

$$\phi''(t) = (\mu + t\sigma^2)^2 \exp\left[\frac{\sigma^2 t^2}{2} + \mu t\right] + \sigma^2 \exp\left[\frac{\sigma^2 t^2}{2} + \mu t\right]$$

So, $\mathbb{E}[X] = \phi'(0) = \mu$, $\mathbb{E}[X^2] = \phi''(0) = \mu^2 + \sigma^2$, $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sigma^2$.

Table 5.1: Common probability distributions: Part 1.

	Probability distribution	**pmf/pdf**	**Moment generating function**, $\phi(t)$	**Mean**	**Variance**
Discrete RV	Bernoulli with parameter p	$p_X(0) = 1 - p$ $p_X(1) = p$	$1 - p + pe^t$	p	$p(1-p)$
	Binomial with parameters (n, p)	$p_X(i) = \binom{n}{i} p^i (1-p)^{n-i}$, $i = 0, 1, \ldots, n$	$(1 - p + pe^t)^n$	np	$np(1-p)$
	Geometric with parameter p	$p_X(i) = p(1-p)^{i-1}$	$\frac{pe^t}{1-(1-p)e^t}$	$1/p$	$(1-p)/p^2$
	Poisson with parameter λ	$p_X(i) = \frac{e^{-\lambda}\lambda^i}{i!}$	$e^{\lambda(e^t-1)}$	λ	λ
Continuous RV	Uniform in the interval $[a, b]$	$f_X(x) = \frac{1}{b-a}$, for $a < x < b$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$(a+b)/2$	$(b-a)^2/12$
	Exponential with parameter λ	$f_X(x) = \lambda e^{-\lambda x}$, for $x > 0$	$\frac{\lambda}{\lambda - t}$	$1/\lambda$	$1/\lambda^2$
	Gaussian with parameters (μ, σ^2)	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$, for $-\infty < x < \infty$	$\exp\left[\frac{\sigma^2 t^2}{2} + \mu t\right]$	μ	σ^2

Example 1.

Seven fair coins are flipped. What is the probability that the outcomes are two heads and five tails?

Denote the random variable X as the number of heads (successes) obtained.

Hence, X is binomial with $n = 7$ and $p = 1/2$.

So,

$$P(X = 2) = \binom{7}{2}(1/2)^2(1 - 1/2)^5 \approx 0.1641$$

Example 2.

An aircraft engine fails with probability $1 - p$ during a flight independent of other engines.

The plane can fly if at least half of its engines are running.

What can you say about p if the the engineer says two-engine plane is safer than a four-engine one?

Let us denote the number of engines running during a flight for a four-engine plane by X_4 and for a two-engine plane by X_2. Note that X_4 is binomial with parameters $n = 4$ and p and X_2 is binomial with parameters $n = 2$ and p.

Hence, the probability that a four-engine plane will complete its flight

$$\begin{aligned}
P(X_4 \geq 2) &= P(X_4 = 2) + P(X_4 = 3) + P(X_4 = 4) \\
&= \binom{4}{2}p^2(1-p)^2 + \binom{4}{3}p^3(1-p)^1 + \binom{4}{4}p^4(1-p)^0 \\
&= 6p^2(1-p)^2 + 4p^3(1-p) + p^4
\end{aligned}$$

Similarly, the probability that a two-engine plane will complete its flight

$$\begin{aligned}
P(X_2 \geq 1) &= P(X_4 = 1) + P(X_4 = 2) \\
&= \binom{2}{1}p(1-p) + \binom{2}{2}p^2(1-p)^0 \\
&= 2p(1-p) + p^2
\end{aligned}$$

Hence, the two-engine plane is safer than a four-engine one if

$$\begin{aligned}
&P(X_2 \geq 1) \geq P(X_4 \geq 2) \\
&2p(1-p) + p^2 \geq 6p^2(1-p)^2 + 4p^3(1-p) + p^4 \\
&3p^3 - 8p^2 + 7p - 2 \leq 0 \\
&(p-1)^2(3p-2) \leq 0 \\
&3p - 2 \leq 0 \qquad [\text{since } p \neq 1] \\
&p \leq 2/3
\end{aligned}$$

Example 3.

The probability that a traffic signal will malfunction is 0.01. Calculate the probability that in a city with 100 traffic signals five or more will malfunction.

The random variable X denotes the number of malfunctioning traffic signals.

Hence, X is binomial with parameters $n = 100$ and $p = 0.01$ (i.e., n large, p small).

Using the Poisson approximation of binomial, X is approximately Poisson distributed with parameter $\lambda = np = 1$.

Hence,

$$\begin{aligned} P(X \geq 5) = 1 - P(X < 5) \approx 1 - \sum_{i=0}^{4} \frac{\lambda^i}{i!} e^{-\lambda} \\ &= 1 - e^{-1} \left[1 + \frac{1}{1!} + \frac{1^2}{2!} + \frac{1^3}{3!} + \frac{1^4}{4!}\right] \\ &= 0.0037 \end{aligned}$$

Example 4.

In a fast-food joint, during rush-hour customer arrives at a rate of α per minute. It is given that the arrival of the customer during a time period is Poisson distributed. Find the probabilities that there are no customers and more than 10 customers in T minutes during rush-hour.

Denote the number of customers by X in T minutes during rush-hour.

Hence, X is Poisson distributed with parameter $\lambda = \alpha T$.

So, the probability that there are no customers in T minutes during rush-hour

$$P(X = 0) = \frac{(\alpha T)^0}{0!} e^{-\alpha T} = e^{-\alpha T}$$

The probability that there are more than 10 customers in T minutes during rush-hour

$$P(X \geq 10) = 1 - P(X < 10) = 1 - \sum_{i=0}^{10} \frac{(\alpha T)^i}{i!} e^{-\alpha T}$$

Example 5.

$X_i, i = 1, \ldots, 10$ are independent Poisson random variables with mean 1.

Get a bound on $P\left(\sum_{i=1}^{10} X_i \geq 15\right)$.

Using Markov inequality,

$$\begin{aligned} P\left(\sum_{i=1}^{10} X_i \geq 15\right) &\leq \frac{\mathbb{E}[\sum_{i=1}^{10} X_i]}{15} \\ &= \frac{\sum_{i=1}^{10} \mathbb{E}[X_i]}{15} \\ &= \frac{10 \cdot 1}{15} = \frac{2}{3} \end{aligned}$$

Example 6.

X and Y are independent binomial random variables with parameters (n, p) and (m, p), respectively. Show that $X + Y$ is also binomial with $(n + m, p)$.

$$
\begin{aligned}
P(X+Y=k) &= \sum_{i=0}^{k} P(X=i, Y=k-i) \\
&= \sum_{i=0}^{k} P(X=i)P(Y=k-i) \qquad \text{[by independence]} \\
&= \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m-k+i} \\
&= p^k (1-p)^{n+m-k} \sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i} \\
&= \binom{n+m}{k} p^k (1-p)^{n+m-k}
\end{aligned}
$$

Example 7.

X and Y are independent Poisson random variables with parameters λ_1 and λ_2, respectively. Show that $X+Y$ is also Poisson with mean $\lambda_1+\lambda_2$.

$$
\begin{aligned}
P(X+Y=k) &= \sum_{i=0}^{k} P(X=i, Y=k-i) \\
&= \sum_{i=0}^{k} P(X=i)P(Y=k-i) \qquad \text{[by independence]} \\
&= \sum_{i=0}^{k} e^{-\lambda_1} \frac{\lambda_1^i}{i!} e^{-\lambda_2} \frac{\lambda_2^{(k-i)}}{(k-i)!} \\
&= e^{-(\lambda_1+\lambda_2)} \sum_{i=0}^{k} \frac{\lambda_1^i \lambda_2^{(k-i)}}{i!(k-i)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda_1^i \lambda_2^{(k-i)} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} (\lambda_1+\lambda_2)^k
\end{aligned}
$$

Example 8.

X and Y are independent exponential random variables with parameters λ and (m,p), respectively. Estimate the probability density of $Z=X+Y$.

$$
\begin{aligned}
F_Z(z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-y} f_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty} F_X(z-y)f_Y(y)dy \\
f_Z(z) = \frac{d}{dz}F_Z(z) &= \frac{d}{dz}\int_{-\infty}^{z} F_X(z-y)f_Y(y)dy \\
&= \int_{-\infty}^{\infty} \frac{d}{dz}F_X(z-y)f_Y(y)dy \\
&= \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy
\end{aligned}
$$

Hence,

$$
\begin{aligned}
f_Z(z) &= \int_0^z \lambda^2 e^{-\lambda(z-y)}e^{-\lambda y}dy \qquad 0 < y < z \\
&= \int_0^z \lambda^2 e^{-\lambda z}dy \\
&= \lambda^2 z e^{-\lambda z} \\
&= \frac{(\lambda z)^{2-1}}{\Gamma(2)}\lambda e^{-\lambda z}
\end{aligned}
$$

So, $X + Y \sim Gamma(\lambda, 2)$.

Example 9.

X and Y are independent uniform random variables on $(0, 1)$. Estimate the probability density of $Z = X + Y$.

The probability density of X and Y are

$$
f_X(x) = f_Y(y) = \begin{cases} 1, & 0 < x, y < 1 \\ 0, & \text{otherwise} \end{cases}
$$

Hence,

$$
f_Z(z) = \int_0^1 f_X(z-y)f_Y(y)dy = \int_0^1 f_X(z-y)dy
$$

For $0 \le z \le 1$,

$$
f_Z(z) = \int_0^z dy = z
$$

For $1 < z < 2$,

$$
f_Z(z) = \int_{z-1}^1 dy = 2 - z
$$

This gives a triangular density

$$f_Z(z) = \begin{cases} z, & 0 \le z \le 1 \\ 2 - z, & 1 < z < 2 \\ 0, & \text{otherwise} \end{cases}$$

Example 10.

Order statistics $X_1, X_2, \ldots, X_n$ are independent and identically distributed with CDF $F(x)$ and pdf $f(x)$. If $X_{(i)}$ is the ith smallest RV then determine the pdf of $X_{(i)}$.

$$\begin{aligned}
F_{X_{(i)}}(x) = P(X_{(i)} \le x) &= \sum_{k=i}^{n} [F(x)]^k [1 - F(x)]^{n-k} \\
\Rightarrow \quad f_{X_{(i)}}(x) = \frac{d}{dx} F_{X_{(i)}}(x) &= f(x) \sum_{k=i}^{n} \binom{n}{k} k [F(x)]^{k-1} [1 - F(x)]^{n-k} \\
&\quad - f(x) \sum_{k=i}^{n} \binom{n}{k} (n-k) [F(x)]^n [1 - F(x)]^{n-k-1} \\
&= f(x) \sum_{k=i}^{n} \frac{n!}{(n-k)!k!} k [F(x)]^{k-1} [1 - F(x)]^{n-k} \\
&\quad - f(x) \sum_{k=i}^{n} \frac{n!}{(n-k)!k!} (n-k) [F(x)]^k [1 - F(x)]^{n-k-1} \\
&= f(x) \sum_{k=i}^{n} \frac{n!}{(n-k)!(k-1)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} \\
&\quad - f(x) \sum_{k=i}^{n} \frac{n!}{(n-k-1)!k!} [F(x)]^k [1 - F(x)]^{n-k-1} \\
&= f(x) \sum_{k=i}^{n} \frac{n!}{(n-k)!(k-1)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} \\
&\quad - f(x) \sum_{j=i+1}^{n} \frac{n!}{(n-j)!(j-1)!} [F(x)]^{j-1} [1 - F(x)]^{n-j} \\
&= \frac{n!}{(n-i)!(i-1)!} f(x) [F(x)]^{i-1} [1 - F(x)]^{n-i}
\end{aligned}$$

Example 11.

If $Z_1, Z_2, \ldots, Z_n$ are standard Gaussian random variables (i.e., with zero mean and standard deviation 1) and the random variables $X_1, X_2, \ldots, X_m$ are given by

$$\begin{aligned}
X_1 &= a_{11}Z_1 + \cdots + a_{1n}Z_n + \mu_1 \\
X_2 &= a_{21}Z_1 + \cdots + a_{2n}Z_n + \mu_2 \\
&\vdots \\
X_i &= a_{i1}Z_1 + \cdots + a_{in}Z_n + \mu_i \\
&\vdots \\
X_m &= a_{m1}Z_1 + \cdots + a_{mn}Z_n + \mu_m
\end{aligned}$$

i.e., $\mathbf{X} = \mathbf{AZ} + \boldsymbol{\mu}$ where $\mathbf{X} = [X_1, X_2, \ldots, X_m]^T$, $\mathbf{A} = [a_{ij}]$, $Z = [Z_1, Z_2, \ldots, Z_n]^T$, and $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_m]$.

Hence,

$$\begin{aligned}
\mathbb{E}[X_i] &= \mu_i \\
\text{Var}(X_i) &= \sum_{j=1}^{n} a_{ij}^2
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\mathbf{X}] &= \boldsymbol{\mu} \\
Cov(\mathbf{X}) &= \mathbf{AA}^T
\end{aligned}$$

In general, when $\mathbf{Y} = \mathbf{AX}$ with $Cov(\mathbf{X}) = \mathbf{\Sigma}$

$$\begin{aligned}
Cov(\mathbf{Y}) = Cov(\mathbf{AX}) &= \mathbb{E}[(\mathbf{A}X - \mathbb{E}[\mathbf{AX}])(\mathbf{AX} - \mathbb{E}[\mathbf{AX}])^T] \\
&= \mathbb{E}[(\mathbf{AX} - \mathbf{A}\mathbb{E}[\mathbf{X}])(\mathbf{AX} - \mathbf{A}\mathbb{E}[X])^T] \\
&= \mathbb{E}[\mathbf{A}(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T\mathbf{A}^T] \\
&= \mathbf{A}\,\underbrace{\mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]}_{=\Sigma}\,\mathbf{A}^T \\
&= \mathbf{A\Sigma A}^T
\end{aligned}$$

Chapter 6

Common Probability Distributions: Part 2

6.1 Discrete Random Variables

6.1.1 Hypergeometric Distribution

Let the random variable X denote the number of acceptable items among n selected items from a pool of N acceptable and M unacceptable items. Then the probability mass function is given by

$$P(X = i) = p_X(i) = \frac{\binom{N}{i}\binom{M}{n-i}}{\binom{M+N}{n}}, \qquad i = 0, 1, \ldots, \min(n, N)$$

We can write $X = \sum_{j=0}^{n} X_i$ where

$$X_j = \begin{cases} 1, & \text{if } j\text{th selected item is acceptable} \\ 0, & \text{if } j\text{th selected item is unacceptable} \end{cases}$$

and $\mathbb{E}[X_i] = P(X_i = 1) = \frac{N}{M+N}$

The mean of X is

$$\mathbb{E}[X] = \sum_{j=0}^{n} \mathbb{E}[X_i] = \sum_{j=0}^{n} \frac{N}{M+N} = \frac{nN}{M+N}$$

$$\mathrm{Var}(X) = \frac{nN}{M+N}\left(1 - \frac{N}{M+N}\right)\left(1 - \frac{n-1}{M+N-1}\right)$$

6.1.2 Negative Binomial Distribution

If X denotes the number of trials needed to obtain r successes X has a probability mass function

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

where each trial results in a success with a probability p.

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

6.2 Continuous Random Variables

6.2.1 Gamma Distribution

The probability density function is given by

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

where the Gamma function is defined by

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

For integer α, $\Gamma(\alpha) = (\alpha - 1)!$.

The mean of the Gamma distribution

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty x f(x) dx \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty x(\lambda x)^{\alpha-1} \lambda e^{-\lambda x} dx \\
&= \frac{1}{\lambda\Gamma(\alpha)} \int_0^\infty (\lambda x)^{\alpha} \lambda e^{-\lambda x} dx \\
&= \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} \int_0^\infty \frac{(\lambda x)^{\alpha}}{\Gamma(\alpha+1)} \lambda e^{-\lambda x} \\
&= \frac{\Gamma(\beta)}{\lambda\Gamma(\beta-1)} \int_0^\infty \frac{(\lambda x)^{\beta-1}}{\Gamma(\beta)} \lambda e^{-\lambda x} \qquad [\beta = \alpha + 1] \\
&= \frac{(\beta-1)\Gamma(\beta-1)}{\lambda\Gamma(\beta-1)} \cdot 1 \\
&= \frac{\beta-1}{\lambda} = \frac{\alpha}{\lambda}
\end{aligned}$$

Similarly,

$$
\begin{aligned}
\mathbb{E}[X^2] &= \int_0^\infty x^2 f(x)dx \\
&= \frac{1}{\Gamma(\alpha)} \int_0^\infty x^2 (\lambda x)^{\alpha-1} \lambda e^{-\lambda x} dx \\
&= \frac{1}{\lambda^2 \Gamma(\alpha)} \int_0^\infty (\lambda x)^{\alpha+1} \lambda e^{-\lambda x} dx \\
&= \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} \int_0^\infty \frac{(\lambda x)^{\alpha+1}}{\Gamma(\alpha+2)} \lambda e^{-\lambda x} \\
&= \frac{\Gamma(\beta)}{\lambda^2 \Gamma(\beta-2)} \int_0^\infty \frac{(\lambda x)^{\beta-1}}{\Gamma(\beta)} \lambda e^{-\lambda x} \qquad [\beta = \alpha + 2] \\
&= \frac{(\beta-1)(\beta-2)\Gamma(\beta-2)}{\lambda^2 \Gamma(\beta-2)} \cdot 1 \\
&= \frac{(\beta-1)(\beta-2)}{\lambda^2} \\
&= \frac{\alpha(\alpha+1)}{\lambda^2}
\end{aligned}
$$

Hence,

$$
\mathrm{Var}(X) = \mathbb{E}[X^2] - \left(\mathbb{E}[X]^2\right)^2 = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}
$$

The moment generating function is given by

$$
\begin{aligned}
\phi(t) &= \mathbb{E}[e^{tX}] \\
&= \left(\frac{\lambda}{\lambda - t}\right)^\alpha
\end{aligned}
$$

Check:

$$
\begin{aligned}
\phi'(t) &= \frac{d}{dt}\phi(t) = \frac{\alpha \lambda^\alpha}{(\lambda - t)^{\alpha+1}} \\
\phi'(0) &= \mathbb{E}[X] = \alpha/\lambda \\
\phi''(t) &= \frac{d^2}{dt^2}\phi(t) = \frac{\alpha(\alpha+1)\lambda^\alpha}{(\lambda - t)^{\alpha+2}} \\
\phi''(0) &= \mathbb{E}[X^2] = \frac{\alpha(\alpha+1)}{\lambda^2}
\end{aligned}
$$

Incomplete Gamma Function

Incomplete Gamma functions are defined as

$$
I_G(u, \alpha) = \frac{\int_0^u y^{k-1} e^{-y} dy}{\Gamma(\alpha)}
$$

Using $y = \alpha x$,

$$\begin{aligned}
P(a < X \leq b) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_a^b x^{\alpha-1} e^{-\lambda x} dx \\
&= \frac{1}{\Gamma(\alpha)} \left[\int_0^{\lambda b} y^{\alpha-1} e^{-y} dy - \int_0^{\lambda a} y^{\alpha-1} e^{-y} dy \right] \\
&= I_G(\lambda b, \alpha) - I_G(\lambda a, \alpha)
\end{aligned}$$

6.2.2 Beta Distribution

The Beta function is defined as

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\ dx, \quad \alpha, \beta > 0$$

This is related to Gamma function

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

The Beta random variable has a probability density function given by

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\int_0^1 x^{\alpha}(1-x)^{\beta-1}\ dx}{B(\alpha, \beta)} \\
&= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \\
&= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\alpha}{\alpha+\beta} \qquad\qquad [\Gamma(x+1) = x\Gamma(x)]
\end{aligned}$$

Similarly,

$$\begin{aligned}
\mathbb{E}[X^2] &= \frac{B(\alpha+2, \beta)}{B(\alpha, \beta)} \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}
\end{aligned}$$

Hence,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

6.2.3 Rayleigh Distribution

The probability density function is given by

$$f(x) = \begin{cases} \frac{x}{\sigma^2} e^{-x^2/2\sigma^2} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The cumulative distribution function is given by

$$\begin{aligned} F(x) &= \int_0^x \frac{a}{\sigma^2} e^{-\frac{a^2}{2\sigma^2}} da \\ &= \int_0^x e^{-\frac{a^2}{2\sigma^2}} d\left(\frac{a^2}{2\sigma^2}\right) \\ &= 1 - e^{-\frac{x^2}{2\sigma^2}} \end{aligned}$$

$$\begin{aligned} E[X] &= \int_0^\infty \frac{x^2}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \int_0^\infty \sqrt{2} t e^{-t} t^{-\frac{1}{2}} \sigma dt \qquad [\text{where } t = \frac{x^2}{2r^2}] \\ &= \sqrt{2}\sigma \int_0^\infty t^{\frac{3}{2}-1} e^{-\sigma} dt \\ &= \sqrt{2}\sigma\Gamma\left(\frac{3}{2}\right) \qquad [\text{using the definition of } \Gamma(\cdot) \text{ function}] \\ &= \frac{\sigma}{\sqrt{2}}\Gamma\left(\frac{1}{2}\right) \\ &= \sigma\sqrt{\pi/2} \end{aligned}$$

Similarly, $\mathrm{Var}(X) = \frac{(4-\pi)\sigma^2}{2}$

6.2.4 Cauchy Distribution

The probability density function is given by

$$f(x) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x-\mu}{\gamma}\right)^2\right)}, \qquad \gamma > 0, -\infty < x < \infty$$

Cauchy distribution does not have mean and variance.

6.2.5 χ^2 Distribution

$Z_1, Z_2, \ldots, Z_n$ are standard Gaussian random variable (i.e., with zero mean and standard deviation 1) then the random variable $X = \sum_{i=1}^n Z_i^2$ is χ squared distributed. χ-square distribution with n degrees of freedom is identical to Gamma distribution with parameters $n/2$ and $1/2$.

$$f(x) = \frac{\frac{1}{2} e^{x/2} \left(\frac{x}{2}\right)^{n/2-1}}{\Gamma(n/2)}, \quad x > 0$$

Hence, $\mathbb{E}[X] = n$ and $\mathrm{Var}(X) = 2n$.

6.2.6 Student's t Distribution

Let us define the random variable

$$X = \frac{Z}{\sqrt{\chi_n^2/n}}$$

where Z is a standard Gaussian random variable and χ_n^2 is a chi-squared random variable with n d.o.f. Then X has a Student's t distribution with n degree-of-freedom. This distribution is symmetric with $\mathbb{E}[X] = 0, n > 1$ and $\mathrm{Var}(X) = n/(n-2), n > 2$.

6.2.7 F Distribution

Define

$$X = \frac{\chi_n^2/n}{\chi_m^2/m}$$

where χ_n^2, χ_m^2 are chi-squared distributed with degrees-of-freedom n and m, respectively. Then X is F distributed with degrees-of-freedom n and m.

6.2.8 Lognormal Distribution

In lognormal distribution, the logarithm of the random variable has a normal distribution. Lognormal distribution has a probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\xi x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\xi}\right)^2\right] \qquad x \geq 0$$

The parameters ξ and λ are related to the mean and variance of the distribution.

$$\begin{aligned}
\mathbb{E}[X] = \mu_X &= \exp\left(\lambda + \frac{1}{2}\xi^2\right) \\
\mathrm{Var}(X) = \sigma_X^2 &= \mu_X^2(e^{\xi^2} - 1)
\end{aligned}$$

This gives

$$\begin{aligned}
\lambda &= 2\ln\mu_X - \frac{1}{2}\ln(\sigma_X^2 + \mu_X^2) \\
\xi^2 &= -2\ln\mu_X + \ln(\sigma_X^2 + \mu_X^2) \\
&= \ln\left[1 + \left(\frac{\sigma_X}{\mu_X}\right)^2\right]
\end{aligned}$$

Table 6.1: Common probability distributions: Part 2a.

	Probability distribution	**pmf/pdf**	**Moment generating function**, $\phi(t)$	**Mean**	**Variance**
Discrete RV	Hypergeometric	$p_X(i) = \frac{\binom{N}{i}\binom{M}{n-i}}{\binom{M+N}{n}}$ $i = 0, 1, \ldots, \min(n, N)$	—	$\frac{nN}{M+N}$	$\frac{nN}{M+N}\left(1 - \frac{N}{M+N}\right)$ $\cdot \left(1 - \frac{n-1}{M+N-1}\right)$
	Negative binomial with parameters (r, p)	$p_X(i) = \binom{i-1}{r-1}(1-p)^{i-r}p^r$, $i = r, r+1, \ldots, \infty$	$\left(\frac{1-p}{1-pe^t}\right)^r, \quad t < -\ln p$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Continuous RV	Gamma with parameters (α, λ)	$f_X(x) = \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}\lambda e^{-\lambda x}$, for $x \geq 0$	$\left(\frac{\lambda}{\lambda - t}\right)^\alpha$	α/λ	α/λ^2
	Beta with parameters (α, β)	$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$, for $\alpha, \beta > 0$	$1 + \sum_{k=1}^{\infty} \frac{t^k}{k!} \prod_{n=1}^{k-1} \frac{\alpha+n}{\alpha+\beta+n}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
	Rayleigh with parameter σ^2	$f_X(x) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}$, for $x \geq 0$	$1 + \sigma t e^{\sigma^2 t^2/2} \sqrt{\pi/2} \left(\text{erf}\left(\frac{\sigma t}{\sqrt{2}}\right) + 1\right)$ †	$\sigma\sqrt{\pi/2}$	$\frac{(4-\pi)\sigma^2}{2}$

† $\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$

Table 6.2: Common probability distributions: Part 2b.

	Probability distribution	**pmf/pdf**	**Moment generating function**, $\phi(t)$	**Mean**	**Variance**
Continuous RV	Cauchy with parameters (μ, γ)	$f_X(x) = \frac{1}{\pi\gamma\left(1+\left(\frac{x-\mu}{\gamma}\right)^2\right)}$, for $\gamma > 0, -\infty < x < \infty$	—	—	—
	χ^2 with d.o.f. n	$f_X(x) = \frac{\frac{1}{2}e^{-x/2}\left(\frac{x}{2}\right)^{n/2-1}}{\Gamma(n/2)}$, for $\alpha, \beta > 0$	$(1-2t)^{-n/2}$ $t < 1/2$	n	$2n$
	Student's t with parameter ν	$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$, for $\nu > 0$	—	0 for $\nu > 1$	0 for $\nu > 2$
	F with d.o.f.s n, m	$f_X(x) = cx^{n/2-1}\left(1+\frac{n}{m}x\right)^{-(n+m)/2}$, † for $x \in [0, \infty)$ and $n, m > 0$	—	$\frac{m}{m-2}$ $m > 2$	$\frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ $m > 4$
	Lognormal with parameters (λ, ξ)	$f_X(x) = \frac{1}{\sqrt{2\pi}\xi x}\exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\xi}\right)^2\right]$, for $x \geq 0$	—	μ_X $= e^{\left(\lambda + \frac{1}{2}\xi^2\right)}$	$\mu_X^2(e^{\xi^2}-1)$

† $c = \left(\frac{n}{m}\right)^{n/2}\frac{1}{B\left(\frac{n}{2}, \frac{m}{2}\right)}$

Example 1.

If Gamma random variable X with mean 15 psf and coefficient of variation (COV) 25% is used to represent loads on building then what is the probability that the load will exceed 25 psf?

$$\begin{aligned} \mathbb{E}[X] &= \frac{\alpha}{\lambda} = 15 \\ COV = \frac{\sqrt{\text{Var}(X)}}{\mathbb{E}[X]} &= \frac{\sqrt{\alpha}/\lambda}{\alpha/\lambda} = \frac{1}{\sqrt{\alpha}} = 0.25 \\ \alpha &= 16 \\ \lambda &= \frac{16}{15} = 1.067 \end{aligned}$$

Hence,

$$\begin{aligned} P(X > 25) &= 1 - P(X \leq 25) \\ &= 1 - I_G(\lambda \cdot 25, \alpha) \\ &= 1 - I_G(26.67, 16) \\ &= 1 - 0.671 = 0.329 \end{aligned}$$

Example 2.

Four earthquakes in last 50 years with magnitude more than 7.

Modeling the occurrences as Bernoulli random variable with $p = 4/50 = 0.08$, the probability that at least one earthquake will occur in 20 years

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \binom{20}{0}(0.08)^0(0.92)^20 \\ &= 0.811 \end{aligned}$$

Modeling the occurrences as Poisson process with rate $\nu = 4/50 = 0.08$. Hence, at least one earthquake will occur in the next 20 years with probability

$$\begin{aligned} P(X_{20} \geq 1) &= 1 - P(X_{20} = 0) \\ &= 1 - \frac{(0.08 \times 20)^0}{0!}e^{-0.08\times 20} \\ &= 0.798 \end{aligned}$$

Example 3.

In the last 125 years 16 earthquakes with a magnitude larger than 6 occurred. If the occurrence of the earthquakes are Poisson distributed what is the probability that the one will occur in the next 2 tears?

The rate $\nu = 16/125 = 0.128$ earthquakes/year. Define X_t as the number of earthquakes in the next t years and T_n as the time up to when nth earthquake occurs. Hence, X_t is Poisson distributed and T_n is exponentially distributed. This can be used to write

$$\begin{aligned} P(X_2 \geq 1) &= 1 - P(X_2 = 0) \\ &= 1 - \frac{(0.128 \times 2)^0}{0!} e^{-0.128 \times 2} \\ &= 0.226 \end{aligned}$$

This is equivalent to

$$P(T_1 \leq 2) = 1 - e^{-0.128 \times 2} = 0.226$$

What is the probability that there are no earthquakes in the next 10 years?

$$P(X_{10} = 0) = \frac{(0.128 \times 10)^0}{0!} e^{-0.128 \times 10} = 0.278$$

This is equivalent to

$$P(T_1 > 10) = e^{-0.128 \times 10} = 0.278$$

Chapter 7

Sampling Statistics

Assume the population has mean μ and variance σ^2. n samples from this population are $X_1, X_2, \ldots, X_n$.

7.1 Sample Mean

Define the sample mean as

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

This estimator of population mean is unbiased.

$$\begin{aligned}
\mathbb{E}[\bar{X}] &= \mathbb{E}\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] \\
&= \frac{1}{n}\left(\mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n]\right) \\
&= \frac{1}{n} \cdot n\mu \\
&= \mu
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) \\
&= \frac{1}{n^2}\left[\sum_{i=1}^{n} \mathrm{Var}(X_i)\right] \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

7.2 Sample Variance

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

This estimator of population variance is unbiased.

$$\begin{aligned}
(n-1)\mathbb{E}[S^2] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] - n\mathbb{E}[\bar{X}^2] \\
&= n\mathbb{E}[X_i^2] - n\mathbb{E}[\bar{X}^2] \\
&= n\left[\text{Var}(X_i) + (\mathbb{E}[X_i])^2\right] - n\left[\text{Var}(\bar{X}) + (\mathbb{E}[\bar{X}^2])^2\right] \\
&= n\sigma^2 + n\mu^2 - n\left(\frac{\sigma^2}{n}\right) - n\mu^2 \\
&= (n-1)\sigma^2
\end{aligned}$$

Hence, $\mathbb{E}[S^2] = \sigma^2$.

7.3 Central Limit Theorem

$X_1, X_2, \ldots, X_n$ is a sequence of independent and identically distributed random variables with mean μ and variance σ^2. The distribution of $\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ tends to standard normal as $n \to \infty$.

Hence, for the sample mean $\bar{X}$: $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is approximately a standard normal random variable.

If the samples are from a normal population then $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is a standard normal random variable, $(n-1)S^2/\sigma^2$ is a χ^2 random variable with $n-1$ degree-of-freedom and $\sqrt{n}(\bar{X}-\mu)/S$ has a t distribution with $n-1$ degree-of-freedom.

Example 1.

In bags of potatoes from a certain company the weight is written as normally distributed with mean 1.5 lb. with a standard deviation of 0.25 lb.

(a) What is the probability that the potato bag you picked to buy weighs more than 1.7 lb.?

Denote the weight of the randomly picked bag by X.

$$\begin{aligned}
P(X > 1.7) &= P\left(\frac{X-1.5}{0.25} > \frac{1.7-1.5}{0.25}\right) \\
&= P(Z > 0.8) \\
&= 1 - \Phi(0.8) \\
&= 1 - 0.7881 = 0.2119
\end{aligned}$$

(b) What is the probability that the potato bag you picked to buy weighs in between 1.3lb. and 1.7 lb.?

$$
\begin{aligned}
P\left(1.3 < X \leq 1.7\right) &= P\left(\frac{1.3-1.5}{0.25} < \frac{X-1.5}{0.25} \leq \frac{1.7-1.5}{0.25}\right) \\
&= P(-0.8 < Z \leq 0.8) \\
&= P(Z \leq 0.8) - P(Z \leq -0.8) \\
&= \Phi(0.8) - \Phi(-0.8) \\
&= 2\Phi(0.8) - 1 \\
&= 2 \times 0.7881 - 1 = 0.5762
\end{aligned}
$$

(c) If you pick 25 bags and observe their average weight what is the probability that their mean is more than 1.55 lb.?

$$
\begin{aligned}
P\left(\bar{X} > 1.7\right) &= P\left(\frac{\bar{X}-1.5}{0.25/\sqrt{25}} > \frac{1.55-1.5}{0.25/\sqrt{25}}\right) \\
&= P(Z > 1) \\
&= 1 - \Phi(1) \\
&= 1 - 0.8413 = 0.1587
\end{aligned}
$$

(d) What is the probability that their mean is in between 1.45 lb. and 1.55 lb.?

$$
\begin{aligned}
P\left(1.45 < X \leq 1.55\right) &= P\left(\frac{1.45-1.5}{0.25} < \frac{X-1.5}{0.25/\sqrt{25}} \leq \frac{1.55-1.5}{0.25/\sqrt{25}}\right) \\
&= P(-1 < Z \leq 1) \\
&= P(Z \leq 1) - P(Z \leq -1) \\
&= \Phi(1) - \Phi(-1) \\
&= 2\Phi(1) - 1 \\
&= 2 \times 0.8413 - 1 = 0.6826
\end{aligned}
$$

Example 2.

Team A in a cricket match scores with mean 250 and standard deviation 25. If you watch 10 such games what is the probability that the sample variance calculated using those 10 games will exceed 30?

$n = 10, \sigma^2 = 625$. Hence, $\frac{(n-1)S^2}{\sigma^2} = \frac{9S^2}{625}$.

$$
\begin{aligned}
P(S^2 > 900) &= P\left(\frac{9S^2}{625} > \frac{9}{625} \cdot 900\right) \\
&= P\left(\chi^2_9 > 12.96\right) \\
&= 1 - P\left(\chi^2_9 \leq 12.96\right) \\
&= 1 - 0.8356 = 0.1644
\end{aligned}
$$

Chapter 8

Parameter Estimation

Random samples from a probability distribution $F(x)$ are: $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$. The probability distribution has a parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_m]^T$.

Estimator: Statistic used to estimate unknown $\boldsymbol{\theta}$.

Estimate: Observed value of the estimator.

8.1 Maximum Likelihood Estimator

The likelihood for independent samples $\mathbf{x}$ is defined as

$$L(\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$$

The maximum likelihood estimator is defined as

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} L(\mathbf{x}; \boldsymbol{\theta})$$

To estimate the value of $\boldsymbol{\theta}$ that maximizes L or equivalently $\ln L$ we will set

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \qquad \text{for } i = 1, 2, \ldots, m$$

Example 1.

For Bernoulli distribution,

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

Hence, among n observations, the likelihood is defined as

$$\begin{aligned} L(\mathbf{x}; p) &= \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i} \\ &= p^{\sum_1^n x_i} (1-p)^{\sum_1^n 1 - x_i} \\ &= p^{n\bar{x}} (1-p)^{n(1-\bar{x})} \end{aligned}$$

The log-likelihood is

$$\ln L = n\bar{x} \ln p + n(1-\bar{x}) \ln(1-p)$$

Taking derivative with respect to the parameter p

$$\begin{aligned}\frac{d\ln L}{dp} &= \frac{n\bar{x}}{p} - \frac{n(1-\bar{x})}{1-p} = 0\\ (1-p)\bar{x} &- (1-\bar{x})p = 0\\ \Rightarrow \quad \hat{p} &= \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}\end{aligned}$$

Hence, the ML estimator is $\hat{p} = \bar{x}$

Example 2.

For Poisson distribution

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

Hence, among n observations, the likelihood is defined as

$$\begin{aligned}L(\mathbf{x}; p) &= \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda)\\ &= \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \exp(-n\lambda)\end{aligned}$$

The log-likelihood is

$$\ln L = n\bar{x}\ln\lambda - n\lambda - \sum_{i=1}^{n} \ln(x_i!)$$

Taking derivative with respect to the parameter λ

$$\begin{aligned}\frac{d\ln L}{d\lambda} &= \frac{n\bar{x}}{\lambda} - n = 0\\ \Rightarrow \quad \hat{\lambda} &= \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}\end{aligned}$$

Hence, the ML estimator is $\hat{\lambda} = \bar{x}$.

Example 3.

For Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Hence, among n observations, the likelihood is defined as

$$\begin{aligned}L(\mathbf{x}; \mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]\\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\right]\end{aligned}$$

The log-likelihood is

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

Taking derivative with respect to the parameter μ

$$\frac{\partial \ln L}{\partial \mu} = -\sum_{i=1}^{n}\frac{(x_i-\mu)}{\sigma^2} = 0$$
$$\Rightarrow \quad \hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Hence, the ML estimator is $\hat{\mu} = \bar{x}$.

Taking derivative with respect to the parameter σ^2

$$\frac{\partial \ln L}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^4} = 0$$
$$\Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=0}^{n}(x_i-\hat{\mu})^2$$

Hence, the ML estimator is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=0}^{n}(x_i-\hat{\mu})^2$.

Example 4.

For Gamma distribution

$$f(x) = \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x}$$

Hence, among n observations, the likelihood is defined as

$$L(\mathbf{x};\alpha,\lambda) = \prod_i \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x}$$
$$= \frac{1}{\Gamma(\alpha)^n}\lambda^{n\alpha}\left(\prod_{i=1}^{n} x_i^{\alpha-1}\right)e^{-\lambda\sum_{i=1}^{n} x_i}$$

The log-likelihood is

$$\ln L = (\alpha-1)\sum_{i=1}^{n}\ln x_i - \lambda\sum_{i=1}^{n} x_i + (n\alpha)\ln\lambda - n\ln\Gamma(\alpha)$$

Taking derivative with respect to the parameter λ

$$\frac{\partial \ln L}{\partial \lambda} = -\sum_{i=1}^{n} x_i + \frac{n\alpha}{\lambda} = 0$$
$$\Rightarrow \quad \hat{\lambda} = \frac{\hat{\alpha}}{\frac{1}{n}\sum_{i=1}^{n} x_i}$$

Hence, the ML estimator is $\hat{\lambda} = \frac{\hat{\alpha}}{\frac{1}{n}\sum_{i=1}^{n} x_i}$.

Taking derivative with respect to the parameter α

$$\frac{\partial \ln L}{\partial \alpha} = \sum_{i=1}^{n} \ln x_i + n \ln \lambda - \frac{n\Gamma'(\alpha)}{\Gamma(\alpha)} = 0$$

$$\Rightarrow \quad \ln \hat{\alpha} - \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} = \ln \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) - \frac{1}{n} \sum_{i=1}^{n} \ln x_i$$

This is a nonlinear equation needed to be solved to get $\hat{\alpha}$.

Example 5.

If the observations $\{0.3, 0.2, 0.5, 0.8, 0.9\}$ are obtained from a distribution with $f(x) = \theta x^{\theta-1}, x \geq 0$ then estimate the value of θ using Maximul Likelihood method.

The likelihood is defined as

$$L(\mathbf{x}; \theta) = \prod_{i=1}^{5} \theta x_i^{\theta-1}$$

The log likelihood is

$$\ln L = 5 \ln \theta + (\theta - 1) \sum_{i=1}^{5} \ln x_i$$

Taking derivative of $\ln L$ with respect to θ

$$\frac{\partial \ln L}{\partial \theta} = \frac{5}{\theta} + \sum_{i=1}^{5} \ln x_i = 0$$

$$\Rightarrow \quad \hat{\theta} = -\frac{5}{\sum_{i=1}^{5} \ln x_i} = 1.3038$$

Example 6.

For Uniform distribution in $(0, \theta)$

$$f(x) = \frac{1}{\theta}, \qquad 0 < x < \theta$$

Hence, among n observations, the likelihood is defined as

$$L(\mathbf{x}; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n}$$

The log-likelihood is

$$\ln L = -n \ln \theta$$

This is maximized when θ is minimum but $\theta \geq \max(x_1, x_2, \ldots, x_n)$.

Hence, the ML estimator is $\hat{\theta} = \max(x_1, x_2, \ldots, x_n)$.

8.2 Interval Estimate

Let $X_1, X_2, \ldots, X_n$ are samples from a Gaussian distribution with mean μ and variance σ^2. The point estimator $\bar{X}$ is Gaussian with mean μ and variance σ^2/n. Hence,

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$
$$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

Based on the observations, with 95% we can say that the population mean μ lies within the interval $\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$ — known as the 95 percent confidence interval estimate of μ.

In general, $100(1-\alpha)$ percent two-sided confidence interval for μ is $\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$.
One-sided upper confidence interval for μ is $\left(\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, +\infty\right)$.
One-sided lower confidence interval for μ is $\left(-\infty, \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$.

8.2.1 Sample size:

If we want the $100(1-\alpha)$ percent two-sided confidence interval for μ to be within $(\bar{x} \pm \Delta x)$ we need a sample size

$$n = \left(\frac{2z_{\alpha/2}\sigma}{\Delta x}\right)^2$$

8.2.2 Quick reference:

$100(1-\alpha)\%$ two-sided confidence interval:
90% confidence: $\alpha = 10$, $z_{\alpha/2} = 1.65$
95% confidence: $\alpha = 5$, $z_{\alpha/2} = 1.96$
98% confidence: $\alpha = 2$, $z_{\alpha/2} = 2.33$
99% confidence: $\alpha = 1$, $z_{\alpha/2} = 2.58$

Similarly, the following Table shows a variety of cases for samples from a normal population: Note that, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Table 8.1: Different cases.

Case	Parameter	Confidence interval	Lower interval	Upper interval
σ^2 known	μ	$\left(\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$	$\left(-\infty, \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$	$\left(\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, +\infty\right)$
σ^2 unknown	μ	$\left(\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$	$\left(-\infty, \bar{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}}\right)$	$\left(\bar{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}}, +\infty\right)$
μ unknown	σ^2	$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$	$\left(0, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$	$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, +\infty\right)$

Example 6.

Estimate the sample size needed for mean to be within ± 0.25 where $\sigma = 2$ and a confidence interval of 95%.

$$\begin{aligned}
n &= \left(\frac{2z_{\alpha/2}\sigma}{\Delta x}\right)^2 \\
&= \left(\frac{2\times 1.96\times 2}{0.25}\right)^2 \\
&\approx 984
\end{aligned}$$

Example 7.

The lifetime X of light bulbs are exponentially distributed. Based on observation of 81 light bulbs we obtain their average lifetime is 200 hours. Estimate the 95% confidence interval for the mean lifetime.

For exponentially distributed random variable X,

$$f(x) = \lambda e^{-\lambda x}$$

The mean of X is $1/\lambda$ and variance is $1/\lambda^2$. For large number of samples n, the sample mean is Gaussian with mean $1/\lambda$ and variance $\frac{1}{n\lambda^2}$.

Hence, we can write

$$\begin{aligned}
P\left(-1.96 < \frac{\bar{X}-\frac{1}{\lambda}}{\frac{1}{\lambda\sqrt{n}}} < 1.96\right) &= 0.95 \\
P\left(\frac{1}{\lambda} - 1.96\frac{1}{\lambda\sqrt{n}} < \bar{X} < \frac{1}{\lambda} + 1.96\frac{1}{\lambda\sqrt{n}}\right) &= 0.95 \\
P\left\{\frac{1}{\lambda}\left(1-\frac{1.96}{\sqrt{n}}\right) < \bar{X} < \frac{1}{\lambda}\left(1+\frac{1.96}{\sqrt{n}}\right)\right\} &= 0.95 \\
P\left(\frac{\bar{X}}{1+1.96/\sqrt{n}} < \frac{1}{\lambda} < \frac{\bar{X}}{1-1.96/\sqrt{n}}\right) &= 0.95
\end{aligned}$$

Hence, the 95% confidence interval for the mean lifetime of the bulbs is $\frac{200}{1+1.96/\sqrt{81}} < \frac{1}{\lambda} < \frac{200}{1-1.96/\sqrt{81}}$ or $164 < \frac{1}{\lambda} < 256$.

Example 8.

For Poisson distributed random variable get the $100(1-\alpha)$ confidence interval.

The p.m.f. is given by

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

The mean $\mathbb{E}[X] = \lambda = \mathrm{Var}(X)$. Hence, for large n $\bar{X}$ is approximately Gaussian with mean λ and variance λ/n. This helps in writing

$$P\left(\bar{X} - 1.96\sqrt{\frac{\lambda}{n}} < \lambda < \bar{X} + 1.96\sqrt{\frac{\lambda}{n}}\right) = 0.95$$

$$P\left(|\bar{X} - \lambda| < 1.96\sqrt{\frac{\lambda}{n}}\right) = 0.95$$

$$P\left\{\left(\bar{X} - \lambda\right)^2 < \frac{(1.96)^2}{n}\lambda\right\} = 0.95$$

Therefore, the confidence interval is the two solutions of the following quadratic equation

$$(\bar{x} - \lambda)^2 = \frac{(1.96)^2}{n}\lambda$$

Chapter 9

Hypothesis Testing

Random samples from a probability distribution $F(x)$ are: $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$. The probability distribution has a parameter vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_m]^T$. Hypothesis tests allow you to test some hypotheses on the unknown parameters. For the assumption $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, write the two competing hypotheses are:

Null hypothesis, $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$

Alternative hypothesis, $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$

The critical region C is used in the following:

accept H_0 if $(X_1, X_2, \ldots, X_n) \notin C$

reject H_0 if $(X_1, X_2, \ldots, X_n) \in C$

Two kinds of errors are encountered in hypothesis testing:

(i) **Type I error:** This error occurs if H_0 is true but $(X_1, X_2, \ldots, X_n) \in C$. The probability of occurring of such errors is

$$P\left\{(X_1, X_2, \ldots, X_n) \in C | H_0 \text{ is true}\right\} = \alpha$$

where α is called the significance level of the test.

(ii) **Type II error:** This error occurs if H_0 is false but $(X_1, X_2, \ldots, X_n) \notin C$. The probability of such an error is a function of $\boldsymbol{\theta}$ and is denoted by

$$P\left\{(X_1, X_2, \ldots, X_n) \notin C | H_0 \text{ is false}\right\} = \beta(\boldsymbol{\theta})$$

- The power of the test $P(\boldsymbol{\theta})$ is defined by the probability that H_0 is rejected when it is false, i.e.,

$$P(\boldsymbol{\theta}) = 1 - \beta(\boldsymbol{\theta}) = P\left\{(X_1, X_2, \ldots, X_n) \in C | H_0 \text{ is false}\right\}$$

- The most powerful test has minimum $\beta(\boldsymbol{\theta})$. In general, C for a most powerful test depends on $\boldsymbol{\theta}$ but if it is same for every $\boldsymbol{\theta} \in \Theta$ the test is called uniformly most powerful.

9.1 The Mean of a Normal Population

9.1.1 Known Variance

Two-tailed Test

Null hypothesis, $H_0 : \mu = \mu_0$

Alternate hypothesis, $H_1 : \mu \neq \mu_0$
Assume the critical region for an α significance level test is given by

$$C = \big\{(X_1, X_2, \ldots, X_n) : |\bar{X} - \mu_0| > c\big\}$$

Hence,

$$\begin{aligned} P(|\bar{X} - \mu_0| > c) &= \alpha \\ P\left(\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} > \frac{c\sqrt{n}}{\sigma}\right) &= \alpha \\ P\left(|Z| > \frac{c\sqrt{n}}{\sigma}\right) &= \alpha \\ \Rightarrow \quad z_{\alpha/2} &= \frac{c\sqrt{n}}{\sigma} \\ c &= \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \end{aligned}$$

The critical region thus becomes

$$C = \left\{(X_1, X_2, \ldots, X_n) : \left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu_0\right| > \frac{z_{\alpha/2}\sigma}{\sqrt{n}}\right\}$$

This helps in writing the hypothesis testing as

$$\boxed{\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| \leq z_{\alpha/2}; \quad \text{reject } H_0 \text{ if } \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| > z_{\alpha/2}}$$

The p-value is defined as

$$p = P\left(|Z| \geq \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0|\right) = 2P\left(Z \geq \frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0|\right)$$

Example 1.

You went to a grocery store and weighed 15 bags of potatoes. Your observations in lb. are: 1.51, 1.55, 1.44, 1.43, 1.61, 1.45, 1.65, 1.54, 1.46, 1.50, 1.59, 1.53, 1.57, 1.62, 1.64. Assume you know their standard deviation $\sigma = 0.25$. Use $\alpha = 5\%$ significance level.

Your hypotheses about mean of the potato bags are:
Null hypothesis, $H_0 : \mu = 1.5$ lb.
Alternate hypothesis, $H_1 : \mu \neq 1.5$ lb.
The mean of the observations is $\bar{X} = 1.54$ lb.
Hence, the test statistic is $\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| = \frac{\sqrt{15}}{0.25}|1.54 - 1.5| = 0.6197 < z_{0.025} = 1.96$.
Accept the null hypothesis, H_0.
The p value is $2P\left(Z \geq 0.697\right)$

One-sided Tests

(a) **Null hypothesis,** $H_0 : \mu = \mu_0$ (or $\mu \le \mu_0$)

Alternate hypothesis, $H_1 : \mu > \mu_0$

$$\boxed{\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) \le z_\alpha; \quad \text{reject } H_0 \text{ if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) > z_\alpha}$$

The p-value is defined as

$$p = P\left(Z \ge \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0)\right)$$

(b) **Null hypothesis,** $H_0 : \mu = \mu_0$ (or $\mu \ge \mu_0$)

Alternate hypothesis, $H_1 : \mu < \mu_0$

$$\boxed{\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) \ge -z_\alpha; \quad \text{reject } H_0 \text{ if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) < -z_\alpha}$$

The p-value is defined as

$$p = P\left(Z \le \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0)\right)$$

9.1.2 Unknown Variance

For the case of unknown variance we need to use the t test:

Null hypothesis, $H_0 : \mu = \mu_0$

Alternate hypothesis, $H_1 : \mu \neq \mu_0$

$$\boxed{\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{S}|\bar{X} - \mu_0| \le t_{\alpha/2,n-1}; \quad \text{reject } H_0 \text{ if } \frac{\sqrt{n}}{S}|\bar{X} - \mu_0| > t_{\alpha/2,n-1}}$$

where

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

The p-value is defined as

$$p = P\left(|T_{n-1}| \ge \frac{\sqrt{n}}{S}|\bar{X} - \mu_0|\right) = 2P\left(T_{n-1} \ge \frac{\sqrt{n}}{S}|\bar{X} - \mu_0|\right)$$

One-sided Tests

(a) **Null hypothesis,** $H_0 : \mu = \mu_0$ (or $\mu \le \mu_0$)

Alternate hypothesis, $H_1 : \mu > \mu_0$

$$\boxed{\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{S}(\bar{X} - \mu_0) \le t_{\alpha,n-1}; \quad \text{reject } H_0 \text{ if } \frac{\sqrt{n}}{S}(\bar{X} - \mu_0) > t_{\alpha,n-1}}$$

The p-value is defined as

$$p = P\left(T_{n-1} \ge \frac{\sqrt{n}}{S}(\bar{X} - \mu_0)\right)$$

(b) **Null hypothesis,** $H_0 : \mu = \mu_0$ (or $\mu \geq \mu_0$)

Alternate hypothesis, $H_1 : \mu < \mu_0$

$$\boxed{\text{accept } H_0 \text{ if } \frac{\sqrt{n}}{S}(\bar{X} - \mu_0) \geq -t_{\alpha,n-1}; \quad \text{reject } H_0 \text{ if } \frac{\sqrt{n}}{S}(\bar{X} - \mu_0) < -t_{\alpha,n-1}}$$

The p-value is defined as

$$p = P\left(T_{n-1} \leq \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0)\right)$$

9.2 The Variance of a Normal Population

Null hypothesis, $H_0 : \sigma^2 = \sigma_0^2$
Alternate hypothesis, $H_1 : \sigma^2 \neq \sigma_0^2$

$$\boxed{\text{accept } H_0 \text{ if } \chi^2_{1-\alpha/2,n-1} \leq \frac{(n-1)S^2}{\sigma_0^2} \leq \chi^2_{\alpha/2,n-1}; \quad \text{reject } H_0 \text{ otherwise}}$$

The p-value is defined as

$$p = 2\min\left\{P\left(\chi^2_{n-1} < \frac{(n-1)S^2}{\sigma_0^2}\right), 1 - P\left(\chi^2_{n-1} < \frac{(n-1)S^2}{\sigma_0^2}\right)\right\}$$

Example 2.

Based on $n = 25$ observations, sample average velocity of vehicles on a freeway is $\bar{V} = 110.12$ km/hr. Use $\alpha = 5\%$ significance level. Your hypothesis about the velocity of the vehicles is

Null hypothesis, $H_0 : \mu_V = 110$
Alternate hypothesis, $H_1 : \mu_V \neq 110$
$\sigma = 0.4$
The test statistic

$$\frac{\sqrt{n}}{\sigma}|\bar{X} - \mu_0| = \frac{\sqrt{25}}{0.4}|110.12 - 110| = 1.5 < z_{0.025} = 1.96$$

Accept H_0.
If σ is unknown and you estimate $s = 0.6$.
The test statistic

$$\frac{\sqrt{n}}{S}|\bar{X} - \mu_0| = \frac{\sqrt{25}}{0.6}|110.12 - 110| = 1 < t_{0.025,24} = 2.06$$

Accept H_0.
If your estimate $s = 0.25$.
The test statistic

$$\frac{\sqrt{n}}{S}|\bar{X} - \mu_0| = \frac{\sqrt{25}}{0.25}|110.12 - 110| = 2.4 > t_{0.025,24} = 2.06$$

Reject H_0.

Example 3. The concrete supplier claims his concrete has a mean compressive strength of 38 N/mm^2. On-site you tested randomly selected cubes and got a sample mean 37.5 N/mm^2. Use $\alpha = 5\%$ significance level.

If you want to test the following hypotheses about variance of the concrete cubes with 41 samples giving sample standard deviation $s = 3.75$ N/mm^2:

Null hypothesis, $H_0 : \sigma^2 = 9$
Alternate hypothesis, $H_1 : \sigma^2 \neq 9$
The test statistic

$$\frac{(n-1)S^2}{\sigma_0^2} = \frac{40 \times (3.75)^2}{9} = 62.5 > \chi^2_{0.025,40} = 59.3$$

Reject H_0.

Chapter 10

Regression

10.1 Single variate case

We want to fit a linear regression curve $Y = \alpha + \beta x$ using data $\{x_i, Y_i\}_{i=1}^n$. The coefficients are estimated as A and B using

$$B = \frac{\sum_{i=1}^n x_i Y_i - \bar{x} \sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$
$$A = \bar{Y} - B\bar{x}$$

Estimate of variance of the noise present is

$$s^2 = \frac{SS_R}{n-2}$$

The 95% confidence interval of the **mean response** is

$$A + Bx_0 \pm \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}} \sqrt{\frac{SS_R}{n-2}} t_{a/2,n-2}$$

The 95% confidence interval of **future response** is

$$A + Bx_0 \pm \sqrt{\frac{n+1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}} \sqrt{\frac{SS_R}{n-2}} t_{a/2,n-2}$$

10.2 Multivariate case

In a multivariate case $Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K$, the coefficients can be estimated using

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} \\ 1 & x_{1,2} & x_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

and

$$\mathbf{B} = \begin{bmatrix} A \\ B_1 \\ B_2 \\ \vdots \\ B_K \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y})$$

Example 1.

(a) Fit a linear regression curve $Y = \alpha + \beta x$ to the following data.

No.	x_i	Y_i	x_iY_i	x_i^2	$(Y_i - A - Bx_i)^2$
1	1.11	0.52	0.58	1.23	0.0313
2	1.17	0.40	0.47	1.37	0.0009
3	1.79	0.97	1.74	3.20	0.1110
4	5.62	2.92	16.40	31.60	0.4000
5	1.13	0.17	0.19	1.28	0.0328
6	1.54	0.19	0.29	2.37	0.1158
7	3.19	0.76	2.43	10.15	0.2360
8	1.73	0.66	1.14	2.99	0.0023
9	2.09	0.78	·	·	·
10	2.75	1.24	·	·	·
11	1.20	0.39	·	·	·
12	1.01	0.30			
13	1.64	0.70			
14	1.57	0.77			
15	1.54	0.59			
16	2.09	0.95			
17	3.54	1.02			
18	1.17	0.39			
19	1.15	0.23			
20	2.57	0.45			
21	3.57	1.59			
22	5.11	1.74			
23	1.52	0.56			
24	2.93	1.12			
25	2.93	0.64			
Sum	53.89	20.05	59.24	153.44	1.7350

$$\bar{x} = 53.89/25 = 2.16, \quad \bar{Y} = 20.05/25 = 0.80$$

Hence,

$$B = \frac{\sum_{i=1}^{n} x_iY_i - \bar{x}\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} = \frac{59.24 - 2.16 \times 20.05}{153.44 - 25 \times (2.16)^2} = 0.435$$
$$A = \bar{Y} - B\bar{x} = 0.80 - 0.435 \times 2.16 = -0.14$$

Hence, the linear regression fitted to the data is $Y = -0.14 + 0.435x$.
(b) Estimate of variance of the noise present is

$$s^2 = \frac{SS_R}{n-2} = \frac{1.735}{23} = 0.075$$

(c) Estimate $P(Y > 2|X = 4) =?$ assuming Y given $X = x$ is Gaussian distributed.

$$\mathbb{E}[Y|X = 4] = \mu_{Y|X=4} = -0.14 + 0.435 \times 4 = 1.6$$

Hence,

$$\begin{aligned} P(Y > 2|X = 4) &= 1 - P(Y \leq 2|X = 4) \\ &= 1 - \Phi\left(\frac{2 - \mu_{Y|X=4}}{s}\right) \\ &= 1 - \Phi\left(\frac{2 - 1.6}{\sqrt{0.075}}\right) = 0.072 \end{aligned}$$

(d) Estimate the 95% confidence interval of the **mean response** at $x_0 = 1$.
The 95% confidence interval is

$$\begin{aligned} &A + Bx_0 \pm \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}}\sqrt{\frac{SS_R}{n-2}}t_{a/2,n-2} \\ &= 0.295 \pm \sqrt{\frac{1}{25} + \frac{(1 - 2.16)^2}{153.44 - 25 \times (2.16)^2}} \times \sqrt{\frac{1.7350}{23}} \times 2.069 \\ &= (0.138, 0.452) \end{aligned}$$

(e) Estimate the 95% confidence interval of **future response** at $x_0 = 1$.
The 95% confidence interval is

$$\begin{aligned} &A + Bx_0 \pm \sqrt{\frac{n+1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}}\sqrt{\frac{SS_R}{n-2}}t_{a/2,n-2} \\ &= 0.295 \pm \sqrt{\frac{26}{25} + \frac{(1 - 2.16)^2}{153.44 - 25 \times (2.16)^2}} \times \sqrt{\frac{1.7350}{23}} \times 2.069 \\ &= (-0.295, 0.885) \end{aligned}$$

Example 2.

Fit a regression curve $Y = \alpha + \beta_1 x_1 + \beta_2 x_2$.

No.	x_{1i}	x_{2i}	Y_i
1	2375	39.27	47.5
2	1459	39.00	52.3
3	604	38.35	56.8
4	3242	37.58	48.4
5	550	39.38	54.2
6	675	38.05	55.1
7	635	39.65	54.4
8	2727	38.66	48.8
9	2424	37.97	50.5
10	659	40.10	52.7

Form the matrices

$$\mathbf{X} = \begin{bmatrix} 1 & 2375 & 39.27 \\ 1 & 1459 & 39.00 \\ 1 & 604 & 38.35 \\ 1 & 3242 & 37.58 \\ 1 & 550 & 39.38 \\ 1 & 675 & 38.05 \\ 1 & 635 & 39.65 \\ 1 & 2727 & 38.66 \\ 1 & 2424 & 37.97 \\ 1 & 659 & 40.10 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} 47.5 \\ 52.3 \\ 56.8 \\ 48.4 \\ 54.2 \\ 55.1 \\ 54.4 \\ 48.8 \\ 50.5 \\ 52.7 \end{bmatrix}$$

Hence,

$$\mathbf{B} = \begin{bmatrix} A \\ B_1 \\ B_2 \end{bmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}) = \begin{bmatrix} 121.05 \\ -0.0034 \\ -1.644 \end{bmatrix}$$

The estimated regression curve is $Y = 121.05 - 0.0034x_1 - 1.644x_2$.

Example 3.

Fit a regression curve $Y = \alpha + \beta \log x$.

Assume $z = \log x$. Hence, the regression curve is $Y = \alpha + \beta z$.

$$\bar{z} = 58.3408/10 = 5.8341, \quad \bar{Y} = 6.40/10 = 0.64$$

Hence,

$$B = \frac{\sum_{i=1}^n z_i Y_i - \bar{z}\sum_{i=1}^n Y_i}{\sum_{i=1}^n z_i^2 - n\bar{z}^2} = \frac{37.5453 - 5.8341 \times 6.40}{341.1911 - 10 \times (5.8341)^2} = 0.2514$$
$$A = \bar{Y} - B\bar{z} = -0.8264$$

Hence, the nonlinear regression curve fitted to the data is $Y = -0.8264 + 0.2514 \log x$.

No.	x_i	Y_i	$z_i = \log x_i$	z_iY_i	z_i^2
1	185	0.50	5.2204	2.6102	27.2521
2	310	0.48	5.7366	2.7536	32.9083
3	260	0.51	5.5607	2.8359	30.9212
4	320	0.58	5.7683	3.3456	33.2735
5	480	0.60	6.1738	3.7043	38.1156
6	340	0.67	5.8289	3.9054	33.9766
7	380	0.69	5.9402	4.0987	35.2856
8	540	0.75	6.2916	4.7187	39.5838
9	340	0.82	5.8289	4.7797	33.9766
10	400	0.80	5.9915	4.7932	35.8976
Sum		6.40	58.3408	37.5453	341.1911