5.1. Summary of Probability Distributions

Let $x$ be a r.v. defined on a probability space $(\Omega, B, P)$ associated with a random $\varepsilon$.

$F_x$: distribution function of $x$

$f_x$: p.m.f. / p.d.f. of $x$

(see. p.166/p.156)

The probability distribution of $x$ describes the manner in which the r.v. $x$ takes values on various sets. It may be desirable to have a set of numerical measures that provide a summary of the prominent features of the probability distribution of $x$. We call these measures as descriptive measures. Four prominently used descriptive measures are:

- **Measures of Central Tendency**
  - **Location** (also called average)
    - Give us the idea about central value of the probability distribution around which the value of r.v. $x$ can be centered. Commonly used measures of central tendency are:

  - **Mean**:
    \[ \mu = \mu(x) = \frac{\sum x_i f(x)}{\sum f(x)} \text{ or } \frac{\sum x_i f(x)}{\sum f(x)} \]
    - May or may not exist
    - Whenever it exists it gives us the idea about average observed value of $x$ when $x$ is repeated a large number of times.
Note that if the distribution of $x$ is symmetric about $\mu$ (i.e., $x - \mu \sim N(0, \sigma^2)$) then

$$E(x) = \mu,$$

provided it exists.

Mean, $\mu$, means the best statistical measure of central tendency for symmetric distribution. Because of its tendency for asymmetric distribution, mean is the least commonly used average. Instead, median is the most commonly used average. However, mean may be affected by a few extreme values and also it may not always be defined.

Before defining the median, we briefly introduce the concept of quantile function or quantile: $Q_x$.

The quantile function of a rv $x$ is a function $Q_x: (0, 1) \rightarrow \mathbb{R}$ defined by

$$Q_x(p) = \inf \{ x \in \mathbb{R} : F(x) \geq p \} \quad p \in (0, 1).$$

For a fixed $p \in (0, 1)$, the quantity $Q_x(p)$ is called the quantile of order $p$. Note that

$$F(x)^{(p)} \leq P \leq F(x)(p),$$

and $F(x)(p) = 1$, provided $F(x)$ is continuous at $x^n$.

Also note that:

- $Q_x(F(x)) \leq \mu$, provided $0 < F(x) < 1$
\[ F_X(\theta(x)) \geq P, \quad 0 < P < 1 \]

- If \( X \) is continuous, then \( F_X(\theta(x)) = P \).
- \( \theta(x) \leq x \Rightarrow F_X(\theta(x)) \geq P \).
- \( \theta(x) = F_X^{-1}(P) \) provided \( F_X(x) \) exists.
- \( \theta(x) \leq \theta(x) \), \( 0 < P < 1 \)

The quantile of order 0.5 is called the median of \( X \). If \( \theta(x) \) is the median of \( X \), then:
\[ F_X(\theta(x)) = \frac{1}{2} \leq F_X(\theta(x)) \]

If the random experiment \( E \) is repeated a large number of times, half of the times observed value \( X \) is expected to be less than \( \theta(x) \) and half of the times it is expected to be greater than \( \theta(x) \).

Suppose that the distribution of \( X \) is symmetric about \( \mu \). Then:
\[ X - \mu \leq \mu - x \]

\[ \Rightarrow P(X - \mu \leq 0) \geq P(\mu - x \leq 0) \]

\[ \Rightarrow F_X(\mu) = 1 - F_X(x) \]

\[ \Rightarrow F_X(\mu) = \frac{1}{2} \leq F_X(x) \]

\[ \Rightarrow \mu = E(X) = \text{Med} \quad \text{provided } F_X \text{ is continuous at } \mu. \]

**Median as a Measure of Central Tendency:**

- Unlike mean, it is always defined.
- Median is not affected by a few extreme values of \( X \) as it takes into account only the probabilities with which different values occur and not their numerical values.

As a measure of central tendency, the median is preferred over the mean if the distribution is asymmetric and a few extreme observations occur with positive probability.
Diversity of Median as a measure of central tendency

- Does not at all take into account the numerical values assumed by \( x \).
- For many probability distributions, it is not easy to evaluate.

**Mode**

Roughly speaking, mode \( \mu_o \) of a probability distribution is the value that occurs with highest probability and is defined by

\[
\delta_{\mu_o}(x) = \max \{ \delta_{\mu}(x) : x \in \mathbb{R} \}
\]

If the random experiment \( E \) is repeated a large number of times, then either mode \( \mu_o \) or a value in the neighborhood of \( \mu_o \) is observed with maximum frequency.

Note that mode of a distribution may not be unique. A distribution having single/double/triple/multiple modes/\( \mu_o \) is called a unimodal/bimodal/trimodal/multimodal distribution.
Feature of a mode as a measure of central tendency

It is easy to understand and easy to calculate. Normally, it can be found by just inspection.

Demerits of mode as a measure of central tendency

A probability distribution may have more than one mode which may be far apart.

As a measure of central tendency mode is less preferred than mean and median. Clearly for symmetric unimodal distributions, mean = median = mode.

(iii) Measures of Dispersion

Apart from measures of central tendency other measures are often required to describe a probability distribution. Measures of dispersion give the idea about the scatter (cluster dispersion) of probability mass of the distribution about a measure of central tendency. Some of the measures of dispersion are listed below.

(a) Range

Let $S_X = [a, b]$. The range of distribution of $X$ is defined by

$$R = b - a$$

It does not take into account how the probability mass is distributed over $[a,b]$. For this reason, it is not a preferred measure of dispersion.

(b) Mean Deviation

Let

$$A:$$ Suitable measure of central tendency.
Define

\[ MD(A) = E(|x-A|) \] 
\[ \text{mean deviation about A} \]

\[ MD(M) = E(|x-M|) \] 
\[ \text{mean deviation about mean M} \]

\[ MD(me) = E(|x-me|) \] 
\[ \text{mean deviation about median} \]

It can be shown that

\[ MD(me) \leq MD(A) \leq AEIL \]

For this reason, MD(me) seems to be more affected than MD(A) for any AEIL.

- MD(A) is generally difficult to compute for many distributions.
- MD(A) is sensitive to extreme observations.
- MD(A) may not exist for many distributions.

### Standard Deviation

The standard deviation of distribution of \( x \) is defined by

\[ \sigma = \sqrt{\text{Var}(x)} = \sqrt{E((x-M)^2)} \]

where \( M \) is

\[ \text{mean} \]

Clearly

\[ \sigma \leq \sqrt{E((x-A)^2)} \]

\[ \sigma \leq AEIM \]

Standard deviation gives us the idea of average spread of values of \( x \) around mean \( M \).

- It is simple to compute for many distributions (unlike MD(A), AEIL).

\[ 6/5 \]
SD is
- Most widely measured of dispersion (especially for nearly symmetric distributions)
- For some distributions SD does not exist
- SD is sensitive to extreme observations

(d) Quantile Deviation

Let

\[ q_1 = 50.25 = \text{quantile of order 0.25 (lower quartile of } x) \]
\[ q_2 = 50.5 = \text{quantile of order 0.5 = median} \]
\[ q_3 = 50.75 = \text{quantile of order 0.75 (upper quartile of } x) \]

\[ q_1, q_2, q_3 \text{ divide the probability distribution of } x \]
into 4 parts so that

\[ F_X(q_2) \leq \frac{1}{4} \leq F_X(q_1); \quad F_X(q_2) \leq \frac{3}{4} \leq F_X(q_3) \]

and \[ F_X(q_1) \leq \frac{1}{4} \leq F_X(q_3) \]

Note that \[ q_1, q_2, q_3 \text{ divide the j.d.b. / p.m.f. of } x \text{ into 4 parts so that each of them has 25% probability mass.} \]

Define

\[ IQR = q_3 - q_1 \rightarrow \text{inter-quartile range} \]
\[ QD = \frac{q_3 - q_1}{2} \rightarrow \text{quantile deviation or the semi-interquartile range} \]

\[ 7/5 \]
Unlike SD, QD is not sensitive to extreme values assumed by $x$.

- Does not at all take into account numerical values of $x$.
- Ignores the tail of the probability distribution (contributing 50% of probability distribution on left side of $v_1$ and right side of $v_3$).

QD depends on the unit of measurements of $x$ and thus it may not be appropriate for comparing distributions of two probability distributions having different units of measurements. For this purpose, one may use

$$\text{CQD} = \frac{v_3 - v_1}{v_3 + v_1} \rightarrow \text{Coefficient of quartile deviation}$$

- does not depend on units of measurements.

(d) Coefficient of Variation

Like QD, the SD also depends on units of measurements of $x$ and thus it is not an appropriate measure of dispersion for comparing distributions having different units of measurements. For this purpose, we consider

$$\text{CV} = \frac{\sigma}{\mu} \rightarrow \text{Coefficient of Variation}$$

- does not depend on units of measurement.

Where $\mu = \text{E}(x)$ and $\sigma = \sqrt{\text{Var}(x)}$. Here we assume that $\mu \neq 0$.

- CV measures variation for units of measurement.
- CV is very sensitive to small changes in $\mu$ when $\mu$ is near 0.
Measures of Skewness

Skewness of a probability distribution is a measure of asymmetry (lack of symmetry).

Recall that:

Distribution of $X$ is symmetric about $\mu$

$\iff \quad x - \mu \equiv \mu - x$

$\iff \quad f_X(\mu + x) = f_X(\mu - x) \quad \forall x \in \mathbb{R}$

And in that case:

- $\mu = E(X) = \text{med} \text{ian}$;
- the shape of the p.d.f./p.m.f. on the left of $\mu$ is the mirror image of that on the right side of $\mu$.

Positively skewed distributions:

- Have more probability mass to the right side of p.d.f./p.m.f.
- Have longer tails on the right side of p.d.f.

Unimodal

For positively skewed distribution, normally

Mode < Median < Mean

Since the positive mass to large values of $x$ pulls up the value of mean $\mu$. 
Negatively skewed distributions

- Have more probability mass to the left side of the f.d.f. / p.m.f.
- Have longer tail on the left side of f.d.f.

For negatively skewed distributions, normally

Min < Median < Mode.

Let $\mu = E(x)$, $\sigma = \sqrt{\text{var}(x)}$ and

$$E = \frac{X - \mu}{\sigma} : \text{Standardized variable}$$

(Independent of units)

Define

$$p_1 = E(2^3) = E((X-\mu)^3) = \frac{\mu_3}{\mu_2^{3/2}}$$

where $\mu_3 = E((X-\mu)^3)$

$\to$ Coefficient of Kurtosis

- For asymmetric distributions $p_1 > 0$. Converse may not be true.
- For positively skewed distributions, normally $p_1$ is a large positive quantity.
- For negatively skewed distributions, normally $p_1$ is a small negative quantity.
A measure of skewness can also be based on quantiles. Let

$q_1$: first quartile
$q_2$: median (or second quartile $u_2$)
$q_3$: third quartile

$\mu$: mean

- For symmetric distributions: $q_3 - u = u - q_1$ ($u = \frac{q_1 + q_3}{2}$)
- For positively skewed distributions: $q_3 - u > u - q_1$
- For negatively skewed distributions: $q_3 - u < u - q_1$

Thus, a measure of skewness can be based on $(q_3 - u) - (u - q_1) = q_3 - 2u + q_1$.

Define

$$\beta_2 = \frac{(q_3 - u) - (u - q_1)}{q_3 - q_1} = \frac{q_3 - 2u + q_1}{q_3 - q_1}$$

($\mu$ is the mean)

(L) Yule's coefficient of skewness.

Clearly for positively/negatively skewed distributions $\beta_2 > 0/\beta_2 < 0$ and for symmetric distributions $\beta_2 = 0$.

**Measures of Kurtosis**

For $\mu = \mu_0$ and $\sigma > 0$, let $\mu_0$ be a r.v. having f.d.f.

$$f_{\mu_0}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_0)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

It can be shown that $\mu_0$ has a normal distribution ($\mu_0 \sim N(\mu_0, \sigma^2)$).

- $E(\mu_0) = \mu_0$;
- $\text{Var}(\mu_0) = \sigma^2$;
- $\mu_0 - \mu \sim N(\mu - \mu_0, \sigma^2)$ and hence $\beta_1 = 0$
- $E((\mu_0 - \mu)^4) = 3\sigma^4$. 

11/5
$f(x)$ is unimodal and symmetric.

Kurtosis of the probability distribution of $x$ is a measure of peakedness and thickness of tails of f.d.f. of $x$ relative to that of normal distribution.

A distribution is said to have higher (lower) kurtosis than the normal distribution if its f.d.f. in comparison with f.d.f. of a normal distribution, has a sharper (rounder) peak and longer, fatter (thinner, thinner) tails.

Define $Z = \frac{(X-\mu)}{\sigma}$ (independent of $\mu$ and $\sigma$)

$$D_1 = \text{E}(Z^4) = \frac{\text{E}((X-\mu)^4)}{\sigma^4} = \frac{\mu_4}{\sigma^4}$$

Kurtosis of the probability distribution of $x$.

$D_1$ is used as a measure of kurtosis for unimodal distributions.

For $N(\mu, \sigma^2)$ distribution, $D_1 = 3$.

The quantity

$$D_2 = D_1 - 3$$

is called the excess kurtosis of the distribution of $x$.

$12/5$
Obviously for normal distributions $\nu_2 = 0$.

Leptokurtic distributions: Distributions with $\nu_2 > 0$ (has sharper peak and longer, thinner tails)

Platykurtic distributions: Distributions with $\nu_2 < 0$ (has rounded peak and thicker, fatter tails)

**Example 5.1** For $\alpha \in \mathbb{R}$, let $X_\alpha$ have the p.d.f.

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

Recall that for $\nu \in \{1, 2, \ldots\}$

$$I_{\nu} = \int_0^\infty e^{-x} x^{\nu} \, dx = \frac{\Gamma(\nu+1)}{\nu!} \quad \text{(using integration by parts)}$$

Thus, for $\nu \in \{1, 2, \ldots\}$

$$M_{\nu}(x) = E(X^\nu) = \int_0^\infty x^\nu e^{-x} \, dx + \int_0^\infty (1-x)^\nu x e^{-x} \, dx$$

$$= \left[ (\nu+1)x + 1-x \right] \int_0^\infty x^\nu e^{-x} \, dx$$

$$= \begin{cases} (1-2\alpha) I_{\nu}, & \nu \in \{1, 3, 5, \ldots\} \\ I_{\nu}, & \nu \in \{2, 4, 6, \ldots\} \end{cases}$$

Let $S_p$ be the quantile of order $p \in (0, 1)$. Then

$F_{X_\alpha}(S_p) = p$ when $F_{X_\alpha}$ is the d.f. of $X_\alpha$.  

13/5
Clearly,
\[ F_{\alpha}(x) = \alpha \int_{-\infty}^{x} e^t \, dt = \alpha \]

For \( 0 \leq x < 1 \)
\[ p = F_{\alpha}(s) = \int_{-\infty}^{s} e^t \, dt = \int_{-\infty}^{s} (1-t) e^{-t} \, dt = 1 - (1-x) e^{-s} \]

and for \( x \geq 1 \)
\[ p = \int_{-\infty}^{\infty} e^t \, dt = \alpha e^{s} \]

Thus
\[ s_p = \begin{cases} \frac{\ln(\frac{1-x}{1-p})}{p}, & 0 \leq x < 1 \\ -\ln\left(\frac{\alpha}{p}\right), & x \geq 1 \end{cases} \]

For \( x_1 = 5 \frac{1}{4} \)
\[ q_{11} = \frac{\ln\left(\frac{4(1-x)}{3}\right)}{3}, \quad 0 \leq x < \frac{1}{2} \\ -\ln\left(4x\right), \quad \frac{1}{4} \leq x \leq 1 \]

For \( x_2 = 5 \frac{1}{2} \)
\[ q_{22} = \frac{\ln\left(2(1-x)\right)}{2}, \quad 0 \leq x < \frac{1}{2} \\ -\ln\left(2x\right), \quad \frac{1}{4} \leq x < 1 \]

For \( x_3 = 5 \frac{3}{4} \)
\[ q_{33} = \frac{\ln\left(4(1-x)\right)}{3}, \quad 0 \leq x < \frac{1}{4} \\ -\ln\left(\frac{4x}{3}\right), \quad \frac{3}{4} \leq x < 1 \]

\[ h_1(x, 1 = 2, \lambda = 1 - 2x \]

Mode = \( \ln(\alpha) = \max \{ \delta(x) ; -0 < \lambda < 0 \} = \ln\alpha \{ \delta, 1 - 2x \}

\[ h_2(x) = E(x^2) = 2 \]

\( \sigma(x) = \sqrt{\text{Var}(x)} = \sqrt{1 + 4x - x^2} \)

\[ 14/5 \]
Note that for $0 \leq x < \frac{1}{2}$, $\ln(x) = \ln(2(1-x)) \geq 0$ and for $x > \frac{1}{2}$, $\ln(x) = -\ln(2x) < 0$. Thus for $0 \leq x < 1$ (no that $\ln(x) > 0$)

$$P_D(\ln(x)) = E \left( |x - \ln(x)| \right)$$

$$= \int_0^\infty \left( \ln(x) - x \right) e^{-x} \, dx + \int_0^{\ln(2)} \frac{e^{x-\ln(x)}}{e^{x}} \, dx + \int_{\ln(2)}^\infty \left( x - \ln(x) \right) e^{-x} \, dx$$

$$= \ln(x) + 2x$$

$$= \ln(2(1-x)) + 2x$$

Similarly, for $\frac{1}{2} \leq x \leq 1$ (no that $\ln(x) \leq 0$)

$$P_D(\ln(x)) = E \left( |x - \ln(x)| \right)$$

$$= \int_0^\infty \left( \ln(x) - x \right) e^{-x} \, dx + \int_0^{\ln(2)} \frac{e^{x-\ln(x)}}{e^{x}} \, dx + \int_{\ln(2)}^\infty \left( x - \ln(x) \right) e^{-x} \, dx$$

$$= 2(1-x) - \ln(x)$$

$$= \ln(2(1-x) + 2(1-x)$$

Thus

$$P_D(\ln(x)) = \begin{cases} \ln(2(1-x)) + 2x, & 0 \leq x \leq \frac{1}{2} \\ \ln(2x) + 2x + 1, & \frac{1}{2} \leq x \leq 1 \end{cases}$$

$$\text{IQD} = \text{IQD}(x) = Q_3(x) - Q_1(x)$$

$$= \begin{cases} \ln(3^2), & 0 \leq x \leq \frac{1}{9} \\ \ln(3^2(1-x)), & \frac{1}{9} \leq x \leq \frac{1}{3} \end{cases}$$
$$Q_3(\alpha_1) = \frac{\nu_3(\alpha_1) - \nu_1(\alpha_1)}{2}$$

$$= \begin{cases} 
\ln \sqrt{3}, & 0 \leq \alpha < \frac{1}{2} \\
\ln \left( \frac{4\sqrt{\alpha(1-\alpha)}}{3} \right), & \frac{1}{4} \leq \alpha < \frac{3}{4} \\
\ln \sqrt{3}, & \frac{3}{4} \leq \alpha \leq 1
\end{cases}$$

$$C_3(\alpha) = \frac{\nu_3(\alpha_1) - \nu(\alpha)}{\nu_3(\alpha_1) - \nu_1(\alpha)}$$

$$= \begin{cases} 
\frac{\ln 3}{\ln \left( \frac{16(1-\alpha)^2}{3} \right)}, & 0 \leq \alpha < \frac{1}{4} \\
\frac{\ln \left( \frac{16\alpha(1-\alpha)}{\ln \left( \frac{1-\alpha}{2} \right)} \right)}{\ln \left( \frac{2-\alpha}{3} \right)}, & \frac{1}{4} \leq \alpha \leq \frac{3}{4} \\
\frac{\ln 3}{\ln \left( \frac{16\alpha^2}{3} \right)}, & \frac{3}{4} \leq \alpha \leq 1
\end{cases}$$

For $\alpha > \frac{1}{2}$

$$C_1 = C_1(\alpha_1) = \frac{\sigma_-(\alpha_1)}{\sigma_+(\alpha_1)} = \frac{\sqrt{1+4(1-\alpha)^2}}{1-2\alpha}$$

$$M_3(\alpha_1) = E(\left( X - \mu_1(\alpha_1) \right)^3)$$

$$= M_3(\alpha_1) - 3M_2(\alpha_1)\mu_2(\alpha_1) + 2(M_2(\alpha_1))^3$$

$$= 2(1-2\alpha)^3$$

$$B_1 = B_1(\alpha) = \frac{M_3(\alpha)}{\sigma_+(\alpha)} = \frac{2(1-2\alpha)^3}{\sqrt{1+4(1-\alpha)^2}}$$

$$B_2 = B_2(\alpha) = \frac{\nu_3(\alpha_1) - 2\nu(\alpha_1) + \nu_1(\alpha)}{\nu_3(\alpha_1) - \nu_1(\alpha)}$$

[16/5]
\[
\begin{cases}
\frac{\ln(\frac{3}{5})}{\ln(\frac{2}{5})}, & \text{if } 0 \leq x < \frac{1}{2} \\
\frac{\ln((4x-1)x)}{\ln((16x-1)x)}, & \text{if } \frac{1}{2} \leq x < \frac{3}{4} \\
\frac{\ln((4x-1)x)}{\ln((16x-1)x)}, & \text{if } \frac{3}{4} \leq x \leq 1
\end{cases}
\]

Clearly, for \(0 \leq x < \frac{1}{2}\), \(p(x) > 0\), and, for \(\frac{1}{2} \leq x \leq 1\), \(p(x) < 0\). For \(\alpha = 2\), \(p(x) \geq 0\) for all \(x\).

Thus:
- For \(0 \leq x < \frac{1}{2}\), distribution of \(X_2\) is positively skewed.
- For \(\frac{1}{2} \leq x \leq 1\), distribution of \(X_2\) is negatively skewed.
- For \(x = \frac{1}{2}\), distribution of \(X_2\) is symmetric. (Since in this case \(b(x) = b(-x)\) for all \(x\).)

\[
\mu_2 = \mu_2(x) = \mathbb{E}[(x - \mu_1(x))^2]
\]

\[
= \mu_2(x) - 2 \mu_1(x) \mu_3(x) + 6 \left(\frac{\mu_3(x)}{\mu_1(x)}\right)^2 - 3 \left(\frac{\mu_3(x)}{\mu_1(x)}\right)^3
\]

\[
D_1 = D_1(x) = \frac{\mu_4(x)}{(\mu_2(x))^2} = \frac{24 - 12(1-2x)^2 - 3(1-2x)^2}{2 - (1-2x)^2}^2
\]

and \(D_2 = D_2(x)-3 = \frac{12 - 6(1-2x)^2}{(2 - (1-2x)^2)^2}\)

Clearly, for any \(x \in [0,1]\), \(D_2(x) > 0\). It follows that for any value of \(x \in [0,1]\) the distribution of \(X_2\) is leptokurtic.
5.2. Some Special Discrete Distributions

5.2.1. Bernoulli and Binomial Distribution

**Bernoulli Experiment.** A random experiment with just two possible outcomes (say) Success (S) and Failure (F). Each replication of a Bernoulli experiment is called a Bernoulli trial.

Consider a sequence of \( n \) independent Bernoulli trials with probability of success \( p \) in each trial as \( p \in (0,1) \) (the same for each trial); i.e., \( n \) in some fixed natural number.

Define

\[ X = \text{# of successes in } n \text{ trials}. \]

Then \( X \in \{0, 1, 2, \ldots, n\} \) and for \( k \leq n \),

\[
P(X = k) = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\text{Total of } \binom{n}{k} \text{ terms}}
\]

\[
= \binom{n}{k} p^k (1-p)^{n-k} \quad \text{(independence of trials)}
\]

Thus

\[
x(k) = \begin{cases} 
\binom{n}{k} p^k (1-p)^{n-k}, & k = 0, 1, 2, \ldots, n \\
0, & \text{otherwise}
\end{cases}
\]

\[ \rightarrow \text{Binomial distribution with } n \text{ trials and success probability } p \text{ (denoted by } \text{Bin}(n, p) \text{) and written as } X \sim \text{Bin}(n, p) \]

\[ 18/5 \]
\( \text{Bin}(n, p) : \text{new } p \in (0, 1) \)

- family of probability distributions.
  - has two parameters, \( n \) and \( p \) \( \in (0, 1) \)

\( \text{Bin}(1, p) : \text{Bernoulli distribution} \)

\( \text{Bin}(1, p) : \text{Bernoulli distribution with success probability } p \in (0, 1) \).

**Proof:** Suppose that \( X \sim \text{Bin}(n, p) \), \( n \in \mathbb{N}, p \in (0, 1) \).

\[
\begin{align*}
\mathbb{E}(e^x) &= \sum_{k=0}^{n} e^{tx} \binom{n}{k} p^{k} (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^{k} (1-p)^{n-k} \\
\Pi_{x \rightarrow 1} &= (1-p + pe^t)^n + e \cdot ne^t
\end{align*}
\]

Let \( v = 1 \), and let \( \Pi_{x \rightarrow v} = (v + pe^t)^n, \ v \in \mathbb{R} \).

\[
\begin{align*}
\Pi_{x \rightarrow 1}^{(1)} &= n (v + pe^t)^{n-1} pe^t \\
\Pi_{x \rightarrow 1}^{(2)} &= np (v + pe^t)^{n-1} e^t + n(n-1) (v + pe^t) (pe^t)^t \\
E(x) &= \Pi_{x \rightarrow 0}^{(1)} = np \\
E(x^2) &= \Pi_{x \rightarrow 0}^{(2)} = np + n(n-1) p^2 \\
\text{Var}(x) &= E(x^2) - (E(x))^2 = np(1-p) = npnp \\
\text{Note that if } X \sim \text{Bin}(n, p) \text{ then } \\
\text{Variance} &< \text{Mean}
\end{align*}
\]

It can be seen that

\[
\begin{align*}
\mu_1 &= E(x^1) = np \{ 1+3p + 3np + 3n^2 p + 3n^2 p^2 + n^2 p^3 \} \\
\mu_1^4 &= E(x^4) = np \{ 1+7p + 12np + 18n^2 p^2 + 6n^2 p^3 - 6n^2 p^4 + 4n^3 p^5 - 6n^2 p^6 + n^3 p^7 \}
\end{align*}
\]
\[ \mu_2 = E((x-\mu_1)^2) = \eta \gamma (1-p) (1-p) \]
\[ \mu_4 = E((x-\mu_1)^4) = \eta \gamma (1-p) \left[ 3 \gamma (2-p) + 3 \gamma (3-p) + 1 \right] \]

\[ \beta_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{1 - 2p}{\sqrt{\eta \gamma (1-p)}} \quad \text{(Symmetric for } p = \frac{1}{2}) \]

\[ \alpha_2 = \frac{\alpha_1 - 3}{\eta \gamma } = \frac{1 - 6p \gamma}{\eta \gamma } \quad \text{where } \alpha_1 = \frac{\mu_4}{\mu_2^2}. \]

**Alt.** For \( v \in \mathbb{Z} \), let \( x(v) = x(x+1) \ldots (x-v+1) \). Then

\[ E(x(v)) = \sum_{k=0}^{n} \binom{n}{k} (1-p)^{k} (1-p)^{n-k} \]

\[ = n \binom{n}{n-v} \sum_{k=n-v}^{n} \binom{n-v}{k} (1-p)^{n-k} \]

\[ = n \binom{n}{n-1} (1-p)^{n-1} \sum_{k=n}^{n-v} \binom{n-v}{k} (1-p)^{n-k} \]

\[ = n \binom{n}{n-1} (1-p)^{n-1} \left( (1-p)^{n-1} \right) \]

\[ = n \binom{n}{n-1} \]

**Theorem 5.2.1.** Let \( x_1 \ldots x_k \) be independent \( \text{iid} \) with \( X_i \sim \text{Bin}(n_i, p) \), and \( y_i \sim \text{Bin}(n_i, \gamma) \). Then

\[ Y = \sum_{i=1}^{k} X_i \sim \text{Bin}(n, p) \]

**Proof.** For \( t \in \mathbb{R} \)

\[ M_Y(t) = E(e^{tY}) = E\left( e^{\frac{t}{n} \sum_{i=1}^{k} X_i} \right) \]

\[ = E\left( \prod_{i=1}^{k} e^{tX_i} \right) = \prod_{i=1}^{k} E\left( e^{tX_i} \right) \quad \text{(Independence)} \]

\[ = \prod_{i=1}^{k} \prod_{i=1}^{n_i} (1-p + pe^t)^{n_i} \]

\[ = \left( 1 - p + pe^t \right)^{\sum_{i=1}^{k} n_i} \quad \text{m.o.c. of Bin}(\frac{\sum_{i=1}^{k} n_i}{n}, p) \]

\[ \frac{20}{57} \]
By uniqueness of m.g.f. \( T \sim \text{Bin}(n, \frac{1}{2}) \) when \( n \to \infty \).

\[ \text{Example 5.2.1} \]

Let \( X \sim \text{Bin}(n, \frac{1}{2}) \). Then \( X - \frac{n}{2} \overset{d}{=} \frac{n}{2} - X \).

Since \( n - x \overset{d}{=} X \). (Exercise)

\[ \text{Example 5.2.2} \]

A fair die is rolled 5 times (independently).
Find the probability that on 3 occasions we get a 6.

\[ \text{Solution} \]

Consider getting a 6 as success. Then

\[ X = \# \text{ of successes in 5 trials} \]

\[ \sim \text{Bin}(5, \frac{1}{6}) \]

Required probability \( = P(X = 3) \)

\[ = \binom{5}{3} \left( \frac{1}{6} \right)^3 \left( \frac{5}{6} \right)^2 . \]

5.2.2. Negative Binomial Distribution

Consider a sequence of independent Bernoulli trials with probability of success in each trial as \( p \in (0, 1) \). Let \( r \in \{1, 2, \ldots \} \) be a fixed positive integer.

\[ X = \# \text{ of failures preceding the } r \text{-th success} \]

Then \( S_x = \{0, 1, 2, \ldots \} \) and for \( k \in S_x \)

\[ f_x(k) = P(X = k) \]

\[ = P( k \text{ failures precede } r \text{-th success}) \]

\[ = P( \text{r-1 successes in first } k+r-1 \text{ trials and success in } (k+r) \text{-th trial}) \]

\[ = P( \text{r-1 successes in first } k+r-1 \text{ trials} | X \]

\[ P( \text{success in } (k+r) \text{-th trial}) \]

(Independence of trials)

\[ \]
\[\begin{align*}
&\sum_{k=0}^{\infty} \binom{k+r-1}{r-1} t^k = 1 + r + \frac{(rt)^r}{1} + \frac{(rt)^{r+1}}{1} + \frac{(rt)^{r+2}}{1} + \cdots \\
&= (1-t)^{-r}.
\end{align*}\]

The m.s.f. of \( X \sim NB(r, p) \) is

\[\begin{align*}
\mathbb{E}(e^{tx}) &= \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} p^r (1-p)^k \\
&= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} ((1-p)e^t)^k \\
&= p^r (1 - (1-p)e^t)^r, \quad t < -\ln(1-p).
\end{align*}\]
\[ \Pi_{x|H} = \left( \frac{p}{1 - (1-p)e^t} \right)^r \text{ for } t < \ln(p) \]

\[ \psi_{x|H} = \ln \Pi_{x|H} = r \ln p - r \ln (1 - qe^t), \quad t < -\ln(q) \]

\[ \psi_{x|H}^{(1)} = \frac{r q e^t}{1 - qe^t} = r \left[ \frac{1}{1 - qe^t} \right]^t, \quad t < -\ln(q) \]

\[ \psi_{x|H}^{(2)} = \frac{r q e^t}{(1 - qe^t)^2}, \quad t \in H \]

\[ E(x) = \psi_{x|H}^{(0)} = \frac{r q}{p} \]

\[ \text{Variance} > \text{Mean} \]

**All**

For \( m \in \{1, 2, \ldots, I \} \) let \( X(m) = X(x_1, \ldots, x_m) \). Then

\[ E(X(m)) = \sum_{k=0}^{\infty} k(a_{k-1} \cdots (m-1)(r_{m-1}) \left( \frac{a_{m-1}}{q_{m-1}} \right) \text{ for } \frac{1}{p} \]

\[ \text{Var}(X) = \psi_{x|H}^{(2)} = \frac{r q}{p^2} \]

\[ \text{Variance} > \text{Mean} \]

\[ E(X(m)) = \frac{r q}{p} \sum_{k=0}^{\infty} \frac{1}{k!(m-k)!} \left( \frac{p}{1-p} \right)^k \]

\[ = \frac{r q}{p} \left( \frac{1}{m^{\frac{1}{2}}} \right) \sum_{k=0}^{\infty} \frac{1}{k!(m-k)!} \left( \frac{p}{1-p} \right)^k \]

\[ = \frac{r q}{p} \left( \frac{1}{m^{\frac{1}{2}}} \right) \sum_{k=0}^{\infty} \frac{1}{k!(m-k)!} \left( \frac{p}{1-p} \right)^k \]

\[ = \frac{r q}{p} \left( \frac{1}{m^{\frac{1}{2}}} \right) \sum_{k=0}^{\infty} \frac{1}{k!(m-k)!} \left( \frac{p}{1-p} \right)^k \]

\[ = \frac{r q}{p} \left( \frac{1}{m^{\frac{1}{2}}} \right) \sum_{k=0}^{\infty} \frac{1}{k!(m-k)!} \left( \frac{p}{1-p} \right)^k \]

\[ = \frac{r q}{p} \left( \frac{1}{m^{\frac{1}{2}}} \right) \sum_{k=0}^{\infty} \frac{1}{k!(m-k)!} \left( \frac{p}{1-p} \right)^k \]

\[ = \frac{r q}{p} \left( \frac{1}{m^{\frac{1}{2}}} \right) \sum_{k=0}^{\infty} \frac{1}{k!(m-k)!} \left( \frac{p}{1-p} \right)^k \]
\[ M_I = E(x) = \frac{\gamma}{\beta} \]

\[ M_2 = E(x^2) = \frac{\gamma (\gamma + \beta)}{\beta^2} \]

It can be seen that

\[ M_3 = E(x^3) = \frac{\beta^3}{\gamma} \left[ \frac{3\gamma^2 + 3\gamma \beta - \beta^2}{\beta^3} \right] \]

\[ M_4 = E(x^4) = \frac{\beta^4}{\gamma} \left[ \frac{6\gamma^3 + 6\gamma^2 \beta - 3\gamma \beta^2}{\beta^4} \right] \]

\[ M_5 = E(x^5) = \frac{\beta^5}{\gamma} \left[ \frac{15\gamma^4 + 15\gamma^3 \beta - 6\gamma^2 \beta^2 + \gamma \beta^3}{\beta^5} \right] \]

\[ M_6 = E(x^6) = \frac{\beta^6}{\gamma} \left[ \frac{45\gamma^5 + 45\gamma^4 \beta - 15\gamma^3 \beta^2 - 6\gamma^2 \beta^3 + \gamma \beta^4}{\beta^6} \right] \]

\[ p_1 = \frac{M_2}{M_4} = \frac{2 - \frac{\beta}{\gamma}}{\sqrt{\gamma}} \quad \text{(Parameter Known)} \]

\[ d_2 = 0_1 - 3 = \frac{\beta - 2\gamma + 3}{\gamma}, \quad \text{where} \quad 0_1 = \frac{M_4}{M_2^2}. \]

The negative binomial distribution is called a geometric distribution (denoted by \( ge(p) \)), \( 0 < p < 1 \). The pmf of \( \text{ge}(p) \) is given by

\[ p(1) = p(2) = \frac{p^2}{1 - p} \quad \text{for} \quad t = 2, 3, \ldots \]

\[ p(y > m) = \frac{1}{y} \sum_{y=m}^{\infty} = q^m \]

\[ 24/5 \]
**Theorem 5.2.1**

Let $T$ be a discrete type rv with range $S_T = \{0, 1, 2, \ldots\}$. Then $T$ has the 

**Proof**

Obviously,

$$T \sim \text{Ge}(p) \iff \text{Thy is LN property.}$$

Conversely, suppose that $T$ has LN property. Then

$$P(T \geq h+1) = P(T \geq h) P(T \geq k) \quad \forall k \in \{0, 1, 2, \ldots\}$$

Let $P(T=0) = P(T \geq 1)$.

$$P(T \geq h+1) = P(T \geq h) P(T \geq 1)$$

$$= P(T \geq h) (1-p)$$

$$= P(T \geq h+1) (1-p) P(T \geq 1)$$

$$= P(T \geq h+1) (1-p) P(T \geq 1)$$

$$= \cdots$$

$$= P(T \geq 0) (1-p)^{h+1}$$

$$= (1-p)^{h+1}$$

$$= P(T \geq h)$$

$$= P(T \geq h) = P(T \geq k) - P(T \geq k+1)$$

$$= P(1-p)^k, \quad k \in \{0, 1, 2, \ldots\}.$$

$$\Rightarrow \quad T \sim \text{Ge}(p).$$

**Example 5.2.1**

A person repeatedly rolls a fair die independently until an upper face with two or three dots is observed at least twice. Find the probability that the person would require eight rolls to achieve this.

**Solution**

Consider getting 2 or 3 dots as success. Let $Z$ be the number of trials required to get 2 successes.

The probability of success on each trial is $\frac{2}{3}$ and

**Required probability**

$$P(Z=8) = \sum_{k=0}^{7} \binom{k}{2} \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^{k-2} \times \frac{1}{3}$$

$$= \frac{448}{6561}.$$
Consider a population consisting of $N$ units of which $a \in \{1, 2, \ldots, N-1\}$ are labeled as S (success) and $N-a$ are labeled as F (failure). A sample of size $n$ is drawn from this population by drawing one unit at a time, but

$$X = \# \text{ of successes in drawn sample}$$

**Case I.** Draws are independent and sampling is with replacement (i.e., after each draw, the drawn unit is replaced back into the population). In this case, we have a sequence of $n$ independent Bernoulli trials (with probability) of success in each trial as $p = \frac{a}{N}$. Thus

$$X \sim \text{Bin}(n, \frac{a}{N}).$$

**Case II.** Without replacement (i.e., drawn units are not replaced back into the population).

Here

$$\begin{align*}
P(\text{obtaining } S \text{ in first draw}) &= \frac{a}{N}, \\
N-a-1 \\
X &+ \frac{a}{N} = \frac{a}{N}.
\end{align*}$$

In general

$$P(\text{obtaining } S \text{ in } i\text{th trial}) = \frac{a}{N}, \quad (i = 1, \ldots, n)$$

(Exercise)

$$P(\text{obtaining } S \text{ in first and second trial})$$

$$= \frac{a}{N} \cdot \frac{a-1}{N-1}$$

$$= \frac{a}{N} \cdot \frac{a}{N} = P(\text{obtaining } S \text{ in first trial}) \times P(\text{obtaining } S \text{ in second trial})$$

[27/5]
\[
\begin{align*}
\Rightarrow & \quad \text{Drawn are not independent} \\
\text{Then we can not conclude that } & \quad X \sim \text{Bin}(n, \frac{a}{h}) \text{.} \\
\text{let } & \quad x \sim \{ \frac{a}{h} \}
\end{align*}
\]

\[
\begin{align*}
b_x(x) &= P(X=x) = \left\{ \begin{array}{ll}
\binom{a}{x} \binom{n-a}{n-x} & \text{for } x = \min \{ a, n \} \\
0 & \text{otherwise}
\end{array} \right.
\end{align*}
\]

\[ \Rightarrow \quad \text{Hypergeometric distribution (Hyp}(a, \frac{y}{N})) \text{.} \]

For real \( y \) let \( X(y) = \sum_{x=\min \{ a, n \}}^{\min \{ a, n \}} x \). Then

\[
E(X(y)) = \frac{1}{N} \sum_{k=\min \{ a, n \}}^{\min \{ a, n \}} k \binom{k}{a} \binom{n-k}{n-a}.
\]

Clearly for \( y > \min \{ a, n \} \), \( E(X(y)) > 0 \). For \( 1 \leq y \leq \min \{ a, n \} \),

\[
\begin{align*}
E(X(y)) &= \frac{1}{N} \sum_{k=\min \{ a, n \}}^{\min \{ a, n \}} k \binom{k}{a} \binom{n-k}{n-a} \\
&= \frac{a(v)}{N} \sum_{k=\min \{ a, n \}}^{\min \{ a, n \}} \binom{k}{v} \binom{n-k}{n-a-v} \\
&= \frac{a(v)}{N} \sum_{k=\min \{ a, n \}}^{\min \{ a, n \}} \binom{n-k+v}{n-v} \binom{n-k}{n-v-k} \\
&= \frac{a(v)}{N} \sum_{k=\min \{ a, n \}}^{\min \{ a, n \}} \binom{n-k}{n-v} \binom{n-k}{n-v-k} \\
&= \frac{a(v)}{N} \frac{(N-v)}{n-v} \binom{n-v}{a(v)}
\end{align*}
\]
\[ \sum_{k=0}^{m} \binom{b}{k} \binom{n-b}{m-k} = \binom{n}{m}. \]

Thus, for \( r \in \mathbb{N} \)
\[ E(X(r)) = \begin{cases} \frac{r \cdot n}{N} & \text{if } r < \min(n, a) \\ 0 & \text{if } r \geq \min(n, a) \end{cases} \]

In particular
\[ E(X) = E(X(1)) = \frac{n \cdot a}{N} = np \left( \frac{a}{N} \right), \quad \text{where } p = \frac{a}{N}. \]
\[ E(X(X-1)) = E(X(2)) = \frac{n(n-1)}{N(N-1)} a(a-1) \]

\[ \text{Var}(X) = E(X^2) - (E(X))^2 \]
\[ = E(X(X-1) + E(X) - (E(X))^2 \]
\[ = n \left( \frac{a}{N} \right) \left( 1 - \frac{a}{N} \right) \frac{n-1}{N-1} \]
\[ = np(1-p) \left( 1 - \frac{n-1}{N-1} \right) \ldots \ldots \quad (\ast) \]

**Remarks:**
In case of sampling with replacement we have \( X \sim \text{Bin}(n, p) \).
\[ E(X) = np \text{ and } \text{Var}(X) = np(1-p), \text{ where } p = \frac{a}{N}. \]

The factor \( \left( 1 - \frac{n-1}{N-1} \right) \), which on multiplying to variance of \( \text{Bin}(n, p) \) distribution yields the variance of \( \text{Hyp}(a, n, N) \) distribution (see (1) in the finite population correction (f.p.c.)). Clearly if the sample size is significantly smaller than the population size \( N \) \( (n \ll N) \) then f.p.c. will be close to 1 and variances of \( \text{Bin}(n, p) \) and \( \text{Hyp}(a, n, N) \) distributions will be very close. In fact, when \( n \ll N \) and \( n < a \leq \frac{a}{N} \) we have \( \frac{a}{N} \) is a fixed quantity (i.e., as \( N \rightarrow \infty \), \( a \rightarrow 0 \)) and
\[ \frac{\frac{a}{N}}{\frac{N}{N}} = 1, \quad \text{where } \frac{a}{N} \text{ is a fixed quantity} \]

Thus \( \text{Bin}(n, \frac{a}{N}) \) distribution provides an approximation to \( \text{Hyp}(a, n, N) \) distribution. Regarding choice of sample
Let \( X_{a_n, N} \sim \text{hypergeom}(a_n, N) \), where \( a_n \) depends on \( N \).

**Theorem 2.2**

Let \( X_{a_n, N} \sim \text{binomial}(a_n, N) \), where \( a_n \) depends on \( N \). Then

\[ \lim_{N \to \infty} \frac{X_{a_n, N}}{N} = p \in (0, 1). \]

Let \( \tilde{b}_{a_n, N}(k) \) denote the p.m.f. of \( X_{a_n, N} \). Then

\[ \lim_{N \to \infty} \tilde{b}_{a_n, N}(k) = \lim_{N \to \infty} \frac{\text{binomial}(a_n, N)}{N} = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \in \{0, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}. \]

i.e., for large \( N \) and fixed \( a_n \), \( \tilde{b}_{a_n, N}(k) \) is a fixed quantity, \( \text{binomial}(a_n, N) \) probabilities can be approximated by \( \binom{n}{k} p^k (1-p)^{n-k} \) probabilities.

**Proof:**

\[ S_x = \sum_{i=1}^{N} \mathbb{1}(X_{a_n, N} = k) \]

\[ = \frac{N}{N} \begin{pmatrix} N \end{pmatrix} \begin{pmatrix} a_n \end{pmatrix} \begin{pmatrix} N-a_n \end{pmatrix} \]

\[ \to \begin{pmatrix} N \end{pmatrix} \begin{pmatrix} k \end{pmatrix} \begin{pmatrix} N-k \end{pmatrix} \]

\[ = \binom{n}{k} p^k (1-p)^{n-k} \]

\[ \lim_{N \to \infty} \tilde{b}_{a_n, N}(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k \in \{0, \ldots, n\} \\ 0 & \text{otherwise} \end{cases}. \]

The m.f.d. of \( X_{a_n, N} \), although finite (since \( X_{a_n, N} \) is finite), can not be expressed in closed form.
5.2.4. The Poisson Distribution

Some event E (e.g., accidents in a crossing) is occurring randomly over a fixed time.

\[ X = \# \text{ of times event E has occurred in an unit interval } (0,1) \]

To model probability distribution of \( X \), partition the unit interval into a large number \((N)\) of infinitesimal sub-intervals \((\frac{k}{N}, \frac{k+1}{N})\), \(k=1, \ldots, N\) of length \(\frac{1}{N}\) each. It may be relevant to assume that

1. For each infinitesimal interval \((\frac{k-1}{N}, \frac{k}{N})\), \(k=1, \ldots, N\), the probability that E will occur in this interval is \(\lambda\) and that it will not occur in this interval in \(1-\lambda\); hence \(\lambda \rightarrow 0\) as \(N \rightarrow \infty\).

2. Chance of two or more occurrences of E in any infinitesimal sub-interval \((\frac{k-1}{N}, \frac{k}{N})\), \(k=1, \ldots, N\), is negligible that it can be neglected.

3. Occurrences of E in two disjoint infinitesimal intervals are independent.

\[ X \equiv X = \# \text{ of times event E occurs in } (0,1) \]

\[ \sim \text{ Poisson}(\lambda N) \]

The p.m.f. of \( X \) is

\[ f_X(k) = \binom{N}{k} \lambda^k (1-\lambda)^{N-k} I_{k=1}^{N} \]

\[ = \frac{1}{N^k} (1-\lambda)^{N-k} \ldots (1-\frac{N-1}{N}) \lambda^k (1-\frac{1}{N})^{N-k} I_{\{0,1,\ldots,N\}} \]

\[ \rightarrow e^{-\lambda} \frac{\lambda^k}{k!} I_{\{0,1,\ldots,N\}} \]

\[ \text{as } N \rightarrow \infty \]
\[ n \approx \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \in \{0, 1, 2, \ldots\} \]

Note that \( \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} = e^\lambda \).

A rv \( X \) is said to have a Poisson distribution with parameter \( \lambda > 0 \) (written as \( X \sim \text{Po}(\lambda) \)) if its pmf is given by

\[ b_X(k) = P(X = k) = \begin{cases} \frac{e^{-\lambda} \lambda^k}{k!}, & k \geq 0, \text{ integer} \\ 0, & \text{otherwise} \end{cases} \]

**Theorem 5.2.4** (Poisson Approximation to Binomial Distribution)

Let \( X_n \sim \text{Bin}(n, p_n) \), \( n \geq 1, \ldots \), where \( p_n \in (0, 1) \), and \( \lim_{n \to \infty} (n p_n) = \lambda \), for some \( \lambda > 0 \). Then

\[ \lim_{n \to \infty} b_X(k) = \lim_{n \to \infty} P(X_n = k) = \begin{cases} \frac{e^{-\lambda} \lambda^k}{k!}, & k \geq 0, \text{ integer} \\ 0, & \text{otherwise} \end{cases} \]

**Proof**

An above.

**Remark 5.2.4**

If \( n \) is large and \( p \) is small (\( n p \to 0 \) as \( n \to \infty \)) such that \( n p \) is a fixed quantity in \( (0, \infty) \), then Poisson distribution provides a good approximation to Binomial distribution.

**Example 5.2.4**

Consider a person who plays 2500 games independently. If the probability of person winning any game is 0.002, find the probability that the person will win at least two games.

**Solution**

Let

\[ X = \# \text{ of wins ( successes) in 2500 games played by person} \]

\[ 32/5 \]
Clearly $X \sim \text{Bin}(2500, 0.002)$ where $n = 2500$ and
$\eta = 5$ ($\lambda, \lambda_0$) is fixed. Therefore

$$P(X \geq 2) \approx P(Y \geq 2),$$

where $Y \sim \text{Po}(5)$.

Thus

$$P(X \geq 2) \approx 1 - (P(Y=0) + P(Y=1))$$

$$= 1 - (e^{-5} + 5e^{-5}) = 0.9590$$

Suppose $X \sim \text{Po}(\lambda)$ for some $\lambda > 0$. Then for $\nu \geq 1$,

$$E(X^{(\nu)}) = E((X^{(1)}) \cdots (X^{(\nu)}))$$

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

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

$$= \lambda^{\nu}.$$

Thus

$$\mu_1 = E(X) = E(X^{(1)}) = \lambda$$

$$E(X^2) = E(X^{(1)}) + E(X) = \lambda^2 + \lambda$$

$$\Rightarrow \text{Var}(X) = E(X^2) - (E(X))^2 = \lambda (\geq \sigma^2 = \mu_2^{1/2})$$

$\text{Mean = Variance}$

$$\mu_3 = E(X^3) = \lambda (\lambda^2 + 3\lambda + 1)$$

$$\mu_4 = E(X^4) = \lambda (\lambda^3 + 6\lambda^2 + 7\lambda + 1)$$

$$\mu_3 = \lambda \Rightarrow \mu_4 = \lambda (3\lambda + 1)$$

$$\rho_1 = \frac{\mu_3}{\sigma^3} = \frac{1}{\lambda^3} \Rightarrow \rho_2 = \frac{1}{\lambda^3} - 3 = \frac{2}{\lambda^2}$$

$$= \frac{1}{\lambda}.\]
\[ M(x^t) = E(e^{xt}) \]
\[ = \sum_{k=0}^{\infty} e^{xt} \frac{e^{-\lambda} \lambda^k}{k!} \]
\[ = e^{-\lambda} \sum_{k=0}^{\infty} \left(\lambda e^t\right)^k \frac{1}{k!} \]
\[ = e^{-\lambda} e^{\lambda e^t} \]
\[ \Psi_x(t) = \ln M(x^t) = \lambda (e^t - 1) \]
\[ \Psi_x(t; i) = \lambda e^t, \quad i = 1, 2, \ldots, n \]

**Theorem 5.24.1** Let \( x_1, \ldots, x_n \) be independent r.v.'s such that \( x_i \sim \text{Po}(\lambda_i) \) for some \( \lambda_i > 0, \quad i = 1, \ldots, n \). Then \( Y = \sum_{i=1}^{n} x_i \sim \text{Po}(\lambda) \) where \( \lambda = \sum_{i=1}^{n} \lambda_i \).

**Proof.** For \( t \in \mathbb{R} \)
\[ \Psi_Y(t) = E(e^{yt}) = E(e^{t \sum_{i=1}^{n} x_i}) \]
\[ = E \left( \prod_{i=1}^{n} e^{tx_i} \right) \]
\[ = \prod_{i=1}^{n} E(e^{tx_i}) \quad (\text{independence of } x_i) \]
\[ = \prod_{i=1}^{n} e^{\lambda_i (e^t - 1)} \]
\[ = e^{-\lambda} \sum_{k=0}^{\infty} \left(\lambda e^t\right)^k \frac{1}{k!} \]
\[ = e^{-\lambda} e^{\lambda e^t} \]

\[ \Rightarrow Y \sim \text{Po}(\lambda), \quad \text{where } \lambda = \sum_{i=1}^{n} \lambda_i \]
5.25. The Discrete Uniform Distribution

N: a given positive integer

\(\lambda_1 < \lambda_2 < \ldots < \lambda_N: \) given real numbers

A r.v. \(X\) is said to follow a discrete uniform distribution on the set \(\{\lambda_1, \ldots, \lambda_N\}\) (written as \(X \sim U(\{\lambda_1, \ldots, \lambda_N\})\)).

If \(X\) is p.m.f. is given by

\[
P(x = \lambda_i) = \frac{1}{N}, \quad \lambda_i \in \{\lambda_1, \ldots, \lambda_N\}
\]

Suppose that \(X \sim U(\{\lambda_1, \ldots, \lambda_N\})\). Then,

\[\mu_Y = E(X^r) = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^r\]

\[\text{Mean} = \mu_X = \frac{1}{N} \sum_{i=1}^{N} \lambda_i\]

\[\text{Var}(X) = \sigma^2 = E[(X-\mu_X)^2] = \frac{1}{N} \sum_{i=1}^{N} (\lambda_i - \mu_X)^2\]

m.o.e., \(\mu_Y(1) = E(e^{tx}) = \frac{1}{N} \sum_{i=1}^{N} e^{t\lambda_i}\)

Suppose that \(Y \sim U(\{1, 2, \ldots, N+1\})\). Then,

\[\mu_Y = \text{Mean} = E(Y) = \frac{1}{N+1} \sum_{i=1}^{N+1} i = \frac{N+1}{2}\]

\[\mu_2 = E(Y^2) = \frac{1}{N+1} \sum_{i=1}^{N+1} i^2 = \frac{(N+1)(2N+1)}{6}\]

\[\mu_3 = E(Y^3) = \frac{1}{N+1} \sum_{i=1}^{N+1} i^3 = \frac{N(N+1)^2}{4}\]

\[\mu_4 = E((Y-\mu_Y)^2) = \frac{1}{N+1} \sum_{i=1}^{N+1} (i-\mu_Y)^2 = \frac{(N+1)(2N+1)(3N^2+3N-1)}{36}\]

\[\mu_5 = E((Y-\mu_Y^3)) = 0\]
\[ M_{\alpha} = E((X-\mu)\alpha) = \frac{5N^{-\alpha}(N^\alpha - 1)}{240} \]

\[ \beta_1 = \frac{\mu_3}{\mu_2} = 0 \quad (\text{Coefficient of skewness}) \]

Kurtosis = \[ 0 \]
\[ = \frac{\mu_4}{\mu_2} = \frac{3}{5} \cdot \frac{3N^{1-\alpha}}{N^{2-\alpha}} \]

The m.g.f. of \[ X \sim U(1, 2, ..., N) \] is given by
\[ \Pi_X(t) = E(e^{tX}) = \frac{1}{N} \sum_{x=1}^{N} e^{tx} \]
\[ = \begin{cases} 
\frac{e^t(N^t - 1)}{e^t - 1}, & 0 < t < \ln(N) \\
1, & t \geq \ln(N) \end{cases} \]

**Examples:** A person has to open a lock whose key is lost among a set of \( N \) keys. Assume that out of these \( N \) keys only one can open the lock. To open the lock, the person tries keys one by one by choosing at each attempt one of the keys at random from the unattempted keys. The unsuccessful keys are not considered for future attempts, let \( Y \) denote the number of attempts the person will have to make to open the lock. Show that \( Y \sim U(1, 2, ..., N) \) and hence find the mean and variance of the r.v. \( Y \).

**Solution.** For \( Y \sim U(1, 2, ..., N) \), we have \( P(Y = r) = \frac{1}{N} \). For \( r \in \{1, 2, ..., N\} \)

\[ P(Y = r) = \frac{N-1}{N} \frac{N-2}{N-1} \cdots \frac{N-(r-1)}{N-(r-2)} \frac{1}{N-(r-1)} = \frac{1}{N} \]

\[ = Y \sim U(1, 2, ..., N) \]

\[ E(Y) = \frac{N+1}{2} \quad \text{and} \quad \text{Var}(Y) = \frac{N^2 - 1}{12} \]
5.3. Some Absolutely Continuous Distributions

5.3.1 Uniform or Rectangular Distribution

Let \(-\infty < \alpha < \beta < \infty\). An absolutely continuous type on \(x\) on \(x\) is said to have a uniform (or rectangular) distribution over the interval \((\alpha, \beta)\) (written as \(X \sim U(\alpha, \beta)\)). Its p.d.f. is given by

\[
 f(x|\alpha, \beta) = \begin{cases} 
 \frac{1}{\beta - \alpha}, & \alpha < x < \beta \\
 0, & \text{otherwise}
\end{cases}
\]

Thus, the family of distributions \(U(\alpha, \beta)\) with \(-\infty < \alpha < \beta < \infty\) corresponds to different classes of \(\alpha\) and \(\beta\) \((-\infty < \alpha < \beta < \infty\))

Suppose that \(X \sim U(\alpha, \beta)\), for some \(-\infty < \alpha < \beta < \infty\). Then,

\[
\mathbb{E}[X^r] = \mathbb{E}(X^r) = \int_{\alpha}^{\beta} \frac{x^r}{\beta - \alpha} \, dx
\]

\[
= \frac{\beta^{r+1} - \alpha^{r+1}}{(r+1)(\beta - \alpha)}
\]

\[
= \frac{\beta^r}{r+1} \left[ 1 + \frac{\alpha}{\beta} + \left(\frac{\alpha}{\beta}\right)^2 + \cdots + \left(\frac{\alpha}{\beta}\right)^r \right]
\]

\[
\mathbb{E}[(X - \mu)^r] = \mathbb{E}[(X - \mu)^r]
\]

\[
= \int_{\alpha}^{\beta} \left( x - \frac{\alpha + \beta}{2} \right)^r \frac{1}{\beta - \alpha} \, dx
\]

\[
= \frac{\beta^{r+2} - \alpha^{r+2}}{r+2(\beta - \alpha)}
\]

\[
= \frac{\int_{\alpha}^{\beta} \left( \frac{\beta - x}{\beta - \alpha} \right)^r \, dx}{\frac{2}{r+2}}
\]

\[
= \left\{ \begin{array}{ll}
0, & \text{if } r = 2, 4, 6, \ldots \\
\frac{(\beta - \alpha)^r}{2^{r+1}(r+1)}, & \text{if } r = 3, 5, \ldots
\end{array} \right.
\]

\[37/5\]
Also,
\[ bX \left( \frac{x - \alpha + \beta}{2} \right) = bX \left( \frac{\alpha + \beta - x}{2} \right) = \begin{cases} \frac{1}{\beta - \alpha}, & -\frac{\beta - \alpha}{2} < x < \frac{\beta - \alpha}{2} \\ 0, & \text{otherwise} \end{cases} \]

\[ X - \frac{\alpha + \beta}{2} = \frac{\alpha + \beta - x}{2} - x \]

Distribution of \( X \) is symmetric about its mean \( \mu' = \frac{\alpha + \beta}{2} \).

- Mean: \( \mu = E(X) = \frac{\alpha + \beta}{2} \)
- Variance: \( \sigma^2 = \text{Var}(X) = \frac{\beta - \alpha}{12} \)
- Coefficient of skewness: \( \beta_1 = \frac{\frac{\beta^2 - 1}{\sigma^2}}{\mu^3} = 0 \)
- Kurtosis: \( \beta_2 = \frac{\mu_4}{\mu^2} = \frac{9}{5} = 1.8 \)

The d.f. of \( X \sim U(x, p) \) is given by
\[ f(x|\alpha, \beta) = \begin{cases} 0, & x < \alpha \\ \frac{2x - \beta + \alpha}{\beta - \alpha}, & \alpha \leq x < p \\ 1, & x \geq p \end{cases} \]

**Theorem 5.3.1:**
Let \(-\infty < x < \infty\) and let \( X \) be a rv of continuous type with \( P(\alpha < X < \beta) = 1 \). Then
\[ X \sim U(x, p) \iff P(X \in I) = P(X \in J) \text{ for any pairs of intervals } I, J \subseteq (\alpha, \beta) \text{ having the same length.} \]
Proof: Suppose that \( x \in U(x, p) \). Then, for \( a \leq c < b \leq p \),

\[
P(x \in (a, b)) = P(x \in [a, b]) = P(x \in (a, b]) = P(x \in (a, b)) = \frac{b - a}{b - a}
\]

\[\rightarrow\] depends only on length \( b - a \) of interval \((a, b) \cup (a, b) \cup (a, b) \cup (a, b)\).

Conversely, suppose that

\[P(x \in I) = P(x \in J), \quad \text{for all pairs } I, J \subseteq (a, b) \text{ having the same length}.
\]

For \( 0 \leq \lambda \leq 1 \), let

\[G(\lambda) = P(\alpha \leq x \leq \omega + (p - \alpha) \lambda) = F(\alpha + (p - \alpha) \lambda | \alpha, p).
\]

Then, for \( 0 < \lambda_1, \lambda_2 \leq 1 \),

\[G(\lambda_1 + \lambda_2) = P(\alpha \leq x \leq \omega + (p - \alpha) (\lambda_1 + \lambda_2))
\]

\[= P(\alpha \leq x \leq \omega + (p - \alpha) \lambda_1) + P(\omega + (p - \alpha) \lambda_1 \leq x \leq \omega + (p - \alpha) \lambda_2)
\]

\[\rightarrow\] depends only on length \((p - \alpha) \lambda_2 \cup (p - \alpha) \lambda_1 \cup (\omega + (p - \alpha) \lambda_1) \cup \omega + (p - \alpha) \lambda_2\).

\[= G(\lambda_1) + G(\lambda_2).
\]

By induction, for \( 0 < \lambda_1 \leq \lambda \leq \lambda_n \), \( 0 < \sum \lambda_i \leq 1 \), we have

\[G(\lambda_1 + \ldots + \lambda_n) = G(\lambda_1) + \ldots + G(\lambda_n)
\]

\[\Rightarrow G(\lambda \lambda) = \lambda G(\lambda), \quad \text{for } 0 < \lambda \leq 1, \ldots, \ldots \quad (A1)
\]

\[G(\lambda) = \lambda G(\frac{\lambda}{n} + \ldots + \frac{\lambda}{n}) = n G(\frac{\lambda}{n}) \quad \ldots \quad (A2)
\]

\[39/5\]
For $\lambda \in \{1, 2, \ldots, n\}$, we have

$$
G\left( \frac{\mu}{n} \right) = G\left( \frac{1}{n} + \ldots + \frac{1}{n} \right)
$$

where $n$ is the number of terms.

$$
= n g\left( \frac{1}{n} \right) \quad \text{(using $A_1$)}
$$

$$
= \frac{\lambda}{\lambda - \Xi} g\left( 1 \right) \quad \text{(using $A_2$)}
$$

$$
= \frac{\lambda}{\lambda - \Xi} F\left( p_{1|p} \right)
$$

$$
= \frac{\lambda}{\lambda - \Xi}
$$

Thus, $G(v) = \lambda$, $\forall v \in \mathcal{B} \cap \{0, 1\}$,

where $\mathcal{B}$ denotes the set of rational numbers. Now let $\lambda \in \mathcal{B} \cap \{0, 1\}$. Then there exists a sequence $\{v_n\}$ in $\mathcal{B} \cap \{0, 1\}$

such that $v_n \downarrow \lambda$ (rational numbers dense in $\{0, 1\}$).

Then, since $a$ is continuous,

$$
G(\lambda) = \lim_{n \to \infty} G(v_n)
$$

$$
= \lim_{n \to \infty} \lambda
$$

$$
= \lambda
$$

Thus, $G(\lambda) = \lambda$, $\forall \lambda \in \mathcal{B} \cap \{0, 1\}$.

$$
F\left( \alpha \left( \beta - \alpha \right) x \left| \alpha \left| p \right. \right. \right) = \lambda, \quad \forall \lambda \in \mathcal{B} \cap \{0, 1\}
$$

$$
F(\lambda | x, p) = \frac{\lambda - \alpha}{\beta - \alpha}, \quad \forall \lambda \in \mathcal{B} \cap \{0, 1\}
$$

$$
F(2 | x, p) = \begin{cases} 
0, & \lambda < x \\
\frac{\lambda - \alpha}{\beta - \alpha}, & \alpha \leq \lambda < \beta \\
1, & \lambda \geq \beta
\end{cases}
$$

$$
= x \in U(k | p).
$$
Theorem 5.3.2 Let \( X \sim U(\alpha, \beta) \) \(-\alpha < x < \beta < \infty\). Then

(i) for \( a > 0 \) and \( b \in \mathbb{R} \), \( Y = aX + b \sim U(a\alpha + b, a\beta + b) \)
(ii) for \( a < 0 \) and \( b \in \mathbb{R} \), \( Y = aX + b \sim U(a\alpha + b, a\beta + b) \)
(iii) \( Z = \frac{X - \alpha}{\beta - \alpha} \sim U(0, 1) \)

Proof: Straightforward

Recall that quantile function is defined as

\[ Q_x(p) = \inf \{ \lambda \in \mathbb{R} : F_x(\lambda) \geq p \} \quad 0 < p < 1 \]
Theorem 5.1. Let \( X \) be a r.v. with d.f. \( F \) and quantile function \( Q(\cdot) \). Then,

(i) (Probability Integral Transform).

\( X \) is of continuous type \( \Rightarrow F(x) \sim U(0,1) \)

(ii) \( U \sim U(0,1) \Rightarrow Q(U) \sim X \).

Proof (i) Let \( G \) be the d.f. of \( Y \equiv F(X) \). Then,

\[ G(y) = P(F(X) \leq y), \quad y \in \mathbb{R}. \]

Clearly, for \( y < 0 \), \( G(y) = 0 \) and for \( y \geq 1 \), \( G(y) = 1 \).

For \( y \in [0,1) \),

\[ G(y) = \{ x \in \mathbb{R} : F(x) \geq y \} = \{ x \in \mathbb{R} : x \geq Q(y) \} \]

\[ \Rightarrow P(F(X) \geq y) = P(X \geq Q(y)) \]

\[ = P(F(X) < y) = P(X < Q(y)) \]

\[ = P(F(X) < y) = P(X \leq Q(y)) = F(Q(y)) = y \]

Since \( X \) is of continuous type,

\[ P(F(X) = y) = 0 \quad \text{for some } y \in \mathbb{R} \] with \( F(y) = F(y + 1) \).

Thus

\[ P(F(X) \leq y) = y, \quad y \in [0,1) \]

\[ G(y) = \begin{cases} 0, & y < 0 \\ y, & 0 \leq y < 1 \\ 1, & y \geq 1 \end{cases} \]

\[ \Rightarrow U \sim U(0,1). \]
(1c) Let $U \sim U(0, 1)$ and let $Z = \Theta(U)$. Then the d.f.
of $Z$ is

\[ H(z) = P(Z \leq z) = P(\Theta(U) \leq z) = P(\Theta(U) \leq z, 0 < U < 1) \]

Note that, for $z \in (0, 1)$

\[ \{ z \in \mathbb{R} : \Theta(U) \leq z \} = \{ x \in \mathbb{R} : F(\lambda) \geq x \} \]

Thus, for $z \in (0, 1)$

\[ H(z) = P(F(\lambda) \geq z, 0 < U < 1) = P(U \leq F(\lambda)) = F(\lambda) \]

\[ \Rightarrow Z = \Theta(U) \equiv X. \]

**Remark 53.1.** The above theorem provides a method to generate observations from an arbitrary distribution using $U(0, 1)$. Suppose that we require an observation $X$ from a distribution having d.f. $F$ and quantile function $\Theta$. To do so, the above theorem suggests that, generate an observation $U$ from $U(0, 1)$ distribution and take $X = \Theta(U)$.

**5.3.2. Gamma and Related Distributions**

Gamma Function: $\Gamma: (0,\infty) \rightarrow (0,\infty)$

\[ \Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \quad x > 0 \]

Gamma Function $\rightarrow$ Converges for any $x > 0$.
Intervallation by Marta Hedly

\[ \Gamma_{n+1} = \sqrt{2} \Gamma_n, \quad n \geq 0 \]

\[ \Gamma_1 = 1 \]

For \( n \in \mathbb{N} \),

\[ \Gamma_n = \frac{\Gamma_{n-1}}{n} \]

\[ \beta_2 = 2 \int_0^\infty \frac{dx}{1 + e^x} = 2 \int_0^\infty e^{-x} \, dx \]

\[ (\beta_2)^2 = \frac{4}{\pi} \int_0^{\pi/2} \frac{e^{-\tan^2 \theta}}{1 + e^{-\tan^2 \theta}} \, d\theta \]

\[ = \frac{4}{\pi} \int_0^{\pi/2} e^{-\tan^2 \theta} \, d\theta 
\]

\[ = \frac{4}{\pi} \int_0^{\pi/2} e^{-r^2} \, r \, dr \, dr \]

\[ = \frac{4}{\pi} \int_0^{\pi/2} e^{-r^2} \, dr \]

\[ = \frac{4}{\pi} \left[ -\frac{1}{2} e^{-r^2} \right]_0^{\pi/2} \]

\[ = \frac{4}{\pi} \left( 1 - 0 \right) \]

\[ = \frac{4}{\pi} \]

\[ \beta_2 = \sqrt{\pi} \]

\[ \beta_2 = \frac{1}{2} \beta_2 = \frac{\beta_2}{2} \]

\[ \beta_4 = \frac{1}{2} \beta_2 = \frac{1}{2} \sqrt{\pi} \]

\[ \sqrt{\frac{2\pi}{2}} = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \quad \Gamma_n = \frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2^n} \]

Clench

\[ \int_0^\infty e^{-\frac{x^2}{2}} \, dx = \sqrt{2\pi} \quad \alpha \geq 0 \]

**Definition 5.3.1**: A rv \( X \) is said to have a gamma distribution with shape parameter \( \alpha \) and scale parameter \( \theta \) (written as \( X \sim \text{Gamma}(\alpha, \theta) \)) if its p.d.f. is given by

\[ f(x; \alpha, \theta) = \begin{cases} \frac{1}{\theta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\theta}} x^{\alpha-1}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \]

\[ \rightarrow \text{family of distributions} \{ \text{Gamma}(x; \alpha, \theta) : \alpha \geq 0, \theta > 0 \} \]
\[ X \sim \text{GAM}(\alpha, \beta) \Rightarrow \frac{X}{\beta} \sim \text{GAM}(\alpha, 1) \quad \text{(it is called a rate parameter)}\]

Since the distribution of \( \frac{X}{\beta} \) does not depend on \( \theta \).

The p.d.f. of \( Z \sim \text{GAM}(\alpha, 1) \) is

\[
b(z) = \begin{cases} 
\frac{e^{-z}z^{\alpha-1}}{\Gamma(\alpha)} & z > 0 \\
0 & \text{otherwise}
\end{cases}
\]

Let \( Z \sim \text{GAM}(\alpha, 1) \). Then

\[
E(Z^\alpha) = \frac{1}{\beta} \int_0^{\infty} y^{\alpha-1} e^{-\frac{y}{\beta}} \, dy = \frac{\alpha}{\beta}, \quad \beta > \alpha, \quad \alpha > 0
\]

\[
= \alpha(\alpha+1) \cdots (\alpha+r-1), \quad r \in \mathbb{R}
\]

Let \( X \sim \text{GAM}(\alpha, \theta) \), \( \alpha > 0 \). Then \( Z = \frac{X}{\theta} \sim \text{GAM}(\alpha, 1) \)

\[
E \left( \left( \frac{X}{\theta} \right)^\alpha \right) = E(Z^\alpha) = \frac{\alpha}{\theta}
\]

\[
E(X^\alpha) = \frac{\Gamma(\alpha+\theta)}{\Gamma(\alpha)}
\]

\[
= \alpha(\alpha+1) \cdots (\alpha+r-1), \quad r \in \mathbb{R}
\]

\[
\text{Mean} = E(X) = \alpha \theta
\]

\[
M_2 = E(X^2) = \alpha(\alpha+1) \theta^2
\]

\[
\mu_2 = \text{Var}(X) = E(X^2) - (E(X))^2 = \alpha \theta^2
\]

\[
\mu_3 = E((X-M_1)^3) = \mu_3 - 3\mu_1 \mu_2 + 2(M_1)^3 = 2 \alpha \theta^3
\]

\[
\mu_4 = E((X-M_1)^4) = M_4 - 4M_1^3 \mu_2 + 6(M_1^2 \mu_2 - 3M_1^3) = 2 \alpha(1+2\theta^2) \theta^4
\]

\[
\text{Coefficient of skewness} = \beta_1 = \frac{M_3}{M_2^{3/2}} = \frac{2}{\sqrt{\alpha}}
\]

\[
\text{Kurtosis} = \psi_3 = \frac{\mu_4}{\mu_2^2} = 6 + \frac{6}{\alpha}
\]

For \( \alpha > 1 \), \( M_1 \), \( M_2 \), and \( \mu_4 \) are all positive. \( \mu_4 \) is 10 \( \alpha^{-1-\theta} \) and \( \mu_2 \) is \( \frac{\alpha^{-1} \theta^{-1}}{(\alpha-1)\theta} \)
\[ f(x) \sim 0.5 \]

\[ \lambda = 2.1 \]

\[ \sum_{i=1}^{k} e^{x_i} = E(e^{x_i}) = E(e^{1.1x_i}) \quad (\lambda = \frac{x}{\theta}) \]

\[ \sum_{i=1}^{k} \frac{1}{\theta} e^{x_i} e^{\frac{1}{\lambda} x_i} \]

\[ \sum_{i=1}^{k} \frac{1}{\theta} e^{-(1+\lambda) x_i} \frac{b^i}{i!} \]

\[ = (1+\lambda)^{-\lambda} \quad \lambda < 1 \]

**Theorem 5.3.1**

Let \( x_1, \ldots, x_k \) be independent and let \( x_i \sim \text{Gam}(\mu_i, \theta) \), \( \alpha(\mu_i, \theta) \in (0, \infty) \). Then \( \gamma = \sum_{i=1}^{k} x_i \sim \text{Gam}(\sum_{i=1}^{k} \lambda_i, \theta) \)

**Proof.** \[ M(\gamma) = \prod_{i=1}^{k} M(x_i) = \prod_{i=1}^{k} \frac{(1+\lambda_i)^{-\lambda_i}}{\lambda_i} \]

\[ = (1+\lambda)^{-\lambda} \quad \lambda < 1 \]

5.3.21 (Relationship Between Gamma and Poisson Distribution)

**Theorem**

For \( n \in \mathbb{N} \), \( \theta > 0 \) and \( t > 0 \), let \( X \sim \text{Gam}(n, \theta) \) and \( Y \sim \text{Po}(\frac{t}{\theta}) \). Then

\[ P(X > t) = P(n \leq y - 1) \]

i.e.

\[ \sum_{y=0}^{\infty} \frac{e^{-\frac{t}{\theta}} \left( \frac{t}{\theta} \right)^y}{y!} \]

**Proof.** Use integration by parts.

**Remark**

For \( n \in \mathbb{N} \), \( \theta > 0 \), let \( X \sim \text{Gam}(n, \theta) \). Then

\[ \sum_{j=0}^{n} e^{rac{x}{\theta}} \left( \frac{x}{\theta} \right)^j \quad \sim \text{Unit} \]
and \[ \sum_{j=0}^{\infty} \frac{e^{-x} (x/\theta)^j}{j!} \sim \text{U} (21) \] \( \Rightarrow 1 - \text{U} (31) \)

**Definition 5.3.22**

For a **GAM** (Gamma) distribution, it is called an exponential distribution with scale parameter \( \theta \) (denoted by \( \text{E}(\theta) \)).

The **p.d.f**. of \( \text{T}_{\text{E}(\theta)} \) is given by

\[ f_T(t) = \frac{1}{\theta} e^{-t/\theta}, \quad t > 0 \]

and its **d.f.** is given by

\[ F_T(t) = \begin{cases} \frac{t}{\theta}, & \text{for } t \leq \theta \\ 1 - e^{-t/\theta}, & \text{for } t > \theta \end{cases} \]

Mean = \( E(T) \) = \( \theta \)

Variance = \( \mu_2 = \theta^2 \)

\( \mu_1 = E(T') = \frac{1}{\theta} \quad \forall \theta \in \text{IN} \)

Coefficient of skewness = \( \beta_1 = \frac{0}{\mu_2} = 0 \)

Kurtosis = \( \beta_2 = \frac{\mu_4}{\mu_2^2} = 3 \)

**P.G.B.**

\[ P(T > t) = (1 - e^{-t/\theta})^t < \theta \]

\[ P(T > t) = \begin{cases} 1, & \text{for } t < 0 \\ e^{-t/\theta}, & \text{for } t > 0 \end{cases} \]

For \( \lambda > 0 \), \( \gamma > 0 \),

\[ P(T > \lambda + \gamma | T > \lambda) = \frac{P(T > \lambda + \gamma)}{P(T > \lambda)} = e^{-\gamma/\theta} = P(T > \gamma) \]

\[ \Rightarrow P(T > \lambda + \gamma) = P(T > \lambda) P(T > \gamma), \quad \forall \lambda, \gamma > 0 \]

---

Lack of Memory Property

Let \( T \) denote the lifetime of a system. Given that the system has survived \( \lambda \) units of time, the probability that it will survive \( \gamma \) additional units of time is the same as the probability that a fresh unit (of age 0) will survive \( \gamma \) units of time. In other words, the system has no memory of its current age or if it is not aging with time.
Theorem 53.23

Let \( Y \) be a rv of continuous type with \( dY \sim F \) such that \( F(0) = 0 \). Then \( Y \) has Laplace property if \( \lim_{n \to \infty} F(t/n) \) is the distribution of \( \text{Exp}(\theta) \) for some \( \theta \geq 0 \).

Proof. Let \( \theta \neq \text{Exp}(\theta) \). Then obviously \( Y \) has Laplace property.

Now suppose that \( F(0) = 0 \) and \( Y \) has Laplace property. Then

\[
F(\lambda t) = F(t)F(\lambda) \quad \lambda > 0
\]

Let \( \lambda = \frac{1}{n} \).

\[
F(1/n) = F(1/n + \cdots + 1/n) = F(1/n)F(1/n) \cdots F(1/n) \quad n \to \infty
\]

Let \( \lambda = F(1/n) \) and \( \lambda \to \infty \).

\[
F(1) = \lim_{n \to \infty} F(1/n)
\]

Let \( \lambda = F(1) \) such that \( \lambda = 0 \).

\[
\lambda = 0 \Rightarrow F(1) = 0, \quad \lambda < \infty
\]

\[
\lambda = 1 \Rightarrow F(1) = F(1/1) = 1, \quad \lambda > \infty
\]

\[
\lim_{n \to \infty} F(1/n) = 0
\]

Thus \( \lambda \in (0, 1) \). Let \( \lambda = e^{-\theta} \), \( 0 < \theta \) (\( \theta = -\frac{1}{\lambda} \)).

The \( \text{min}(A_1) \)

\[
F(1) = e^{-\theta/\lambda}, \quad 0 < \lambda \leq \lambda_{10}
\]

Let \( \lambda \in (0, 0) \). Then there exists a sequence \( \{y_n\}_{n=1}^{\infty} \) of \( 0 < y_n < \lambda \).

\[
F(1) = F(\lim y_n) = \lim F(y_n) = \lim e^{-y_n/\lambda} = e^{-\lambda_0} = e^{-\theta/\lambda}
\]

Thus \( F(1) = \int_{0}^{\theta} e^{-y_0} dy_0 = \frac{1 - e^{-\theta}}{\theta} \geq 0 \)

Thus \( \lambda \sim \text{Exp}(\theta) \).
Example 5.3.21

X: Waiting time for occurrence of an event E

Suppose that $X \sim \text{Exp}(3)$. Then the conditional probability that the waiting time for occurrence of E is at least 5 given that it has not occurred in first two hours is $P(X > 5 | X > 2) = 1(X > 5) = e^{-1}$.

Chi-Squared Distribution

Let $n > 0$. The $\chi^2(n)$ distribution is a chi-squared distribution with $n$ degrees of freedom (denoted by $\chi^2$).

Let $X \sim \chi^2$. The pdf of $X$ is

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2 - 1} e^{-x/2} & x > 0, \\ 0 & \text{otherwise} \end{cases}$$

Mean = $E(X) = n$,

Variance = $\mu^2 = \sigma^2 = 2n$

Coefficient of skewness = $\beta_1 = 2 \sqrt{2/n}$

Kurtosis = $\beta_2 = 3 + \frac{12}{n}$

Theorem 5.3.2.7

Let $X_1, \ldots, X_k$ be independent with $X_i \sim \chi^2(n_i)$.

The $\sum_{i=1}^{k} X_i \sim \chi^2(n)$ where $n = \sum_{i=1}^{k} n_i$.

For various values of $n$ and $\chi^2(0.1)$, tables for $\chi^2(n)$ quantiles of $\chi^2$ distribution (i.e., $x_0.1$ satisfying $P(X \leq x_0.1) = 0.1$) are available in various textbooks.
For $\alpha > 0$, $\beta > 0$

$$\Gamma(\alpha, \beta) = \int_0^\infty e^{-(\alpha u + \beta)} u^{\alpha-1} + \beta u^{\beta-1} du$$

$$= \int_0^\infty e^{-(\alpha u + \beta)} (\alpha u + \beta)^{\beta-1} du$$

$$= \Gamma(\alpha) \int_0^\infty u^{\alpha-1} (1-u)^{\beta-1} du$$

$$= \Gamma(\alpha) \Gamma(\beta) \Gamma(\alpha + \beta)$$

$$= \frac{\Gamma(\alpha)}{\alpha}$$

$$B(\alpha, \beta) \rightarrow \text{Beta function} \left( \frac{\alpha}{\beta} \right)$$

Note: $B(\alpha, \beta) = B(\beta, \alpha)$

**Definition:** For given $\alpha > 0$ and $\beta > 0$, a $\theta \in [0, 1]$ is said to have a Beta distribution with parameters $(\alpha, \beta)$ (written as $X \sim \text{Be}(\alpha, \beta)$) if the pdf of $X$ is given by

$$f(x | \alpha, \beta) = \left\{ \begin{array}{ll}
\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \\
0 & \text{otherwise}
\end{array} \right.$$  

Suppose that $X \sim \text{Be}(\alpha, \beta)$, for some $\alpha > 0$ and $\beta > 0$. Then

$$E(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)} = \frac{\alpha}{\alpha+\beta}$$

$$\text{Mean} = M'_1 = E(x) = \frac{\alpha}{\alpha+\beta}$$

$$M_2 = E(x^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

$$\text{Var}(x) = \text{Var}(x) = E(x^2) - (E(x))^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$
Node: \( N_0 = \frac{a-1}{a-1} p_{-1} \), \( a \geq 1 \) and \( a + p \geq 2 \)

Statement: \( p_1 = \frac{a_b}{m_2} = \frac{2(b-a) \sqrt{a+b+1}}{\sqrt{a+b} (a+b+2)} \)

Kurtosis: \( J_1 = \frac{m_4}{m_2^2} = 6 \left[ \frac{(b-p)^2 (a+b+2) - b^2 (a+b+2)}{a+b} \right] \)

Let \( X \sim Be(\alpha, \beta) \) \( \alpha > 0, \beta > 0 \)

\[ f(x|\alpha, \beta) = \begin{cases} \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \]

Thus, if \( X \sim Be(\alpha, \beta) \) then the distribution of \( \frac{X}{\alpha + \beta} \) is symmetric about \( \frac{\alpha}{\alpha + \beta} \).

**Sasi Relationship between Beta and Bimomial Distribution**

**Theorem** For \( \mu \in \mathbb{N} \) and \( \alpha \in (0, 1) \), let \( X \sim Be(\mu, \mu) \) and \( Y \sim Bin(\mu + \mu - 1, \alpha) \). Then

\[ P(X \leq x) = P(Y \geq \mu) \]

i.e., \[ \frac{1}{B(\mu, \mu)} \int_0^x t^{\mu-1} (1-t)^{\mu-1} dt = \sum_{j=\mu}^{\mu+\mu-1} \binom{\mu+\mu-1}{j} x^j (1-x)^{\mu+\mu-1-j} \]

**Proof** Fix \( \mu \in \mathbb{N} \) and \( \alpha \in (0, 1) \). Let

\[ LHS = \frac{(\mu+\mu-1)}{\mu} \int_0^x t^{\mu-1} (1-t)^{\mu-1} dt \]

= \[ \left( \frac{\mu+\mu-1}{\mu} \right) x^\mu (1-x)^{\mu-1} + \left( \frac{\mu+\mu-1}{\mu} \right) \int_0^x t^{\mu-1} (1-t)^{\mu-1} dt \]

\[ = \left( \frac{\mu+\mu-1}{\mu} \right) x^\mu (1-x)^{\mu-1} + LHS_{\mu+\mu-1, \mu} \]

\[ \frac{1}{B(\mu, \mu)} \int_0^x t^{\mu-1} (1-t)^{\mu-1} dt = \sum_{j=\mu}^{\mu+\mu-1} \binom{\mu+\mu-1}{j} x^j (1-x)^{\mu+\mu-1-j} \]

\[ \frac{1}{B(\mu, \mu)} \int_0^x t^{\mu-1} (1-t)^{\mu-1} dt = \sum_{j=\mu}^{\mu+\mu-1} \binom{\mu+\mu-1}{j} x^j (1-x)^{\mu+\mu-1-j} \]
**Example 53.1** Time (in hours) to finish a job follows beta distribution with mean $\frac{1}{3}$ hr. and variance $\frac{2}{63}$ hr. Find the probability that the job will be finished in 30 minutes.

**Proof**

Define

$$X = \text{time to finish job (in hours)}$$

$$\sim B(\alpha, \beta), \ \forall \gamma.$$  

$$E(X) = \frac{1}{3}, \ \text{Var}(X) = \frac{2}{63}$$

$$\alpha = \frac{\alpha}{\alpha + \beta} = \frac{1}{3}, \ \frac{\alpha \beta}{(\alpha + \beta)^2} = \frac{1}{63}$$

$$\Rightarrow \alpha = 2, \ \beta = 4 \Rightarrow X \sim B(2, 4)$$

Required probability:

$$P(X < \frac{1}{2}) = \frac{1}{B(2, 4)} \int_0^{1/2} x(1-x)^3 \, dx$$

$$= \frac{13}{16}.$$

**Theorem 53.2** Let $X_1$ and $X_2$ be independent gamma with $X_i \sim \text{Ga}(\lambda, \theta), \ \alpha_i \gamma_0 \theta \gamma_0 \leq 1/2$. Define $Y_1 = X_1 + X_2$ and $Y_2 = \frac{X_1}{X_1 + X_2}$. Then $Y_1$ and $Y_2$ are independently distributed with $Y_1 \sim \text{Ga}(\alpha_1 + \alpha_2, \theta)$ and $Y_2 \sim \text{Be}(\alpha_1, \alpha_2)$.

(b) Let $X_1$ and $X_2$ be i.i.d. Exp(\theta) rv. Then $Y_2 = \frac{X_1}{X_1 + X_2} \sim \text{Unif}(0, 1)$. $\frac{1}{52/5}$
Proof (a) The joint pdf of \( X \equiv (X_1, X_2) \) is

\[
\begin{align*}
    f_{X_1}(x_1, x_2) &= \frac{1}{\alpha_1 \Gamma(\theta_1)} \left( \frac{x_1}{\theta_1} \right)^{x_1-1} e^{-\frac{x_1}{\theta_1}} I_{[0, \infty)}(x_1) \\
    f_{X_2}(x_2) &= \frac{1}{\alpha_2 \Gamma(\theta_2)} \left( \frac{x_2}{\theta_2} \right)^{x_2-1} e^{-\frac{x_2}{\theta_2}} I_{[0, \infty)}(x_2)
\end{align*}
\]

\[
\begin{align*}
    f_{X_1}(x_1, x_2) &= \frac{1}{\alpha_1 \Gamma(\theta_1) \theta_2 \alpha_2} \left( \frac{x_1}{\theta_1} \right)^{x_1-1} e^{-\frac{x_1}{\theta_1}} \left( \frac{x_2}{\theta_2} \right)^{x_2-1} e^{-\frac{x_2}{\theta_2}} I_{[0, \infty)}(x_1) I_{[0, \infty)}(x_2)
    = f_{X_1}(x_1) f_{X_2}(x_2), \quad 0 < x_1, x_2 < \infty,
\end{align*}
\]

Hence \( S_X = (0, \infty)^2 \).

Let

\[
    h_1(x_1, x_2) = y_1 = x_1 + x_2
\]
and

\[
    h_2(x_1, x_2) = y_2 = \frac{x_1}{x_1 + x_2}
\]

Then \( h : \mathbb{R}^2 \to \mathbb{R}^2 \) is 1-1 with inverse

\[
    \begin{align*}
        h_1^{-1}(y_1, y_2) &= y_1, \\
        h_2^{-1}(y_1, y_2) &= y_1 (1 - y_2)
    \end{align*}
\]

\[
    J = \begin{vmatrix} y_1 & y_1 \\ y_2 & -y_2 \end{vmatrix} = -y_1
\]

\[
    E[S_X^k] = \int_{E_2} y_1 y_2^k d(1 - y_2) = \int_{(0, \infty)} y_1 (1 - y_2)^k d(1 - y_2) = \int_{(0, 1)} y_1 I_{(0, 1)}(y_1) d(1 - y_2)
\]

Then the joint pdf of \( Y \equiv (Y_1, Y_2) \) is

\[
    f_Y(y_1, y_2) = \frac{1}{\alpha_1 \Gamma(\theta_1) \theta_2 \alpha_2} \left( \frac{y_1}{\theta_1} \right)^{y_1-1} \left( \frac{1 - y_2}{\theta_2} \right)^{y_2-1} e^{-\frac{y_1}{\theta_1} - \frac{y_2}{\theta_2}} I_{[0, \infty)}(y_1) I_{[0, \infty)}(1 - y_2)
\]

\[
\begin{align*}
    f_Y(y_1, y_2) &= \frac{y_1 y_2^{y_2-1}}{\theta_1 \Gamma(\theta_1) \theta_2 \Gamma(\theta_2)} e^{-\frac{y_1}{\theta_1} - \frac{y_2}{\theta_2}} I_{[0, \infty)}(y_1) I_{[0, \infty)}(1 - y_2)
    = f_{Y_1}(y_1) f_{Y_2}(y_2),
\end{align*}
\]

where \( Y_1 \sim \text{Ga}(\alpha_1 + x_1, \theta_1) \) and \( Y_2 \sim \text{Be}(x_1, x_2) \).

Clearly \( Y_1 \) and \( Y_2 \) are independent.
5.3.4. Normal Distribution

Recall that
\[ \sqrt{\pi} = \frac{1}{\sqrt{2}} = \int_0^\infty e^{-t^2} \, dt \]
\[ = 2 \int_0^\infty e^{-t^2} \, dt = \int_0^\infty e^{-t^2} \, dt \]
\[ = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-t^2} \, dt \]
\[ = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = 1 \quad \text{a.e. in } \mathbb{R} \text{ and } \sigma > 0. \]

**Definition** Let \( \mu \in \mathbb{R} \) and \( \sigma > 0 \) be given constants. An absolutely continuous type of \( x \) is said to follow a normal distribution with parameters \( \mu \in \mathbb{R} \) and \( \sigma > 0 \) (written as \( X \sim N(\mu, \sigma^2) \)) if its probability density function is given by
\[ f(x | \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty. \]

The \( N(0,1) \) distribution is called the standard normal distribution. The p.d.f. and d.f. of standard normal distribution are denoted by \( \phi(3) \) and \( \Phi(3) \), respectively. No that
\[ \Phi(3) = \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} \, dx = \frac{3}{2} \quad \text{and} \quad \Phi(3) = \frac{3}{2\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} \, dx. \]

\[ \therefore \quad X \sim N(\mu, \sigma^2) \Rightarrow f(x | \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in \mathbb{R} \]
\[ \therefore \quad X - \mu \sim N(0, \sigma^2) \quad ( \text{Distr. of } X \text{ is independent of } \mu ) \]
\[ \therefore \quad E(X) = \mu \quad \text{and} \quad F(\mu | \mu, \sigma) = \frac{1}{2}. \]
Moreover

\( P(x - \mu \leq \lambda) = P(\mu - x \leq \lambda) \quad \forall x \in \mathbb{R} \)

\( \Rightarrow \quad F'(\mu + x | \mu, \sigma) = 1 - F(\mu - x | \mu, \sigma) \quad \forall x \in \mathbb{R} \)

In particular

\( \Phi(0) = \frac{1}{2} \quad \text{and} \quad \Phi(1) = \Phi(-1) \quad \forall \, x \in \mathbb{R} \).

The histogram \( B(x | \mu, \sigma) \) is unimodal \((\mu, \sigma)\) and \( \nabla \mu (\mu, \sigma) = 0 \).

Thus \( \text{Mode} = \mu_0 = \mu \)

The Mean = Median = Mode = \( \mu \).

\[ \text{M. 5. 6.} \quad \text{Let} \quad X \sim N(\mu, \sigma^2). \text{ Then} \]

\[ P_x^{+1} = E(e^{tx}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \]

\[ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(M+\sigma^2/2)} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (\frac{2\mu^2}{\sigma^2} > 2) \]

\[ = e^{\mu t + \frac{\sigma^2 t^2}{2}} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(3-\sigma^2/2)}{2}} dx \]

\[ = e^{\mu t + \frac{\sigma^2 t^2}{2}} + \text{error} \]

\[ \text{Let} \quad X \sim N(\mu, \sigma^2). \text{ Then} \]

\[ Z = \frac{X - \mu}{\sigma} \sim N(0, 1). \]

\[ P_Z^{+1} = E(e^{t \frac{Z}{\sqrt{2}}} \) = e^{-\frac{\mu t}{\sqrt{2}}} \times \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = e^{\mu t} + \text{error} \]

\[ \text{M. 5. 6.} \quad \text{of } N(0, 1) \]

\[ 55/5 \]
Theorem 5.3.1

Let \( X \sim N(\mu, \sigma^2) \) and \( \Lambda \sim N(0, I) \)

(a) for \( a \neq 0 \), be \( \gamma = ax + b \sim N(a\mu + b, a^2\sigma^2) \)

(b) \( Z \overset{d}{=} \frac{X - \mu}{\sigma} \sim N(0, I) \)

(c) \[ E(Z^r) = \begin{cases} 0, & r = 1, 3, 5, \ldots \\ \frac{1}{2} \Gamma \left( \frac{r+1}{2} \right), & r = 2, 4, 6, \ldots \end{cases} \]

(d) Mean = \( \mu' = E(X) = \mu \)

Variance = \( \sigma^2 = \sigma^2 \)

Coefficient of skewness = \( \beta_1 = 0 \)

Kurtosis = \( \beta_2 = 3 \)

(e) \( Z^2 \sim X_1^2 \)

Proof.

(a) \[ \Pi_1 H_1 = E(e^{t(aX+b)}) \]

\[ = e^{bt} \Pi_1 H_1(a) \]

\[ = e^{bt} e^{\sigma^2 a^2 t^2} \]

\[ = e^{t(a\mu + b)} + \frac{e^{(a\mu + b)^2 + \sigma^2 a^2 t^2}}{2} \]

\[ \rightarrow \text{m.r.o. of } N(a\mu + b, a^2\sigma^2) \]

\[ \gamma \sim N(a\mu + b, a^2\sigma^2) \]

(b) Follows from (a) by taking \( a = \frac{1}{\sigma} \) and \( b = -\mu \)

(c) \[ \Pi_2 H_1 = E^{1/2} \bar{X} \cdot X_{12} \]

\[ = \frac{2^{2k}}{\Gamma(k)} \sum_{k=0}^{\infty} \frac{t^{2k}}{2^k \Gamma(k)} \]
\[ E(2^x) = \text{Coefficient of } \frac{t^x}{x!} \text{ in the expansion of } e^{t/H} \]

\[
\begin{cases}
0, & x = 1, 2, 3, \ldots \\
\frac{t^x}{x!}, & x = 2, 3, 4, \ldots
\end{cases}
\]

(a) \[ E\left( x \frac{M}{x} \right) = E(2) = 0 \quad \Rightarrow \quad M_1 = E(x) = M \]
\[ E\left( \left( x \frac{M}{x} \right)^2 \right) = E(2^2) = 1 \quad \Rightarrow \quad M_2 = \cdots \quad E\left( (x-M)^2 \right) = 0 \]
\[ M_3 = E\left( \left( x \frac{M}{x} \right)^3 \right) = E(2^3) = 0 \quad \Rightarrow \quad M_3 = E\left( (x-M)^3 \right) = 0 \]
\[ M_4 = E\left( \left( x \frac{M}{x} \right)^4 \right) = E(2^4) = 0 \quad \Rightarrow \quad M_4 = 3 \gamma \]
\[ = 3 \]

Coeff. of Skewness = \[ \beta_1 = \frac{M_3}{M_2^\frac{3}{2}} = 0 \]

Kurtosis = \[ \beta_2 = \frac{M_4}{M_2^2} = 3 \]

(b) Let \( r = 2^Z \). Then
\[ M_i H_1 = E(e^{rZ}) \]
\[ = \int e^{rZ} e^{-rZ^2/2} = \sqrt{2\pi} \]
\[ = \frac{1}{\sqrt{2\pi}} \int e^{-\left( 1-2t \right)Z^2/2} dZ = \left( 1-2t \right)^{-Z^2} - \frac{Z^2}{m} \]
\[ \approx \sqrt{2} \quad \text{and} \quad x \]

\[ \text{Corollary:} \quad \frac{1}{\sum_{i=1}^{k} (x_i - \bar{x})^2} \sim \chi_k \]

Let \( x_1, \ldots, x_k \) be independent and let \( X_i \sim N\left( \mu_i, \sigma_i^2 \right) \)

\[ \sum_{i=1}^{k} \frac{(x_i - \bar{x})^2}{\sigma_i^2} \sim \chi_k \]
Remark (1) In $\mathcal{N}(\mu, \sigma^2)$ distribution, the parameters $\mu$ and $\sigma^2$ are the mean and the variance of the distribution, respectively.

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then $P(\xi < x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{2} (1 - \Phi\left(\frac{x - \mu}{\sigma}\right))$.

Let $Y_\alpha$ be the $(1-\alpha)$-th quantile of $\mathcal{N}(\mu, \sigma^2)$ distribution.

The

$$P(-Y_\alpha < X < Y_\alpha) = 1 - \Phi(Y_\alpha) = \alpha$$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0.001</th>
<th>0.01</th>
<th>0.025</th>
<th>0.05</th>
<th>0.1</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_\alpha$</td>
<td>3.092</td>
<td>2.5758</td>
<td>2.326</td>
<td>1.96</td>
<td>1.64499</td>
<td>1.282</td>
</tr>
</tbody>
</table>

Table: $(1-\alpha)$-th quantiles of $\mathcal{N}(\mu, \sigma^2)$ distribution

Note: For values of $\Phi^{-1}$ for different values of $\alpha$ are available in various text books.
Example 5.3.4.1: Let $X \sim N(\mu, \sigma^2)$. Find $P(X \leq 0)$, $P(|X| \geq 2)$, $P(1 < X \leq 3)$, and $P(X \leq 3 | X > 1)$.

Solution

$P(X \leq 0) = \Phi \left( \frac{0-\mu}{\sigma} \right) = \Phi(-1) = 1 - \Phi(1) = 0.1587$

$P(|X| \geq 2) = P(X \leq -2) + P(X \geq 2)$

$= \Phi \left( \frac{-2-\mu}{\sigma} \right) + 1 - \Phi \left( \frac{2-\mu}{\sigma} \right)$

$= \Phi(-2) + 1 - \Phi(2)$

$= 0.0228 + 0.5 = 0.5228$

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

$= \Phi \left( \frac{3-\mu}{\sigma} \right) - \Phi \left( \frac{1-\mu}{\sigma} \right)$

$= \Phi(0.5) - \Phi(-0.5)$

$= 2\Phi(0.5) - 1 = 2 \times 0.6915 - 1 = 0.383$

$P(X \leq 3 | X > 1) = \frac{P(1 < X \leq 3)}{P(X > 1)} = \frac{0.383}{1 - \Phi \left( \frac{1-\mu}{\sigma} \right)} = 0.383 / 0.5 = 0.766$

Theorem 5.3.4.2: Let $x_1, \ldots, x_k$ be independent $N(\mu_i, \sigma^2_i)$ and let $X \sim N(\mu, \sigma^2)$. Then

$\sum_{i=1}^{k} a_i x_i \sim N \left( \sum_{i=1}^{k} a_i \mu_i, \frac{\sum_{i=1}^{k} a_i^2 \sigma^2_i}{\sigma^2} \right)$.

Proof

$\prod_i H_i = E \left( e^{\sum_{i=1}^{k} a_i x_i} \right)$

$= E \left( \prod_{i=1}^{k} e^{a_i x_i} \right)$

$= \prod_{i=1}^{k} E \left( e^{a_i x_i} \right)$ (independence)

$= \prod_{i=1}^{k} e^{\mu_i + \frac{1}{2} a_i^2 \sigma^2_i}$

$= \prod_{i=1}^{k} e^{\mu_i + \frac{1}{2} a_i^2 \sigma^2_i}$

$= \prod_{i=1}^{k} e^{\mu_i + \frac{1}{2} a_i^2 \sigma^2_i} \frac{e^{\frac{1}{2} a_i^2 \sigma^2_i}}{e^{\frac{1}{2} a_i^2 \sigma^2_i}}$

$\sim N \left( \sum_{i=1}^{k} a_i \mu_i, \frac{\sum_{i=1}^{k} a_i^2 \sigma^2_i}{\sigma^2} \right)$
By uniqueness of m.g.f.'s $\gamma \sim N\left( \frac{k}{c_i}, \frac{c_i}{\sigma_i} \right)$.

**Theorem 5.5.3**

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution, where $\mu \in \mathbb{R}$ and $\sigma > 0$. Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2$ denote the sample mean and sample variance, respectively. Then,

(i) $\overline{X} \sim N(\mu, \sigma^2/n)$.

(ii) $\bar{X}$ and $S^2$ are independent

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

(iv) $E(S^2) = \sigma^2$, $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$, $E(S^2) = \frac{1}{E(S^2)} - 1$.

**Proof.** (i) Follows from last theorem by taking $a_i = \frac{1}{n}$, $c_i = \mu$, $\sigma_i = \sigma$, $x_i = x_i - \bar{X}$.

(ii) Let $Y_i = X_i - \bar{X}$, $i = 1, \ldots, n$ and let $Z = (Y_1, \ldots, Y_n)$.

Then

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} (X_i - \bar{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i - n\bar{X} = 0$$

$$(n-1) S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} Z_i^2 \quad (a \text{ function of } Z)$$.

The joint m.g.f. of $\left( \bar{X}, S^2 \right)$ is given by

$$\phi_{\bar{X}, S^2}(t) = E \left( e^{t\bar{X} + sS^2} \right) = \left( \frac{\sigma^2}{\sigma^2 + \frac{s}{n-1}} \right)^{\frac{n-1}{2}} e^{\frac{t^2}{2(n-1)}}$$.

$$\sum_{i=1}^{n} t_i \bar{X} + s \sum_{i=1}^{n} X_i$$

$$= \sum_{i=1}^{n} t_i (X_i - \bar{X}) + s \sum_{i=1}^{n} X_i$$

$$= \sum_{i=1}^{n} \left( t_i \bar{X} + \frac{s(t_i - \bar{T})}{n} \right) X_i$$

$\text{where } \bar{T} = \frac{1}{n} \sum_{i=1}^{n} t_i$.

$$= \sum_{i=1}^{n} \left( \frac{1}{n} \sum_{i=1}^{n} t_i \right) X_i = \frac{1}{n} \sum_{i=1}^{n} X_i$$

(continued on next page)
Then \( \sum_{i=1}^{n} U_i = t_{0.01} \) and \( \sum_{i=1}^{n} u_i = \sum_{i=1}^{n} (x_i - \bar{x}) \).

\( \sum_{i=1}^{n} U_i = \sum_{i=1}^{n} \left( x_i - \bar{x} \right) \),

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

\( \Pi_{X,X}(t) = E \left( e^{\sum_{i=1}^{n} u_i x_i} \right) \)

\( = \prod_{i=1}^{n} \Pi_{X_i}(u_i) \)

\( = \prod_{i=1}^{n} e^{\mu u_i + \frac{1}{2} \sigma^2 u_i} \)

\( = e^{\mu t_{0.01} + \frac{1}{2} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \)

\( = \left\{ e^{\frac{\mu t_{0.01} + \frac{1}{2} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}{2}} \right\} \times \left\{ e^{\frac{\sigma^2}{2}} \right\} \)

\( \Pi_{X}(t) = \Pi_{X}(t_1 \ldots t_n) = \Pi_{X}(0 \ldots 0 + t_{0.01}) = e^{\frac{\mu t_{0.01} + \frac{1}{2} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}{2}} \)

\( \Pi_{X}(t) = \Pi_{X}(0 \ldots 0 + t_{0.01}) = e^{\frac{\mu t_{0.01} + \frac{1}{2} \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}{2}} \).

\( \Pi_{Y}(t) = \Pi_{Y}(t_1 \ldots t_n) \Pi_{Y}(t_{0.01}) \), (as \( t_1 \ldots t_n \) and \( t_{0.01} \) are independent)

\( \Pi_{Y}(t) = \Pi_{Y}(t_1 \ldots t_n) \Pi_{Y}(t_{0.01}) \).

\( \Rightarrow \) \( Y \) and \( X \) are independent

\( \Rightarrow \) \( \sum_{i=1}^{n} (x_i - \bar{x})^2 \) and \( X \) are independent.

\( (\text{iii}) \) Let \( Z_1 = \frac{X_1 - \mu}{\sigma} \), \( \ldots \), \( Z_n \). Then \( Z_1, \ldots, Z_n \) are \( \text{i.i.d. } N(0,1) \).

\( \Rightarrow \) Also \( Z = \frac{\sqrt{n}(X - \mu)}{\sigma} \) \( \sim \text{N}(0,1) \) (using (vi)).

Let \( W = Z^2 = \frac{n(X - \mu)^2}{\sigma^2} \) and \( T = \frac{(n - 1)S^2}{\sigma^2} \).

Then, by (iii), \( W \) and \( T \) are independent \( \text{i.i.d.} \). Also
\[ W \sim x_i^1, \quad \text{and} \quad V = \sum_{i=1}^{n} 2x_i \sim x_i^2. \]

\[
V = \sum_{i=1}^{n} 2x_i^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2
\]

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

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

\[
= \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{x} - \mu)^2}{\sigma^2}
\]

\[
= T + W
\]

\[\Rightarrow \quad \frac{\text{MVH}}{\text{MVH}} = \frac{\mu_{Ti}}{\mu_{Wi}} = \frac{(1-2k)^{\frac{2k}{2}}}{}^{+} + \frac{k}{2}
\]

\[\Rightarrow \quad \mu_{Ti} = \frac{\text{MVH}}{\text{MVH}} = \frac{(1-2k)^{\frac{2k}{2}}}{}^{+} + \frac{k}{2}
\]

\[\Rightarrow \quad \mu_{Ti} \sim x_{n-1}^k
\]

\[\Rightarrow \quad T = \frac{(n-1)S^2}{\sigma^2} \sim x_{n-1}^k
\]

(W) \[ T = \frac{(n-1)S^2}{\sigma^2} \sim x_{n-1}^k, \quad \text{where} \quad 0 < n-1. \]

Thus,

\[ E(T) = \int_0^2 \frac{t^2}{2} e^{-\frac{t^2}{2}} + \frac{2^{\frac{1}{2}}}{\Gamma^n} dt
\]

\[ = \frac{1}{2\sigma} \left[ \frac{\sqrt{2\pi}}{2} \right] e^{-\frac{t^2}{2}} + \frac{2^{\frac{1}{2}}}{\Gamma^n} \right]_0^\infty dt
\]

\[ = \frac{2^{\frac{1}{2}}}{2\sigma} \left[ \frac{\sqrt{2\pi}}{2} \right] - \frac{2^{\frac{1}{2}}}{\Gamma^n} \right]_0^\infty \]

\[ \Rightarrow \quad \frac{(n-1)^\frac{5}{2}}{\sigma^2} E(S^2) = 2^{\frac{1}{2}} \frac{\sqrt{2\pi}}{\Gamma^n} \]

\[ \Rightarrow 62/5 \]
\[ E(S^2) = \left( \frac{2}{h-1} \right)^\frac{1}{2} \sqrt{\frac{\sum_{i=1}^{h} x_i^2}{h-1}} \]

\[ E(S') = \left( \frac{2}{h-1} \right)^\frac{1}{2} \sqrt{\frac{\sum_{i=1}^{h} x_i^2}{h-1}} - \frac{\sum_{i=1}^{h} x_i}{h} \]

\[ E(S) = \sqrt{\frac{2}{h-1}} \sqrt{\frac{\sum_{i=1}^{h} x_i^2}{h-1}} \]

\[ E(S') = \left( \frac{2}{h-1} \right)^\frac{1}{2} \sqrt{\frac{\sum_{i=1}^{h} x_i^2}{h-1}} - \frac{\sum_{i=1}^{h} x_i}{h} \]

\[ \text{Var}(S') = E(S') - (E(S'))^2 = \frac{2\sigma^2}{h-1} \]

**Remark 5.3.4.2**

Let \( x_1, \ldots, x_n \) be a random sample from a distribution having p.m.f. / p.d.f. b. \( X \)

\[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \] and \( S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2 \).

\[ E(x_i) = \mu \] and \( \text{Var}(x_i) = \sigma^2 \). Then,

\[ E(\bar{X}) = E\left( \frac{1}{n} \sum_{i=1}^{n} x_i \right) = \mu \quad \text{Var}(\bar{X}) = \text{Var}\left( \frac{1}{n} \sum_{i=1}^{n} x_i \right) = \frac{\sigma^2}{n} \]

\[ E \left( (h-1) S^2 \right) = E \left[ \sum_{i=1}^{h} (x_i - x)^2 \right] \]

\[ = 1 \left( h-1 \right) E(S^2) = E \left[ \sum_{i=1}^{h} (x_i - x)^2 \right] \]

\[ = \sum_{i=1}^{h} E(x_i^2) - h E(x^2) \]

\[ = h \left[ E(x_i^2) - E(x^2) \right] \]

\[ = h \left[ \text{Var}(x_i) + (E(x_i))^2 - \text{Var}(x) - (E(x))^2 \right] \]

\[ = h \left[ \sigma^2 + \mu^2 - \frac{\sigma^2}{h} - \mu^2 \right] = (h-1) \sigma^2 \]

\[ \text{(63)} \]

\[ E(S^2) = \sigma^2 \]
For this reason $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$ is called Sample Variance and not $S^2 = \frac{1}{h} \sum_{i=1}^{n} (x_i - \bar{x})^2$. Note that

$$E(S^2) = E \left( \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right) = \frac{h}{n} \sigma^2 \xi \cdot \sigma^2,$$

i.e. $S^2$ is unbiased with $\sigma^2$.

### 5.3.5 Distributions Based on Sampling From Normal Distribution

**Definition 5.3.5.1**

For a positive integer $m$, a rv $X$ is said to have the Student $t$-distribution with $m$ degrees of freedom (written as $X \sim t_m$) if the p.d.f. of $X$ is given by

$$f(x|m) = \frac{\Gamma \left( \frac{m+1}{2} \right)}{\sqrt{m\pi} \Gamma \left( \frac{m}{2} \right)} \left( 1 + \frac{x^2}{m} \right)^{-\frac{m+1}{2}}, \quad -\infty < x < \infty$$

(b) For positive integers $m_1$ and $m_2$, a rv $X$ is said to have the Snedecor $F$-distribution with $(m_1, m_2)$ degrees of freedom (written as $X \sim F_{m_1, m_2}$) if the p.d.f. of $X$ is given by

$$f(x|m_1, m_2) = \left\{ \begin{array}{ll}
\frac{m_1}{m_2} \frac{m_1 x^{m_1/2}}{B(m_1/2, m_2/2)} & \quad (1 + \frac{m_1 x}{m_2})^{-\frac{m_1+1}{2}} , \quad 0 < x < \infty \\
0 & \quad \text{otherwise}
\end{array} \right.$$  

Remark: (a) $X \sim t_m \Rightarrow f(x|m) = f(-x|m)$, for $x > 0$

(b) $X \sim t_m \Rightarrow$ if $f(x|m)$ $\uparrow$ in $(-\infty, 0)$, $f(0) = \frac{1}{2}$

(c) $X \sim t_m \Rightarrow$ if $f(x|m)$ $\downarrow$ in $(0, \infty)$, $f(0) = \frac{1}{2}$

[6.4.5]
(c1) The distribution is nothing but Cauchy distribution with density

\[ f(x|\alpha) = \frac{1}{\pi \alpha} \left( 1 + \frac{x^2}{\alpha^2} \right)^{-1/2}, \quad -\infty < x < \infty \]

and

\[ \text{E}(x) \text{ does not exist.} \]

(c1) \( X \sim \text{Fr}_m \Rightarrow f(x|\alpha, \beta) = \begin{cases} \frac{m_1 x}{\beta} \left( \frac{m_1 x}{\beta + m_1 x} \right)^{m_1 - 1} \left( \frac{1 - \frac{m_1 x}{\beta}}{1 + \frac{m_1 x}{\beta}} \right)^{\frac{m_1}{2}} \left( \frac{1 + \frac{m_1 x}{\beta}}{1 + \frac{m_1 x}{\beta}} \right)^{-\frac{m_1}{2}} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases} \]

\[ y = \frac{m_1 x}{\alpha} \sim \text{Be} \left( \frac{m_1}{2}, \frac{\alpha}{2} \right) \]

**Theorem (a).** Let \( Z \sim \text{N}(0,1) \) and let \( Y \sim \chi^2 \), \( m \geq 1 \). Let \( Y = X_1 + \ldots + X_m \). Then

\[ T = \frac{Z}{\sqrt{Y/m}} \sim \text{t}_m. \]

(b) For positive integers \( m_1 \) and \( m_2 \), let \( X_1 \sim \chi^2_{m_1} \) and \( X_2 \sim \chi^2_{m_2} \) be independent \( \text{RVs} \). Then

\[ U = \frac{X_1/m_1}{X_2/m_2} \sim \text{Fr}_{m_1/m_2}. \]

(c1) Let \( X \sim \text{t}_m \). Then \( \text{E}(x^r) \) is finite if \( r \neq m \).

\[ \text{For } r \geq 3, \ldots, \text{m}-1 \quad (m \text{ is even}) \]

\[ \text{For } r \geq 3, \ldots, \text{m}-1 \quad (m \text{ is odd}) \]

\[ E(x^r) = \begin{cases} 0 & r > m \\ \frac{m}{m - r} \left( \frac{m - r}{2} \right)^{\frac{m-r}{2}} \left( \frac{2}{r} \right)^{\frac{m-r}{2}} \left( \frac{r}{m} \right)^{\frac{m-r}{2}} & r \text{ is odd} \\ \frac{m}{m - r} \left( \frac{m - r}{2} \right)^{\frac{m-r}{2}} \left( \frac{2}{r} \right)^{\frac{m-r}{2}} \left( \frac{r}{m} \right)^{\frac{m-r}{2}} & r \text{ is even} \end{cases} \]

\[ \text{(65/5)} \]
(a) If \( X \sim \text{univ} \) then
\[
\begin{align*}
\text{Mean} &= \mu_1 = E(X) = 0, \quad m_2 = 2 \times 3 \\
\text{Variance} &= \mu_2 = E((X - \mu_1)^2) = \frac{m_2}{m_2 - 2}, \quad m_2 > 2 \times 4 \\
\text{Coefficient of skewness} &= \beta_1 = \frac{m_3}{m_2^{3/2}} = 0, \quad m_2 > 5 \times 5 \\
\beta_2 &= \text{Kurtosis} = \frac{3\mu_2 - 6}{\mu_2 - 4}, \quad m_2 > 5 \times 5 \\
\end{align*}
\]

(c2) Let \( m_1, m_2 \) and \( y \) be positive integers, and let \( X \sim \text{Fm}_1 \).
Then, for \( y \in \{ 2, 3, \ldots, m_2 \} \) and \( y > m_1 \), \( E(x^y) \) is not finite. For \( y \in \{ 2m_1, 2m_2, 2m_3, \ldots \} \), \( y > m_1 \),
\[
E(x^y) = \left( \frac{m_1^y}{m_2} \right) \prod_{i=1}^y \left( \frac{m_1 + i - 1}{m_2 - i} \right).
\]

(c3) If \( X \sim \text{Fm}_1 \), then
\[
\begin{align*}
\text{Mean} &= \mu_1 = E(X) = \frac{m_2}{m_2 - 2}, \quad m_2 \in \{ 2, 3, \ldots, y \} \\
\text{Variance} &= \mu_2 = E((X - \mu_1)^2) = \frac{2m_2^2 (m_2 - 2)}{m_1 (m_2 - 2)^2 (m_2 - 4)}, \quad m_2 \in \{ 5, 6, \ldots \} \\
\text{Coefficient of skewness} &= \beta_1 = \frac{m_3}{m_2^{3/2}} = \frac{2 (2m_1 m_2 - 2)}{m_2 - 6}, \quad m_2 \in \{ 7, 8, \ldots \} \\
\text{Kurtosis} &= \beta_2 = \frac{12 \left[ (m_2 - 2)^2 (m_2 - 4) + m_1 (m_2 - 2) (m_2 - 6) \right]}{m_1 (m_2 - 6) (m_2 - 8) (m_2 - 10)}
\end{align*}
\]

Proof (c1) The joint pdf of \((Y, Z)\) is given by
\[
\begin{align*}
&y \leq z : h_1(1/2, 1/2) = \begin{cases} \\
\frac{1}{\sqrt{2\pi} \sqrt{\pi}} e^{-yz^2/2} & y > 0, z > 0 \\
0 & \text{otherwise}
\end{cases}
\end{align*}
\]

Let \( U = \sqrt{\frac{Y}{Z}} \)
\[
S_{yz} = (0, 0) \times IR. \quad \text{Let} \quad h_1(1, 1) : [0, q] \times IR \to IR^2 \quad \text{where}
\]
\[
h_1(1, 1) = \begin{cases} \\
\frac{3}{16} & y \geq 0 \\
\frac{\sqrt{2\pi} \sqrt{\pi}}{2\pi} e^{-yz^2/2} & y < 0, z > 0 \\
0 & \text{otherwise}
\end{cases}
\]

The transformation \( h : S_{yz} \to IR \) is 1-1 with unique transformant.
\[ f^t = (h^t, a^t) \quad \text{where} \]
\[ h^t(u) = \ln u, \quad a^t(t, u) = t u \]

\[ J = \left| \begin{array}{cc} 0 & 2m u \\ u & t \end{array} \right| = -2m u t \]

\[ b^t(v) = \{ (t, u) : m u^2 > 0, -a c \leq u c \leq a \} \]
\[ = \{ (t, u) : b c > 0, t \in [x] \} = [12 \times 12] \]

The pdf of \((T, U)\) is given by

\[ b_T(t, u) = \frac{1}{2m} e^{\frac{-(h^t(t, u) - a^t(t, u))}{2}} \int_{12 \times 12} \frac{e^{-(h^t(t, u) - a^t(t, u))}}{2} \, dt \]

\[ = \frac{m}{2 \pi e} \int_{12 \times 12} e^{-(h^t(t, u) - a^t(t, u))} \, dt \]

The marginal pdf of \(U\) is

\[ b_U(u) = \int b_T(t, u) \, dt \]

\[ = \frac{m}{2 \pi e} \int_{12 \times 12} e^{-(h^t(t, u) - a^t(t, u))} \, dt \]

\[ = \frac{1}{\sqrt{m \pi} \sqrt{m \pi}} \int_{12 \times 12} e^{-(h^t(t, u) - a^t(t, u))} \, dt \]

\[ = \frac{1}{\sqrt{m \pi} \sqrt{m \pi}} \int_{12 \times 12} e^{-(h^t(t, u) - a^t(t, u))} \, dt \]

\[ \rightarrow \text{pdf of } U \]

\[ 67/5 \]
(5) The joint p.d.f. of \( x = (x_1, x_2) \) is given by

\[
\begin{align*}
  f_{X_1, X_2}(x_1, x_2) &= f_{X_1}(x_1) f_{X_2}(x_2) \\
  &= \frac{1}{2^{n_1/2} \Gamma(n_{1/2})} e^{-\frac{x_1^2}{2}} \frac{n_{1/2}}{x_1^{n_{1/2}-1}} I_{0}(\frac{1}{2}x_1x_2) \\
  &= \frac{1}{2^{n_2/2} \Gamma(n_{2/2})} e^{-\frac{x_2^2}{2}} \frac{n_{2/2}}{x_2^{n_{2/2}-1}} I_{0}(\frac{1}{2}x_1x_2)
\end{align*}
\]

Let \( V = \frac{x_1}{n_{1/2}} \) and \( S = (0, 0) \times (0, n_{2}) \), Consider the transformation \( h: \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) defined by

\[
  h_1(x, y) = \frac{2y}{2y/n_{2}} \quad \text{and} \quad h_2(x, y) = \frac{2x}{n_{1/2}}
\]

No that \( U = h_1(x, y) \) and \( V = h_2(x, y) \).

The transformation \( h: \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) with inverse transformation \( h^{-1} = (h_1^{-1}, h_2^{-1}) \) where

\[
  h_1^{-1}(y, u) = n_{1/2} u : h_2^{-1}(y, u) = n_{2/2} u
\]

\[
  J = \begin{vmatrix} h_{1u} & h_{1v} \\ h_{2u} & h_{2v} \end{vmatrix} = n_{1/2} n_{2/2}
\]

\[ h(S) = \{ (y, u) : n_{1/2} u > 0, n_{2/2} u > 0 \} = (0, n_{2/2}) \times (0, n_{1/2}) \]

Thus the joint p.d.f. of \( U, V \) is

\[
  f_{U, V}(y, u) = b_{X_1, X_2}(h_1^{-1}(y, u), h_2^{-1}(y, u)) |J| = f_{X_1, X_2}(h_1^{-1}(y, u), h_2^{-1}(y, u)) \frac{1}{2^{n_1/2} \Gamma(n_{1/2})} e^{-\frac{y^2}{2}} \frac{1}{2^{n_2/2} \Gamma(n_{2/2})} e^{-\frac{u^2}{2}} I_{0}(\frac{1}{2}y u) I_{0}(\frac{1}{2}uy)
\]

The p.d.f. of \( U \) is given by

\[
  f_U(u) = \int_0^{n_{2/2}} f_{U, V}(y, u) \, dy
\]

\[
  = \frac{1}{2^{n_{1/2} + n_{2/2}} \Gamma(n_{1/2}) \Gamma(n_{2/2})} \int_0^{n_{2/2}} e^{-\frac{u^2}{2} - \frac{1}{2} y u} I_{0}(\frac{1}{2}u y) \, dy
\]

\[ \int_0^{n_{2/2}} e^{-\frac{1}{2} y u} I_{0}(\frac{1}{2}u y) \, dy \]

\[ \boxed{68/5} \]
\[
\frac{\sqrt{\frac{m}{2}}}{{\frac{m+1}{2}}} \quad \frac{\sqrt{\frac{m}{2}}}{{\frac{m+1}{2}}}
\]

\[
\frac{(\frac{m}{2} - 1)}{(1 + \frac{m}{2} n)}\left(\frac{m}{2}\right)
\]

\[
\text{P}(E_{\text{neg}})
\]

- [a] \text{ d.f.} \text{ of } F_{m, n}

(c) \text{ Fix } m \geq 2, \ldots 1, \text{ Then }

\[
X \sim \frac{Z}{\sqrt{V}}
\]

where \( Z \sim N(0, 1) \) and \( V \sim \chi^2 \) are independent.

\[
E(X^2) = E\left(\left(\frac{Z}{\sqrt{V}}\right)^2\right) = \frac{E(Z^2)}{\text{var}(V)} = E(Z^2) \frac{1}{\text{var}(V)}
\]

\[
E(Z^2) = \int_{0}^{\infty} y \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1
\]

\[
E(\chi^2) = \frac{1}{\text{var}(V)} \int_{0}^{\infty} y \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{\text{var}(V)}
\]

For \( r < m \)

\[
E(\chi^2) = \frac{2^{\frac{m-r}{2}} \Gamma\left(\frac{m-r}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} = \frac{\Gamma\left(\frac{m-r}{2}\right)}{\Gamma\left(\frac{m}{2}\right)}
\]

(i) \text{ E(x)} = \left\{
\begin{array}{ll}
0, & \text{if } r \text{ is even and } r < m \\
\frac{\sqrt{LX}}{2^{\frac{m-r}{2}} \Gamma\left(\frac{m-r}{2}\right)} & \text{if } r \text{ is odd and } r < m
\end{array}
\right.

(A) \text{ Exercise}

(e) \text{ Fix } m_1, m_2 \in \mathbb{N}. \text{ Then }

\[
X \overset{d}{=} \frac{X_1/m_1}{X_2/m_2} = \frac{m_2}{m_1} \frac{X_1}{X_2}
\]

where \( X_1 \sim \chi^2_{m_1} \) and \( X_2 \sim \chi^2_{m_2} \) are independent.

69/5
For $v \in \infty$

\[
E(x^v) = \left( \frac{n_1}{n_1} \right) \cdot E\left( \frac{x_1^v}{X^2} \right) = \left( \frac{n_1}{n_1} \right) \cdot E(x_1^v) \cdot E\left( \frac{1}{X^2} \right)
\]

\[
E(x^v) = \frac{1}{2\pi[I_1]} \int_0^\infty x^{\frac{v+n_1-1}{2}} e^{-\frac{x}{2}} dx
\]

\[
= \frac{2^{\frac{n_1}{2}} \Gamma\left( \frac{n_1}{2} \right)}{2^{\frac{v}{2}} \Gamma\left( \frac{v}{2} \right)} = \frac{\Gamma\left( \frac{n_1}{2} + v \right)}{\Gamma\left( \frac{v}{2} \right)} \cdot \text{if } v > 2
\]

\[
E\left( \frac{1}{X^2} \right) = \begin{cases} 
\frac{2^{\frac{n_1}{2}} \Gamma\left( \frac{n_1}{2} \right)}{2^{\frac{v}{2}} \Gamma\left( \frac{v}{2} \right)}, & \text{if } v > 2 \\
0, & \text{if } v \leq 2
\end{cases}
\]

\[
= \begin{cases} 
\prod_{j=1}^{n_1} (n_1-2x), & \text{if } n_1 > 2 \\
0, & \text{if } n_1 \leq 2
\end{cases}
\]

\[
\Rightarrow E(x^v) = \begin{cases} 
\left( \frac{n_1}{n_1} \right) \cdot \prod_{j=1}^{n_1} \left( \frac{n_1 + 2(n_1-1)}{n_1-2x} \right), & n_1 > 2 \\
0, & \text{if } n_1 \leq 2
\end{cases}
\]

\text{(6) Esterline}

\textbf{Corollary 5.3.1.1}

Let $x_1, \ldots, x_n \ (n > 2)$ be a random sample from \( N(\mu, \sigma^2) \) distribution, where $\mu \in \mathbb{R}$ and $\sigma > 0$. Let $\bar{x} = \frac{1}{n} \sum x_i$ denote the sample mean and

\[ S = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \] denote the sample variance respectively. Then

\[ \sqrt{n} \left( \bar{x} - \mu \right) \sim t_{n-1} \]
Proof. We know that
\[ \bar{X} \sim N \left( \mu, \frac{\sigma^2}{n} \right) \]
\[ \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \]
\[ \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim t_{n-1} \] \(\text{independent} \)
\[ \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \]
\[ \sqrt{\frac{(n-1)s^2}{\sigma^2}} \sim \chi_{n-1} \] \(\text{independent} \)
\[ t = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim t_{n-1} \]
\[ \left( \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \right)^2 \sim \chi^2_{n-1} \]
\[ \left( \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \right)^2 \sim \chi^2_{n-1} \]

Corollary: Let \( X_1, \ldots, X_n \) (\( n \geq 2 \)) and \( Y_1, \ldots, Y_m \) (\( m \geq 2 \)) be independent \( (\text{i.e., } X_1, \ldots, X_n \text{ and } Y_1, \ldots, Y_m \text{ are independent}) \) random samples from \( N(\mu, \sigma^2) \) and \( N(\nu, \tau^2) \) distributions, respectively. Where \( \mu, \nu \in \mathbb{R} \), \( \sigma^2, \tau^2 > 0 \). Let
\[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{m} \sum_{i=1}^{m} Y_i, \quad S_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \text{and} \]
\[ S_2^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2 \]

(a) \[ \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0, 1) \]
(b) \[ \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim t_{\text{num}-1} \]
\[ \sqrt{\frac{(n-1)s^2}{\sigma_1^2} + \frac{(m-1)s^2}{\sigma_2^2}} \]
(c) \[ \frac{\sigma^2_1 s_2^2}{\sigma_1^2 s_2^2} \sim \chi^2_{n-1} \]
Proof. \( \overline{X} \sim N(\mu_1, \frac{1}{w} \sigma^2) \), \( T \sim N(\mu_2, \frac{w}{1} \sigma^2) \), and \( \frac{(w-1) \frac{S_1^2}{\sigma^2}}{\sum} \) are independent r.v.s. Thus:

\[ \overline{X} - T \sim N(\mu_1 - \mu_2, \frac{\sigma^2}{w} + \frac{w}{1} \sigma^2) \]

\[ \frac{(w-1) \frac{S_1^2}{\sigma^2}}{\sum} + \frac{(w-1) \frac{S_2^2}{\sigma^2}}{\sum} \sim \chi^2_{w-1} \]

\[ \overline{X} - \frac{T}{\sqrt{\frac{\sigma^2}{w} + \frac{w}{1} \sigma^2}} \sim N(0, 1) \]

and

\[ \frac{(\overline{X} - T - (\mu_1 - \mu_2))}{\sqrt{\frac{\sigma^2}{w} + \frac{w}{1} \sigma^2}} \sim \n + \chi^2_{w-1} \]

Remark:
(a) \( X \sim \chi^2 \)

\[ X \equiv \frac{N(0, 1)}{\sqrt{\chi^2_{w-1}}} \sim \chi^2_{w-1} \]

\[ X^2 \equiv \frac{(N(0, 1))^2}{\chi^2_{w-1}} \sim \chi^2_{w-1} \]

\[ \frac{X^2}{\chi^2_{w-1}} \sim \chi^2_{w-1} \]

\[ \frac{X^2}{\chi^2_{w-1}} \equiv \chi_{w-1} \]

Thus, \( X \sim \chi^2 \Rightarrow X^2 \sim \chi_{w-1} \)

\[ X \sim \chi^2 \Rightarrow X^2 \sim \chi_{w-1} \]
(b) \( X \sim F_{n_1} \Rightarrow X = \frac{X_{n_1}/n_1}{X_{n_2}/n_2} \sim \text{independent} \)

\[
\frac{1}{X} \sim \frac{X_{n_2}/n_2}{X_{n_1}/n_1} \sim \text{independent}
\]

\( T = \frac{1}{X} \sim F_{n_2/n_1} \)

Thus \( X \sim F_{n_1} \Rightarrow \frac{1}{X} \sim F_{n_2/n_1} \).

(c) \( X \sim t_m \Rightarrow \text{Kurtosis } = \frac{2(m+1)}{m-2} \frac{m-4}{m-6} \)

\( t_m \) distribution \( (m \geq 4) \) is symmetric and leptokurtic (i.e., it has sharper peak and longer fatter tails compared to \( N(0,1) \) distribution)

\[ \text{Figure: Path of } t_m \text{ and } N(0,1) \text{ distributions.} \]

An \( m \rightarrow \infty \Rightarrow t_m \rightarrow \infty \). The results that for large \( m \), \( t_m \) distribution behaves like \( N(0,1) \) distribution.

(d) For various values of \( m \) and \( \alpha \in (0,1) \), the d.f. of \( t_m \) is tabulated in various text books.

(e) For fixed \( m_1 \), \( m_2 \), and \( \alpha \in (0,1) \) let \( b_{m_1/m_2} \) be the \((\alpha)\)-th quantile of \( X \sim F_{m_1/m_2} \). Then

\[
P(X \leq \frac{1}{X} \leq \frac{1}{b_{m_1/m_2}}) = \alpha
\]

\[
\Rightarrow b_{m_1/m_2} = \frac{1}{b_{m_2/m_1}} \quad (\text{Note: } \frac{1}{X} \sim F_{m_2/m_1})
\]

\[75/5\]
Example 5.3.1

Let \( x_1, \ldots, x_n \) be a random sample from \( N(\mu, \sigma^2) \)
distribution, where \( \mu \in \mathbb{R}, \sigma > 0 \) and \( n \in \mathbb{N} \). Let \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \)
and \( S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \) be the sample mean and sample variance, respectively. Evaluate \( \mathbb{E}(\frac{\bar{x}}{S}) \), for \( n \in \mathbb{N} \).

Solution

\[
\bar{x} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{independent} \\
S^2 \sim \chi_{n-1}^2(\frac{\sigma^2}{\sigma^2}) \sim \chi^2_{n-1}
\]

\[
\mathbb{E}(\frac{\bar{x}}{S}) = \sqrt{\frac{n-1}{n}} \mathbb{E}(\bar{x} \cdot \frac{1}{S^2})
\]

\[
= \sqrt{\frac{n-1}{n}} \mathbb{E}(\bar{x}) \mathbb{E}(\frac{1}{S^2}) \quad \text{(independence)}
\]

\[
= \sqrt{\frac{n-1}{n}} \mu \int_0^\infty \frac{e^{-\frac{y}{2} \left( \frac{2}{\sigma^2} \right)}}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)} \, dy
\]

\[
= \sqrt{\frac{n-1}{2}} \frac{\mu}{\sqrt{\frac{n-1}{2}}}
\]

Example 5.3.2

Let \( z_1, \ldots, z_n \) be i.i.d. \( \text{N}(0,1) \) r.v.'s and let \( a_1, b_1, \ldots, a_n, b_n \in \mathbb{R} \) be such that \( \sum_{i=1}^{n} a_i^2 > 0 \) and \( \sum_{i=1}^{n} b_i^2 > 0 \) and \( \sum_{i=1}^{n} a_i b_i = 0 \). Show that

(a) \( y_1 = \frac{\sum_{i=1}^{n} b_i^2}{\sum_{i=1}^{n} a_i^2} \) \( \sim \chi_1^2 \)

(b) \( y_2 = \left( \frac{\sum_{i=1}^{n} a_i z_i}{\sum_{i=1}^{n} b_i z_i} \right)^T \) \( \sim \chi_1 
\)
\[ T_3 = \sqrt{\frac{\sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} b_i^2}} \]

Solution

\[ c_1 \sum_{i=1}^{n} a_i z_i + c_2 \sum_{i=1}^{n} b_i z_i \] 

a linear combination of \( z_1 \ldots z_n \) 

\( \rightarrow \) univariate normal distribution

\[ \Rightarrow \left( \sum_{i=1}^{n} a_i z_i, \sum_{i=1}^{n} b_i z_i \right) \sim N_2 \]

\[ E(\sum_{i=1}^{n} a_i z_i) = 0; \quad \text{Var}(\sum_{i=1}^{n} a_i z_i) = \sum_{i=1}^{n} a_i^2 \]

\[ E(\sum_{i=1}^{n} b_i z_i) = 0; \quad \text{Var}(\sum_{i=1}^{n} b_i z_i) = \sum_{i=1}^{n} b_i^2 \]

\[ \text{Cov}(\sum_{i=1}^{n} a_i z_i, \sum_{i=1}^{n} b_i z_i) = \sum_{i=1}^{n} a_i b_i = 0 \]

\[ \Rightarrow \left( \sum_{i=1}^{n} a_i z_i, \sum_{i=1}^{n} b_i z_i \right) \sim N_2(0, 0; \sum_{i=1}^{n} a_i^2, 0) \]

\[ \Rightarrow \sum_{i=1}^{n} a_i z_i \sim N(0, \sum_{i=1}^{n} a_i^2) \quad \text{and} \quad \sum_{i=1}^{n} b_i z_i \sim N(0, \sum_{i=1}^{n} b_i^2) \]

\[ \Rightarrow \frac{\sum_{i=1}^{n} a_i z_i}{\sqrt{\sum_{i=1}^{n} a_i^2}} \sim N(0, 1) \quad \text{and} \quad \frac{\sum_{i=1}^{n} b_i z_i}{\sqrt{\sum_{i=1}^{n} b_i^2}} \sim N(0, 1) \] are independent

\[ \Rightarrow \frac{\sum_{i=1}^{n} a_i z_i}{\sqrt{\sum_{i=1}^{n} a_i^2}} \sim N(0, 1) \quad \text{and} \quad \frac{\sum_{i=1}^{n} b_i z_i}{\sqrt{\sum_{i=1}^{n} b_i^2}} \sim N(0, 1) \] are independent

\[ \Rightarrow \frac{\sum_{i=1}^{n} a_i z_i}{\sqrt{\sum_{i=1}^{n} a_i^2}} / \sqrt{\sum_{i=1}^{n} b_i^2} \sim t_1 \]

\[ \Rightarrow \frac{\sum_{i=1}^{n} a_i z_i / \sqrt{\sum_{i=1}^{n} a_i^2}}{\sqrt{\sum_{i=1}^{n} b_i^2}} \sim t_1 \]

\[ \Rightarrow T_3 \sim t_1 \]

i.e., \( T_3 \sim t_1 \)

7575
(5) Since \( t_i \equiv F_i \), the result follows on using (9).

(c) \( F_{(n)}(\gamma) = \Phi(\gamma) \)
\[
= P\left( \frac{Z_1}{\sqrt{2}} \leq \gamma \right), \quad \text{where} \quad Z \sim N(0,1)
\]
and are independent.

(clearly)
\[
F_{(n)}(\gamma) = P\left( \frac{Z_1}{\sqrt{2}} \leq \gamma, \ Z_2 \geq 0 \right) + P\left( \frac{Z_1}{\sqrt{2}} \leq \gamma, \ Z_2 < 0 \right)
\]
\[
= P\left( \frac{Z_1}{\sqrt{2}} \leq \gamma, \ Z_2 \geq 0 \right) + P\left( \frac{Z_1}{\sqrt{2}} \leq \gamma, \ Z_2 < 0 \right)
\]
\[
= P\left( \frac{Z_1}{\sqrt{2}} \leq \gamma, \ \gamma < 0 \right)
\]
\[
\Rightarrow \gamma \sim \frac{Z_1}{\sqrt{2}} \sim t_{(n)} \quad \text{by (9.1)}
\]
\[
\Rightarrow \gamma \sim t_{(n)}
\]

5.4. Special Multivariate Distributions

5.4.1. Multinomial Distribution

(A generalization of binomial distribution)

\( E \): a random experiment whose each trial results in
one (and only one) of \( p \) possible outcomes \( E_1, \ldots, E_p \).

\( E_i \), \( \text{for } i = 1, \ldots, p \) and \( \sum_{i=1}^{p} E_i = n \). Let

\[ P(A_i) = \theta_i \in (0,1) \quad (i = 1, \ldots, p) \]

and \( \sum_{i=1}^{p} \theta_i < 1 \), no \( \theta_i = 0 \).

\[ P(E_{(1)}) = 1 - \sum_{i=1}^{p} \theta_i \in (0,1) \]

Consider \( n \) independent trials of \( E \).

Define

\[ X_i = \# \text{ of times } E_i \text{ occurs in } n \text{ trials} \quad (i = 1, \ldots, p) \]

Then

\[\sum_{i=1}^{p} X_i = n, \quad \text{i.e. } X_{(1)} = n - \sum_{i=1}^{p} X_i \quad \text{One way} \]

\[\sum_{i=1}^{p} X_i = n \]
be interested in probability distribution of \( X = (x_1, \ldots, x_p) \).

We have

\[
S_x = \left\{ (x_1, \ldots, x_p) : 2 \leq \sum_{i=1}^{p} x_i \leq n \right\}
\]

\[
\begin{align*}
\frac{1}{S_x} (x_1, \ldots, x_p) &= \frac{1}{\binom{n}{x_1, \ldots, x_p}} \frac{\prod_{i=1}^{p} \binom{n}{x_i}}{\prod_{i=1}^{p} (n-x_i)} \theta_1^{x_1} \cdots \theta_p^{x_p} (1-\sum \theta_i)^{n-\sum x_i}, \\
&= \begin{cases} 
\frac{\prod_{i=1}^{p} \binom{n}{x_i}}{\prod_{i=1}^{p} (n-x_i)} \theta_1^{x_1} \cdots \theta_p^{x_p} (1-\sum \theta_i)^{n-\sum x_i}, & 2 \leq S_x \\
0 & \text{otherwise}
\end{cases}
\end{align*}
\]

\[
\rightarrow \text{Multinomial distribution with } n \text{ trials and cell probabilities } \theta_1, \ldots, \theta_p \text{ denoted by } \text{Mult}(n, \theta_1, \ldots, \theta_p).
\]

\[
\rightarrow \text{a family of distributions with varying } n \in \mathbb{N} \text{ and } \theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p \] \[
= \{(x_1, \ldots, x_p) : \theta_1 \geq 0, \ldots, \theta_p \geq 0, \theta_1 + \cdots + \theta_p = 1\}.
\]

Remark for \( p = 1 \): \text{Mult}(n, \theta_1) \text{ distribution is the same as } \text{Bin}(n, \theta_1).

Theorem 5.4.1: Suppose that \( X = (x_1, \ldots, x_p) \in \text{Mult}(n, \theta_1, \ldots, \theta_p) \), where \( n \in \mathbb{N} \) and \( \theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p \). Then

(a) \( X_i \sim \text{Bin}(n, \theta_i) \), \( \theta_i = \cdots = \theta \)

(b) \( X_i + x_j \sim \text{Bin}(n, \theta_i + \theta_j) \), \( \theta_i = \cdots = \theta \)

(c) \( \text{E}(X_i) = n\theta_i \) and \( \text{Var}(X_i) = n\theta_i(1-\theta_i), \theta_i = \cdots = \theta \)

(d) \( \text{Cov}(X_i, X_j) = -n\theta_i\theta_j, \theta_i = \cdots = \theta \)

Proof: (a) Fix \( \{i \} \in \{1, \ldots, p\} \). A given trial of the experiment treat the occurrence of \( E_i \) as success and its non-occurrence (i.e. occurrence of any other \( E_j, j \neq i \)) as failure. Then we have a sequence of independent
Dennoulli trials with probability of success in each trial as \( P(E_i) = \theta_i \). Thus

\[
X_i = \# \text{ of times } E_i \text{ occur in } n \text{ Denonulli trials} \\
\sim \text{Bin}(n, \theta_i) \quad (i \geq 1).
\]

(b) For \( i \neq j \) etc. In any given trial of \( E \) consider occurrence of \( E_i \) on \( E_j \) as success and occurrence of \( E_i \) or \( E_j \) as failure. Then we have a sequence of independent Bernoulli trials with success probability \( P(E_i \cap E_j) = \theta_i + \theta_j \)

\[
X_i + X_j = \# \text{ of successes in } n \text{ Bernoulli trials} \\
\sim \text{Bin}(n, \theta_i + \theta_j).
\]

(c) Obvious

(d) \[
\begin{align*}
\text{Var} (X_i + X_j) &= n(\theta_i + \theta_j) (1-\theta_i - \theta_j) \\
&= \text{Var}(X_i) + \text{Var}(X_j) + 2 \text{Cov}(X_i, X_j) = n\theta_i \theta_j (1-\theta_i - \theta_j) \\
&= n\theta_i \theta_j (1-\theta_i - \theta_j) + 2 \text{Cov}(X_i, X_j) = n\theta_i \theta_j (1-\theta_i - \theta_j)
\end{align*}
\]

\[\text{MAF} \quad \text{The m.s.b. of } Z = (X_1, \ldots, X_p) \text{ is given by}
\]

\[
\begin{align*}
\pi_Z(t_1, \ldots, t_p) &= E(e^{t_1 X_1 + \cdots + t_p X_p}) \\
&= \sum_{x_1 = 0}^{\infty} \cdots \sum_{x_p = 0}^{\infty} e^{t_1 x_1 + \cdots + t_p x_p} \prod_{j=1}^{p} \left( \frac{1}{1-2\theta_j} \right) \\
&= \sum_{x_1 = 0}^{\infty} \cdots \sum_{x_p = 0}^{\infty} \prod_{j=1}^{p} \left( \frac{1}{1-2\theta_j} \right) \\
&= \left( \frac{1}{1-2\theta_1} \right) \cdots \left( \frac{1}{1-2\theta_p} \right) \left( \frac{1}{1-2\theta_1} \right) \\
&= (\Theta e^{t_1} + \cdots + \Theta e^{t_p} + 1 - \frac{1}{\Theta})^n = \text{m.s.b.}
\end{align*}
\]

Remark 54.6.2

The last theorem can also be proven with m.s.b.

For example, for \( i, j \in \{1, \ldots, p\} \) \( i \neq j \)

\[
\pi_{X_i + X_j}(t_1, \ldots, t_p) = \left( (\theta_i \theta_j)^{t_1} + 1 - \theta_i - \theta_j \right) \\
\sim \text{m.s.b. of } \text{Bin}(n, \theta_i + \theta_j)
\]

178/5
5.4.2. Bivariate Normal Distribution

**Definition 5.4.2.1** A bivariate set \(Z = (X_1, X_2)\) is said to follow a bivariate normal distribution \(N_2(\mu, \Sigma)\) if,

for some \(-\infty < \mu < \infty\), \(\sigma > 2\) and \(-\infty < \mu < \infty\), the joint p.d.f. of \(Z = (X_1, X_2)\) is given by

\[
d x_1 x_2 = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[ \left( \frac{X_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{X_1 - \mu_1}{\sigma_1} \right) \left( \frac{X_2 - \mu_2}{\sigma_2} \right) + \left( \frac{X_2 - \mu_2}{\sigma_2} \right)^2 \right]}
\]

\(\sigma_1 \leq \sigma_2, \quad \sigma_2 \geq \sigma_1 > 0, \quad \rho = \rho(\Sigma)\).

(Clearly) \(d x_1 x_2 \geq 0\) for \(x_1, x_2 \in \mathbb{R}^2\) and on unskewing the transformation

\[
\int x_1 x_2 (d x_1 d x_2) = \frac{1}{2\pi \gamma_2 \sqrt{1-\rho^2}} \int d x_1 \int x_2 e^{-\frac{1}{2(1-\rho^2)} \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2} d x_1 d x_2
\]

\[
= \frac{1}{2\pi \gamma_2 \sqrt{1-\rho^2}} \int e^{-\frac{1}{2(1-\rho^2)} \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2} d x_2 = 1
\]

\[
\Rightarrow d x_1 x_2 (x_1, x_2) \text{ is a p.d.f.}
\]

Note that for \(x = (x_1, x_2) \in \mathbb{R}^2\),

\[
dx_1 x_2 (x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2}
\]

\[
= \left\{ \begin{array}{c}
\frac{1}{\sqrt{2\pi} \sigma_1 \sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_1^2 (1-\rho^2)} \left( 21 - \left( \frac{\mu_1 - \mu_2}{\sigma_1} \right) \left( 21 - \mu_2 \right) \right)^2} \\
\frac{1}{\sqrt{2\pi} \sigma_2} e^{-\frac{1}{2\sigma_2^2} \left( 3 - \mu_2 \right)^2}
\end{array} \right\} d x_1 d x_2
\]

\[
= 1 \times 1 (d x_1 d x_2) \frac{d x_2 (x_2)}{d x_2 (x_2)}
\]

\[79/5\]
Let $x_1$ be the outcome of an event and $x_2$ be another event. We assume:

- $X_1 \sim N\left(\mu_1, \sigma_2^2\right)$
- $X_2 \sim N\left(\mu_2, \sigma_2^2\right)$
- $\mathbb{E}(X_1) = \mathbb{E}(X_2) = 0$
- $\sigma^2 = \text{Var}(X_1)$ and $\sigma^2 = \text{Var}(X_2)$

Then, for any $t \in \mathbb{R}$:

$$
\mathbb{P}(X_1 X_2) = E(e^{tX_1 + tX_2}) = E(E(e^{tX_1 + tX_2} | X_2)) = E(e^{tx_1 + tx_2} E(e^{tX_1} | X_2)) = E(e^{tx_1 + tx_2} e^{t \mathbb{E}(X_1) + t \mathbb{E}(X_2)})
$$

$$
= e^{t \mathbb{E}(X_1) + t \mathbb{E}(X_2)} \mathbb{E}(e^{tX_1 + tX_2})
$$

$$
= e^{t \mathbb{E}(X_1) + t \mathbb{E}(X_2)} \mathbb{E}(e^{tX_1 + tX_2})
$$

$$
= e^{t \mathbb{E}(X_1) + t \mathbb{E}(X_2)} \mathbb{E}(e^{tX_1 + tX_2})
$$

Thus, we have the following theorem.

\[ \text{Conclusion} \]
Theorem 5.2.1

Suppose that \( X = (X_1, X_2) \) with \( \mu_1, \mu_2, \sigma_{11}^2, \sigma_{22}^2, \rho \),

\(-\infty < \mu_1 < \mu_2 < \infty,
\sigma_{11}^2, \sigma_{22}^2 > 0 \)
and \(-1 < \rho < 1\). Then

(a) \( X_1 \sim N(\mu_1, \sigma_{11}^2) \) and \( X_2 \sim N(\mu_2, \sigma_{22}^2) \).

(b) For fixed \( x_2 \in \mathbb{R} \)

\[ \chi_{x_2} \sim N\left( \frac{x_2 - \mu_2}{\sqrt{1 - \rho^2}}, \frac{\sigma_{11}^2}{\sqrt{1 - \rho^2}} \right) \]

and for fixed \( x_1 \in \mathbb{R} \)

\[ \chi_{x_1} \sim N\left( \frac{x_1 - \mu_1}{\sqrt{1 - \rho^2}}, \frac{\sigma_{22}^2}{\sqrt{1 - \rho^2}} \right), \]

(c) The pdf of \( X = (X_1, X_2) \) in

\[ f_{X_1 X_2}(x_1, x_2) = \frac{1}{2\pi \sigma_{11} \sigma_{22} \sqrt{1 - \rho^2}} \exp\left(-\frac{(x_1 - \mu_1)^2}{2\sigma_{11}^2} - \frac{(x_2 - \mu_2)^2}{2\sigma_{22}^2} - \frac{\rho(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_{11} \sigma_{22} \sqrt{1 - \rho^2}} \right) \]

\( \mu_1 < x_1 < \mu_2 \) and \( \mu_1 < x_2 < \mu_2 \),

\( \sigma_{11} < \frac{\sigma_{22}}{\sqrt{1 - \rho^2}} \),

\( \frac{\sqrt{\sigma_{11}^2 + \sigma_{22}^2 - 2\rho \sigma_{11} \sigma_{22} \sqrt{1 - \rho^2}}}{\sqrt{1 - \rho^2}} > \frac{\rho}{\sqrt{1 - \rho^2}} \)

(d) \( P(x_1, x_2) = \text{Cov}(X_1, X_2) = 1 \)

(e) \( X_1 \) and \( X_2 \) are independent if \( \rho = 0 \)

(f) For real constants \( c_1 \) and \( c_2 \) such that \( (c_1, c_2)^T \neq 0 \)

\[ c_1 X_1 + c_2 X_2 \sim N\left( c_1 \mu_1 + c_2 \mu_2, c_1^2 \sigma_{11}^2 + c_2^2 \sigma_{22}^2 + 2c_1 c_2 \sigma_{12} \right) \]

Solution

(a) - (c) All were done

(4) For \( t = (t_1, t_2) \in \mathbb{R}^2 \)

\[ \mathbb{E}(X_1 X_2 (t_1, t_2)) = \mathbb{E}(X_1 X_2) \cdot (t_1, t_2) = \mu_1 \mu_2 + \sigma_{12} \rho \]

\[ \frac{\partial}{\partial t_1} \mathbb{E}(X_1 X_2 (t_1, t_2)) = \mu_1 + 2 \sigma_{12} \rho \]

\[ \frac{\partial^2}{\partial t_1 \partial t_2} \mathbb{E}(X_1 X_2 (t_1, t_2)) = \rho \sigma_{12} \rho \]

\[ \text{Cov}(X_1, X_2) = \left[ \frac{\partial^2}{\partial t_1 \partial t_2} \mathbb{E}(X_1 X_2 (t_1, t_2)) \right]_{t_1 = 0, t_2 = 0} = \sigma_{12} \rho \]

\[ P(X_1, X_2) = \text{Cov}(X_1, X_2) \frac{1}{\sqrt{\text{Var}(X_1) \text{Var}(X_2)}} = \rho \]

Page 85
(e) Obviously if \( x_i \) and \( x_2 \) are independent then
\[
P = \text{Cov}(x_1, x_2) = 0.
\]
Now suppose that \( P \neq 0 \). Then
\[
x_1 \text{ and } x_2 \text{ cannot be independent.}
\]

(b) Let \( Y = C_1 X_1 + C_2 X_2 \). Then
\[
E(Y^2) = E\left( e^{Y^2} \right) = E\left( e^{(C_1^2 + C_2^2 + 2C_1C_2O_i^2)} \right)
\]
\[
\approx \exp \left( \frac{C_1^2 + C_2^2 + 2C_1C_2O_i^2}{2} \right)
\]

\[\Rightarrow \text{ M.G.L. of } N\left( C_1^2 + C_2^2, 2C_1C_2O_i^2 \right)\]
Conversely, suppose that for all \( t \geq (x, t) \in \mathbb{R}^d \times [0, T] \)

\[
\gamma = t x + \alpha t \in \mathbb{R}^{2d}, \quad \mu = \frac{1}{2} \sigma^2 + \frac{1}{2} \alpha^2 + \beta + t \alpha \in \mathbb{R}
\]

Then for \( t = (y, t) \in \mathbb{R}^d \times [0, T] \)

\[
\Pi_{x, x_0} (y, t) = E \left( e^{x t + \alpha t x_0} \right)
\]

\[
= \Pi_{x, x_0} (1)
\]

\[
= e^{\Pi_{x, x_0} (T)} + \frac{1}{2} \sigma^2 + \frac{1}{2} \alpha^2 + \beta + t \alpha \in \mathbb{R}
\]

\[\rightarrow\] \text{ m.s.b. of } \mathcal{N}^2 \left( m, \sigma^2, \alpha, \beta \right)

\[
\Rightarrow \quad T = (x, x_0) \sim \mathcal{N}^2 \left( m, \sigma^2, \alpha, \beta \right).
\]