IC105: Probability and Statistics | 2021-22-M

Lecture 0: Lecture Notes

Instructor: Dr. Kuldeep Kumar Kataria | *Scribe:*

These are the updated notes of Lecture #12 to Lecture #22

0.1. Some Special Discrete Distributions

0.1.1. 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 (S) in each trial as $p \in (0,1)$ (same for each trial); here $n \in \mathbb{N}$ is a fixed natural number.

Define $X =$ the number of success in n trials. Then $S_X = \{0, 1, 2, \dots, n\}$ and for $k \in S_X$

$$\begin{aligned}
P(X = k) &= P(\underbrace{SS\cdots SFF\cdots F}_{k \text{ successes and } n-k \text{ failures}}) + P(\underbrace{SFFS\cdots FFS}_{k \text{ successes and } n-k \text{ failures}}) + \cdots + P(\underbrace{FF\cdots FSS\cdots S}_{k \text{ successes and } n-k \text{ failures}}) \\
&\qquad\qquad\qquad\qquad \left(\text{total of } \binom{n}{k} \text{ terms}\right) \\
&= p^k(1-p)^{n-k} + p^k(1-p)^{n-k} + \cdots + p^k(1-p)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}, \quad (\text{independence of trials}).
\end{aligned}$$

Thus,

$$f_X(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, \dots, n, \\ 0, & \text{otherwise}, \end{cases}$$

is called Binomial distribution with n trials and success probability p (denoted by $Bin(n,p)$ and written as $X \sim Bin(n,p)$). $\{Bin(n,p) : n \in \mathbb{N}, p \in (0,1)\}$ is the family of probability distributions that has two parameters $n \in \mathbb{N}$ and $p \in (0,1)$.

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

Suppose that $X \sim Bin(n,p)$, $n \in \mathbb{N}, p \in (0,1)$. Then

$$\text{m.g.f. } M_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{\infty} \binom{n}{k} (pe^t)^k (1-p)^{n-k} = (1 - p + pe^t)^n, \quad t \in \mathbb{R}.$$

Let $q = 1 - p$, so that $M_X(t) = (q + pe^t)^n$, $t \in \mathbb{R}$. Then

$$
\begin{aligned}
M_X^{(1)}(t) &= n(q + pe^t)^{n-1}pe^t, \\
M_X^{(2)}(t) &= np(q + pe^t)^{n-1}e^t + n(n-1)(q + pe^t)^{n-2}(pe^t)^2, \\
E(X) &= M_X^{(1)}(0) = np, \;\; E(X^2) = M_X^{(2)}(0) = np + n(n-1)p^2, \;\; \mathrm{Var}(X) = np(1-p) = npq.
\end{aligned}
$$

Note that if $X \sim Bin(n, p)$ then Variance $<$ Mean. It can be seen that

$$
\begin{aligned}
\mu_3' &= E(X^3) = np(1 - 3p + 3np + 2p^2 - 3np^2 + n^2p^2), \\
\mu_4' &= E(X^4) = np(1 - 7p + 7np + 12p^2 - 18np^2 + 6n^2p^2 - 6p^3 + 11np^3 - 6n^2p^3 + n^3p^3), \\
\mu_3 &= E((X - \mu_1')^3) = np(1-p)(1-2p), \\
\mu_4 &= E((X - \mu_1')^4) = np(1-p)(3p^2(2-n) + 3p(n-2) + 1), \\
\beta_1 &= \frac{\mu_3}{\mu_2^{3/2}} = \frac{1-2p}{\sqrt{np(1-p)}} = \begin{cases} \text{symmetric for } p = \frac{1}{2} \;, \\ \text{positively skewed for } 0 < p < \frac{1}{2}, \\ \text{negatively skewed for } p > \frac{1}{2}, \end{cases} \\
\nu_2 &= \nu_1 - 3 = \frac{1 - 6pq}{npq}, \;\; \text{where } \nu_1 = \frac{\mu_4}{\mu_2^2}.
\end{aligned}
$$

Also, for $r \in \{1, 2, \dots\}$, let $X_{(r)} = X(X-1)(X-2)\cdots(X-r+1)$, the rth factorial moment is given by

$$
\begin{aligned}
E(X_{(r)}) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} k(k-1)(k-2)\cdots(k-r+1) \\
&= n(n-1)(n-2)\cdots(n-r+1) \sum_{k=r}^{n} \binom{n-r}{k-r} p^k (1-p)^{n-k} \\
&= n(n-1)(n-2)\cdots(n-r+1)p^r \sum_{k=0}^{n-r} \binom{n-r}{k} p^k (1-p)^{n-r-k} \\
&= n(n-1)(n-2)\cdots(n-r+1)p^r(1-p+p)^{n-r} = n(n-1)(n-2)\cdots(n-r+1)p^r.
\end{aligned}
$$

Theorem 0.1. *Let $X_1, X_2, \dots, X_k$ be independent r.v.'s with $X_i \sim Bin(n_i, p)$, $n_i \in \mathbb{N}$, $p \in (0,1)$, $i = 1, 2, \dots, k$. Then $Y = \sum_{i=1}^k X_i \sim Bin(n, p)$, where $n = \sum_{i=1}^k n_i$.*

Proof. For $t \in \mathbb{R}$,

$$
\begin{aligned}
M_Y(t) = E(e^{tY}) = E(e^{t\sum_{i=1}^k X_i}) = E\left(\prod_{i=1}^k e^{tX_i}\right) &= \prod_{i=1}^k E(e^{tX_i}), \quad \text{(independent of } X_i's) \\
&= \prod_{i=1}^k M_{X_i}(t) = \prod_{i=1}^k (1 - p + pe^t)^{n_i} = (1 - p + pe^t)^{\sum_{i=1}^k n_i} \\
&\to \text{m.g.f. of } Bin\left(\sum_{i=1}^k n_i, p\right).
\end{aligned}
$$

By uniqueness of m.g.f. $Y \sim Bin(n, p)$, where $n = \sum_{i=1}^k n_i$. □

Example 0.2. *Let $X \sim Bin(n, 1/2)$, then $X - \frac{1}{2} \stackrel{d}{=} \frac{n}{2} - X$, since $n - X \stackrel{d}{=} X$ (Exercise).*

Example 0.3. *A fair dice is rolled* 5 *times independently. Find the probability that on* 3 *occasions we get a six.*

Solution: Consider getting a six as success. Then $X =$ the number of success in 5 trials $\sim Bin(5, 1/6)$.

So, the required probability $= P(X = 3) = \binom{5}{3}(1/6)^3(5/6)^2$.

0.1.2. Negative Binomial Distribution

Consider a sequence of independent Bernoulli trials with probability of success in each trial as $p \in (0, 1)$. Let $r \in \{1, 2, \dots\}$ be a fixed positive integer. Let X denote the number of failures preceding the rth success. Then $S_X = \{0, 1, 2, \dots\}$ and for $k \in S_X$, we have

$$\begin{aligned}
f_X(k) &= P(X = k) \\
&= P(k \text{ failures precede } r\text{th success}) \\
&= P(r-1 \text{ successes in first } k+r-1 \text{ trials and success in } (k+r)\text{th trial}) \\
&= P(r-1 \text{ successes in first } k+r-1 \text{ trials}) \times P(\text{success in } (k+r)\text{th trial}), \quad (\text{independence of trials}) \\
&= \binom{k+r-1}{r-1} p^{r-1}(1-p)^k p = \binom{k+r-1}{r-1} p^r (1-p)^k.
\end{aligned}$$

Thus,

$$f_X(x) = P(X = x) = \begin{cases} \binom{x+r-1}{r-1} p^r (1-p)^x, & x = 0, 1, 2, \dots, \\ 0, & \text{otherwise.} \end{cases}$$

The distribution of X is called Negative binomial distribution with r success, and success probability $p \in (0, 1)$ (denoted by $NB(r, p)$ and written as $X \sim NB(r, p)$) (has two parameters $r \in \mathbb{N}$ and $p \in (0, 1)$). $\{NB(r, p) : r \in \mathbb{N},\ p \in (0, 1)\}$ is a family of probability distribution.

Remark 0.4. *For* $t \in (-1, 1)$*, we have*

$$\sum_{k=0}^{\infty} \binom{k+r-1}{r-1} t^k = 1 + rt + \frac{(r+1)r}{2!} t^2 + \frac{(r+2)(r+1)r}{3!} t^3 + \cdots = (1-t)^{-r}.$$

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

$$\begin{aligned}
M_X(t) = E(e^{tX}) &= \sum_{k=0}^{\infty} e^{tk} \binom{k+r-1}{r-1} (1-p)^k p^r \\
&= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} ((1-p)e^t)^k = \left(\frac{p}{1-(1-p)e^t} \right)^r, \quad t < -\ln(1-p).
\end{aligned}$$

Thus,

$$\begin{aligned}
\psi_X(t) &= \ln M_X(t) = r \ln p - r \ln(1 - qe^t),\ t < -\ln(1-p), \\
\psi_X^{(1)}(t) &= \frac{rqe^t}{1-qe^t} = r\left(\frac{1}{1-qe^t} - 1 \right), \quad t < -\ln(1-p), \\
\psi_X^{(2)}(t) &= \frac{rqe^t}{(1-qe^t)^2}, \quad t \in \mathbb{R},
\end{aligned}$$

$$E(X) = \psi_X^{(1)}(0) = \frac{rq}{p}, \quad \mathrm{Var}(X) = \psi_X^{(2)}(0) = \frac{rq}{p^2}, \quad \text{Variance} > \text{Mean}.$$

Also, for $m \in \{1, 2, \dots\}$, let $X_{(m)} = X(X-1)(X-2)\cdots(X-m+1)$. Then

$$\begin{aligned}
E(X_{(m)}) &= \sum_{k=0}^{\infty} k(k-1)(k-2)\cdots(k-m+1)\binom{k+r-1}{r-1}p^r(1-p)^k \\
&= p^r \sum_{k=m}^{\infty} k(k-1)(k-2)\cdots(k-m+1)\binom{k+r-1}{r-1}(1-p)^k \\
&= r(r+1)(r+2)\cdots(r+m-1)p^r \sum_{k=m}^{\infty} \frac{(k+r-1)!}{(k-m)!(r+m-1)!}(1-p)^k \\
&= r(r+1)(r+2)\cdots(r+m-1)p^r \sum_{k=0}^{\infty} \frac{(k+m+r-1)!}{k!(r+m-1)!}(1-p)^{k+m} \\
&= r(r+1)(r+2)\cdots(r+m-1)p^r q^m \sum_{k=0}^{\infty} \binom{k+m+r-1}{m+r-1} q^k \\
&= r(r+1)(r+2)\cdots(r+m-1)p^r q^m (1-q)^{-(m+r)} = r(r+1)(r+2)\cdots(r+m-1)(q/p)^m.
\end{aligned}$$

$$\mu_1' = E(X) = \frac{rq}{p}; \quad \mu_2' = E(X^2) = \frac{rq(1+rq)}{p^2}.$$

It can be seen that

$$\begin{aligned}
&\mu_3' = E(X^3) = \frac{q(rp^2 + 3pqr + q^2 r(r+1)}{p^3}, \\
&\mu_4' = E(X^4) = \frac{q(rp^3 + 7p^2 qr + 6pq^2 r(r+1) + q^3 r(r+1)(r+2)}{p^4}, \\
&\mu_2 = E((X-\mu_1')^2) = r(1-p), \\
&\mu_3 = E((X-\mu_1')^3) = \frac{r(p-1)(p-2)}{p^3}, \\
&\mu_4 = E((X-\mu_1')^4) = \frac{r(1-p)(6-6p+p^2+3r-3pr)}{p^4}, \\
&\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{2-p}{\sqrt{rq}} > 0 \quad \text{(positively skewed)}, \\
&\nu_2 = \nu_1 - 3 = \frac{p^2-2p+6}{rq}, \ \text{ where } \nu_1 = \frac{\mu_4}{\mu_2^2}.
\end{aligned}$$

$NB(1,p)$ distribution is called a geometric distribution (denoted by $Ge(p)$, $0 < p < 1$). The p.m.f. of $Y \sim Ge(p)$ is given by

$$f_Y(y) = P(Y = y) = \begin{cases} pq^y, & y = 0, 1, 2, \dots, \\ 0, & \text{otherwise.} \end{cases}$$

$P(Y \geq m) = p\sum_{y=m}^{\infty} q^y = q^m$. This implies that

$$\begin{aligned}
P(Y \geq m+n | Y \geq m) &= \frac{P(Y \geq m+n, Y \geq m)}{P(Y \geq m)} = \frac{P(Y \geq m+n)}{P(Y \geq m)} \\
&= \frac{q^{m+n}}{q^m} = q^n = P(Y \geq n), \ \forall\, m, n \in \{0, 1, \dots\}. \qquad (0.1)
\end{aligned}$$

Also,

$$P(Y \geq m+n) = P(Y \geq m)P(Y \geq n),\ \forall\, m, n \in \{0, 1, \dots\}. \tag{0.2}$$

Remark 0.5. *The property (0.1) possessed by $Ge(p)$ distribution has an interesting interpetation. Suppose that a device can absorb $0, 1, 2, \dots$ shocks before failing. Let T denote the random variable representing the number of shocks that device can absorb before failing.*

$P(T \geq m+n | T \geq m)$: *conditional probability that a system has absorbed m shocks will absorb atleast n additional shocks before failing.*

$P(T \geq n)$: *a new device can survive atleast n shocks before failing.*

Thus if distribution of T has property (0.1) then the age of the device has no effect as the residual (remaining) life of the device (implying that an used device is as good as a new device). The property (0.1) (or equivalently (0.2)) is famously known as Lack of memory (LoM) property.

Theorem 0.6. *Let T be a discrete type r.v. with range $S_T = \{0, 1, 2, \dots\}$. Then T has the lack of memory property if and only if $T \sim Ge(p)$, for some $p \in (0, 1)$.*

Proof. Obviously, $T \sim Ge(p)$, for some $p \in (0,1) \implies T$ has LoM property. Then $P(T \geq j+k) = P(T \geq j)P(T \geq k)\ \forall\, j, k \in \{0, 1, \dots\}$. Let $P(T = 0) = p$. Then $p \in (0, 1)$ and for $j \in \{0, 1, \dots\}$

$$\begin{aligned}
P(T \geq j+1) &= P(T \geq j)P(T \geq 1)\\
&= P(T \geq j)(1-p)\\
&= P(T \geq j-1)(1-p)^2\\
&\ \ \vdots\\
&= P(T \geq 0)(1-p)^{j+1} = (1-p)^{j+1}
\end{aligned}$$

This implies

$$P(T = k) = P(T \geq k) - P(T \geq k+1) = p(1-p)^k,\ \ k = \{0, 1, 2, \dots\} \implies T \sim Ge(p).$$

This completes the proof. $\square$

Example 0.7. *A person repeatedly rolls a fair die independently untill an upper face with two or three dots is observed twice. Find the probability that the person would require eight rolls to achive this.*

Solution: Consider getting 2 or 3 dots as success. Let $Z =$ the number of trials requires to get 2 successes. Then probability of success in each trial is $1/3$ and required probability $= P(Z = 8) = \left\{\binom{7}{1}\frac{1}{3}\left(\frac{2}{3}\right)^6\right\} \times \frac{1}{3} = \frac{448}{6561}$.

0.1.3. Hypergeometric Distribution

Consider a population comprising of N (≥ 2) units out of which $a \in \{1, 2, \dots, N-1\}$ are labelled as S (success) and $N - a$ are labeled as F (failure). A sample of size n is drawn from this population drawing one unit at a time. Let X denotes the number of successes in drawn sample.

Case-I: Drawn are independent and sampling is with replacement (*i.e.* after each draw the drawn units is replaced back into the population)

In this case we have sequence of n independent Bernoulli trials with probability of success in each trial as $p = \frac{a}{N}$. Thus $X \sim Bin(n, \frac{a}{N})$.

<u>Case-II</u>: Without replacement (*i.e.* drawn units are not replaced back into the population).

Here,

$$P(\text{obtaining } S \text{ in first draw}) = \frac{a}{N},$$
$$P(\text{obtaining } S \text{ in second draw}) = \frac{a}{N}\frac{a-1}{N-1} + \frac{N-a}{N}\frac{a}{N-1} = \frac{a}{N}.$$
$$\text{In general, } P(\text{obtaining } S \text{ in } i\text{th trial}) = \frac{a}{N}, \quad i = 1, 2, \ldots, n \text{ (Exercise)},$$
$$\begin{aligned}P(\text{obtaining } S \text{ in first and second trial}) &= \frac{a}{N}\frac{a-1}{N} \\ &\neq \frac{a}{N}\frac{a}{N} = P(\text{obtaining } S \text{ in first trial}) \times P(\text{obtaining } S \text{ in second trial}) \\ &\implies \text{Draws are not independent.}\end{aligned}$$

Thus, we can not conclude that $X \sim Bin(n, \frac{a}{N})$. So,

$$f_X(x) = P(X = x) = \begin{cases} \dfrac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}, & x = \max\{0, n-N+a\}, \ldots, \min\{n, a\}, \\ 0, & \text{otherwise.} \end{cases}$$

The distribution is called Hypergeometric distribution (Hyp(a, n, N)). It has three parameters $N \in \{2, 3, \ldots\}$, $a, n \in \{1, 2, \ldots, N-1\}$.

For $r \in \mathbb{N}$, let $X_{(r)} = X(X-1)(X-2)\cdots(X-r+1)$. Then

$$E(X_{(r)}) = \frac{1}{\binom{N}{n}} \sum_{k=\max\{0,n-N+a\}}^{\min\{n,a\}} k(k-1)(k-2)\cdots(k-r+1)\binom{a}{k}\binom{N-a}{n-k}.$$

Clearly for $r > \min\{n, a\}$, $E(X_{(r)}) = 0$. For $1 \leq r \leq \min\{n, a\}$, we have

$$\begin{aligned}
E(X_{(r)}) &= \frac{1}{\binom{N}{n}} \sum_{k=\max\{r,n-N+a\}}^{\min\{n,a\}} k(k-1)(k-2)\cdots(k-r+1)\binom{a}{k}\binom{N-a}{n-k} \\
&= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max\{r,n-N+a\}}^{\min\{n,a\}} \binom{a-r}{k-r}\binom{N-a}{n-k} \\
&= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max\{0,n-N+a-r\}}^{\min\{n-r,a-r\}} \binom{a-r}{k}\binom{N-a}{n-r-k} \\
&= \frac{a_{(r)}}{\binom{N}{n}} \sum_{k=\max\{0,(n-r)-(N-r)+a-r\}}^{\min\{n-r,a-r\}} \binom{a-r}{k}\binom{(N-r)-(a-r)}{(n-r)-k} = \frac{\binom{N-r}{n-r}}{\binom{N}{n}} a_{(r)}.
\end{aligned}$$

Since $\sum_{k=\max\{0,m-M+b\}}^{\min\{m,b\}} \binom{b}{k}\binom{M-b}{m-k} = \binom{M}{m}$. Thus, for $r \in \mathbb{N}$, we have

$$E(X_{(r)}) = \begin{cases} \frac{\binom{N-r}{n-r}}{\binom{N}{n}} a_{(r)}, & \text{if } r \leq \min\{n, a\}, \\ 0, & \text{if } r > \min\{n, a\}. \end{cases}$$

In particular,

$$E(X) = E(X_{(1)}) = n\frac{a}{N} = np \text{ (say), where } p = \frac{a}{N},$$

$$
\begin{aligned}
E(X(X-1)) = E(X_{(2)}) &= \frac{n(n-1)}{N(N-1)}a(a-1),\\
\mathrm{Var}(X) &= E(X^2) - (E(X))^2\\
&= E(X(X-1)) + E(X) - (E(X))^2\\
&= n\frac{a}{N}(1-\frac{a}{N})\frac{N-n}{N-1} = np(1-p)\left(1-\frac{n-1}{N-1}\right).
\end{aligned}
\tag{0.3}
$$

Remark 0.8. *In case of sampling with replacement we have $X \sim Bin(n,p)$, $E(X) = np$ and $\mathrm{Var}(X) = np(1-p)$, where $p = \frac{a}{N}$. The factor $(1 - \frac{n-1}{N-1})$ which on multiplying to variance of $Bin(n,p)$ distribution yields the variance of $Hyp(a,n,N)$ distribution (see 0.3) is called the finite population correction (f.p.c.). Clearly if the sample size n is significantly smaller than the population size N ($n << N$) then f.p.c. will be close to 1 and variance of $Bin(n,p)$ and $Hyp(a,n,N)$ distribution will be very close. Infact when $n << N$ and $n << a \equiv a_N$ (say) are such that $\frac{a_N}{N}$ is a fixed quantity (i.e. as $N \to \infty$, $a_N \to \infty$ and $\frac{a_N}{N} \to p \in (0,1)$, where $p \in (0,1)$ is a fixed quantity) then $Bin(n, \frac{a}{N})$ distribution provides an approximation to $Hyp(a,n,N)$ distribution. Regarding choice of sample size n for using this approximation a guideline based on various empirical studies, is that the sample size n should not exceed 10% of the population size N.*

Theorem 0.9 (Binomial Approximation to Hypergeometric Distributon)**.** *Let $X_{a_N,n,N} \sim Hyp(a_N,n,N)$, where a_N depends on N and $\lim\limits_{N\to\infty} \frac{a_N}{N} = p \in (0,1)$. Let $f_{a_N,n,N}(\cdot)$ denote the p.m.f. of $X_{a_N,n,N}$. Then*

$$
\lim_{N\to\infty} f_{a_N,n,N}(k) = \lim_{N\to\infty} P(X_{a_N,n,N} = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & k \in \{0,1,2,\ldots,n\},\\ 0, & \textit{otherwise,} \end{cases}
$$

i.e., for large N and large a_N, so that $p = \frac{a_N}{N} \in (0,1)$ is a fixed quantity, $Hyp(a_N,n,N)$ probabilities can be approximated by $Bin(n, \frac{a}{N})$ probabilities.

Proof. $S_X = \{m \in \mathbb{N} : \max\{0, n-N+a_N\} \le m \le \min\{n, a_N\}\}$, $n - N + a_N = N(\frac{n}{N} - 1 + \frac{a_N}{N}) \to \infty$ and $a_N = N\frac{a_N}{N} \to \infty$, as $N \to \infty$. Also for $k \in S_X$,

$$
\begin{aligned}
f_X(k) &= \binom{n}{k}\left\{\prod_{j=0}^{k-1}\left(\frac{a_N - j}{N-j}\right)\right\}\left\{\prod_{j=0}^{n-k-1}\left(\frac{N-a_N-j}{N-j}\right)\right\}\\
&\xrightarrow{N\to\infty} \binom{n}{k}\left\{\prod_{j=0}^{k-1}(p)\right\}\left\{\prod_{j=0}^{n-k-1}(1-p)\right\} = \binom{n}{k}p^k(1-p)^{n-k}\\
\implies \lim_{N\to\infty} f_{a_N,n,N}(k) &= \begin{cases} \binom{n}{k}p^k(1-p)^{n-k}, & k \in \{0,1,2,\ldots,n\}\\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

This completes the proof. □

The m.g.f. of $X \sim Hyp(a,n,N)$, althouh exists (since S_X is finite), can not be expressed in closed form.

0.1.4. Poisson Distribution

Some event E (say number of cars crossing a particular bridge/tunnel) is occuring randomly over a period of time. Let X denotes the number of times E has occured in an unit interval (say $(0,1]$).

To model probability distribution of X, partition the unit interval into a large number (say n where $n \to \infty$) of infinitesimal subintervals $(\frac{i-1}{n}, \frac{i}{n}]$, $i = 1, 2, \ldots, n$ of length $\frac{1}{n}$ each. In many situations, it may be relevant to assume that

(i) For each infinitesimal interval $(\frac{i-1}{n}, \frac{i}{n}]$, $i = 1, 2, \ldots, n$, the probability that E will occur in this interval is p_n and that it will not occur in this interval is $1 - p_n$; here $p_n \to 0$ as $n \to \infty$ and $np_n \to \lambda \in (0, \infty)$ as $n \to \infty$.

(ii) Chance of two or more occurences of E in any infinitesimal interval $(\frac{i-1}{n}, \frac{i}{n}]$, $i = 1, 2, \ldots, n$, is so small that it can be neglected.

(iii) occurrences of E in two disjoint infinitesimal intervals are independent.

$X \equiv X_n =$ the number of times event E occurs in $(0, 1] \sim Bin(n, p_n)$. The p.m.f. of X_n is

$$\begin{aligned}
f_n(k) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} I_{\{0,1,\ldots,n\}}(k) \\
&= \frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)(np_n)^k\left(1-\frac{np_n}{n}\right)^{n-k} I_{\{0,1,\ldots,n\}}(k) \\
&\to \frac{e^{-\lambda}\lambda^k}{k!} I_{\{0,1,\ldots,n\}}(k) \\
&= \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!}, & k \in \{0,1,2,\ldots\}, \\ 0, & \text{otherwise.} \end{cases} \to \text{Poisson distribution } (Po(\lambda) : \lambda > 0) \text{(family of probability distributions).}
\end{aligned}$$

A r.v. X is said to have a Poisson distribution with parameter $\lambda > 0$ (written as $X \sim Po(\lambda)$) if its p.m.f. is given by

$$f_X(k) = P(X = k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!}, & k = 0, 1, 2, \ldots, \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 0.10 (Poisson Approximation to Binomial Distribution)**.** *Let* $X_n \sim Bin(n, p_n)$*,* $n = 1, 2, \ldots,$ *where* $p_n \in (0, 1)$*,* $n = 1, 2, \ldots$ *and* $\lim_{n\to\infty}(np_n) = \lambda$*, for some* $\lambda > 0$*. Then*

$$\lim_{n\to\infty} f_{X_n}(k) = \lim_{n\to\infty} P(X_n = k) = \begin{cases} \dfrac{e^{-\lambda}\lambda^k}{k!}, & k = \{0, 1, 2, \ldots\}, \\ 0, & \textit{otherwise.} \end{cases}$$

Proof. As above. □

Remark 0.11. *If n is large and p is small (*$p_n \to 0$ *as* $n \to \infty$*) so that np is a fixed quantity in* $(0, \infty)$ *(*$np_n \to \lambda > 0$*) then Poisson distribution provides a good approximation to Binomial distribution.*

Example 0.12. *Consider a person who plays a series of* 2500 *games independently. If the probability of person winning any game is* 0.002*, find the probability that the person will win atleast two games.*

Solution: Let X denote the number of wins (successes) in 2500 games played by person.

Clearly $X \sim Bin(2500, 0.002)$, where $n = 2500$ and $np = 5$ (=λ, say) is fixed. Therefore,

$$P(X \geq 2) \approx P(Y \geq 2), \text{ where } Y \sim Po(5).$$

Thus, $P(X \geq 2) \approx 1 - (P(Y = 0) + P(Y = 1)) = 1 - (e^{-5} + 5e^{-5}) = 0.9596$.

Suppose that $X \sim Po(\lambda)$, for some $\lambda > 0$. Then for $r \in \{1, 2, \dots\}$, we have

$$\begin{aligned} E(X_{(r)}) &= E(X(X-1)\cdots(X-r+1)) \\ &= \sum_{k=0}^{\infty} k(k-1)(k-2)\cdots(k-r+1)\frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=r}^{\infty}\frac{\lambda^k}{(k-r)!} = e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^{j+r}}{j!} = \lambda^r e^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda^r. \end{aligned}$$

Thus,

$$\begin{aligned} \mu_1 &= E(X) = E(X_{(1)}) = \lambda, \\ E(X^2) &= E(X_{(2)}) + E(X) = \lambda^2 + \lambda, \quad \mathrm{Var}(X) = E(X^2) - (E(X))^2 = \lambda \ (\sigma^2 = \mu_2) \quad \text{(Mean=Variance)}, \\ \mu_3' &= E(X^3) = \lambda(\lambda^2 + 3\lambda + 1), \\ \mu_4' &= E(X^4) = \lambda(\lambda^3 + 6\lambda^2 + 7\lambda + 1), \ \ \mu_3 = \lambda; \ \mu_4 = \lambda(3\lambda+1), \\ \beta_1 &= \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{\lambda}}, \ \ \nu_2 = \nu_1 - 3 = \frac{\lambda(3\lambda+1)}{\lambda^2} - 3 = \frac{1}{\lambda}, \\ M_X(t) &= E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk}\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^{tk})^k}{k!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}, \ \ t \in \mathbb{R}, \\ \psi_X(t) &= \ln M_X(t) = \lambda(e^t - 1), \quad \psi_X^r(t) = \lambda e^t, \ \ r = 1, 2, \dots, \\ &\implies E(X) = \psi_X^1(0) = \lambda, \ \ \mathrm{Var}(X) = \psi_X^2(0) = \lambda. \end{aligned}$$

Theorem 0.13. *Let $X_1, X_2, \dots, X_k$ be independent r.v.'s such that $X_i \sim Po(\lambda_i)$, for some $\lambda_i > 0$, $i = 1, 2, \dots, k$. Then $Y = \sum_{i=1}^k X_i \sim Po(\lambda)$, where $\lambda = \sum_{i=1}^k \lambda_i$.*

Proof. For $t \in \mathbb{R}$,

$$\begin{aligned} M_Y(t) = E(e^{tY}) = E(e^{t\sum_{i=1}^k X_i}) = E\left(\prod_{i=1}^k e^{tX_i}\right) &= \prod_{i=1}^k E(e^{tX_i}), \quad \text{(independent of } X_i's) \\ &= \prod_{i=1}^k M_{X_i}(t) = \prod_{i=1}^k e^{\lambda_i(e^t-1)} = e^{\lambda(e^t-1)}. \end{aligned}$$

This implies that $Y \sim Po(\lambda)$, where $\lambda = \sum_{i=1}^k \lambda_i$. □

0.1.5. Discrete Uniform Distribution

Let N be a given positive integer and $x_1 < x_2 < \dots < x_N$ be given real numbers. A r.v. X is said to follow a discrete uniform distribution on the set $\{x_1, x_2, \dots, x_N\}$ (written as $X \sim U(\{x_1, x_2, \dots, x_N\})$) if its p.m.f. is given by

$$f_X(x) = P(X = x) = \begin{cases} \dfrac{1}{N}, & x \in \{x_1, x_2, \dots, x_N\}, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that $X \sim U(\{x_1, x_2, \dots, x_N\})$. Then,

$$\mu_r' = E(X^r) = \frac{1}{N}\sum_{i=1}^N x_i^r,$$

$$\text{Mean} = \mu_1' = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

$$\text{Var}(X) = \sigma^2 = E((X-\mu_1')^2) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_1')^2,$$

$$\text{m.g.f. } M_X(t) = E(e^{tX}) = \frac{1}{N}\sum_{i=1}^{N} e^{tx_i}.$$

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

$$\begin{aligned}
\mu_1' &= \text{Mean} = E(Y) = \frac{1}{N}\sum_{i=1}^{N} i = \frac{N+1}{2},\\
\mu_2' &= E(Y^2) = \frac{1}{N}\sum_{i=1}^{N} i^2 = \frac{(N+1)(2N+1)}{6},\\
\mu_3' &= E(Y^3) = \frac{1}{N}\sum_{i=1}^{N} i^3 = \frac{N(N+1)^2}{4},\\
\mu_4' &= E(Y^4) = \frac{1}{N}\sum_{i=1}^{N} i^4 = \frac{(N+1)(2N+1)(3N^2+3N-1)}{30},\\
\mu_2 &= E((Y-\mu_1')^2) = \frac{N^2-1}{12},\\
\mu_3 &= E((Y-\mu_1')^3) = 0,\\
\mu_4 &= E((Y-\mu_1')^4) = \frac{(3N^2-7)(N^2-1)}{240},
\end{aligned}$$

$$\text{Coefficient of skewness} = \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0,$$

$$\text{Kurtosis} = \nu_1 = \frac{\mu_4}{\mu_2^2} = \frac{3}{5}\frac{(3N^2-7)}{N^2-1},$$

$$\text{m.g.f. } M_Y(t) = E(e^{tY}) = \frac{1}{N}\sum_{j=1}^{N} e^{tj} = \begin{cases} \dfrac{e^t(e^{Nt}-1)}{N(e^t-1)}, & t \neq 0,\\ 1, & t = 0.\end{cases}$$

Example 0.14. *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 unsuuccessful 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 $Y \sim U(\{1,2,\dots,N\})$ and hence find the mean and variance of the r.v. Y.*

Solution: For $r \neq \{1,2,\dots,N\}$, we have $P(Y=r)=0$. For $r \in \{1,2,\dots,N\}$, we have

$$P(Y=r) = \frac{N-1}{N}\cdot\frac{N-2}{N-1}\cdots\frac{N-(r-1)}{N-(r-2)}\cdot\frac{1}{N-(r-1)} = \frac{1}{N} \implies Y \sim U(\{1,2,\dots,N\}).$$

This implies $E(Y) = \dfrac{N+1}{2}$ and $\text{Var}(Y) = \dfrac{N^2-1}{12}$.

0.2. Some Special Continuous Distributions

0.2.1. Uniform or Rectangular Distribution

Let $-\infty < \alpha < \beta < \infty$. An absolutely continuous type r.v. X is said to have a uniform (or rectangular) distribution over the interval (α, β) (written as $X \sim U(\alpha, \beta)$) if its p.d.f. is given by

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

$\{U(\alpha,\beta) : -\infty < \alpha < \beta < \infty\}$ is a family of distributions corresponding to different choices of α and β ($-\infty < \alpha < \beta < \infty$).

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

$$\mu_r' = E(X^r) = \int_\alpha^\beta \frac{x^r}{\beta-\alpha}\mathrm{d}x = \frac{\beta^{r+1}-\alpha^{r+1}}{(r+1)(\beta-\alpha)} = \frac{\beta^r}{r+1}\left[1+\frac{\alpha}{\beta}+\left(\frac{\alpha}{\beta}\right)^2+\cdots+\left(\frac{\alpha}{\beta}\right)^r\right],$$

$$E(X) = \frac{\alpha+\beta}{2} = \mu_1',$$

$$\mu_r = E(X-\mu_1')^r = \int_\alpha^\beta \left(x-\frac{\alpha+\beta}{2}\right)^r \frac{1}{\beta-\alpha}\mathrm{d}x = \int_{-\frac{\beta-\alpha}{2}}^{\frac{\beta-\alpha}{2}} \frac{t^r}{\beta-\alpha}\mathrm{d}t = \begin{cases} 0, & r = 1,3,5,\ldots, \\ \frac{(\beta-\alpha)^r}{2^r(r+1)}, & r = 2,4,6,\ldots. \end{cases}$$

Also,

$$f_X\left(x-\frac{\alpha+\beta}{2}\right) = f_X\left(\frac{\alpha+\beta}{2}-x\right) = \begin{cases} \frac{1}{\beta-\alpha}, & -\frac{\beta-\alpha}{2} < x < \frac{\beta-\alpha}{2}, \\ 0, & \text{otherwise} \end{cases} \implies X - \frac{\alpha+\beta}{2} \stackrel{d}{=} \frac{\alpha+\beta}{2} - X.$$

This implies distribution of X is symmetric about its mean $\mu_1' = \frac{\alpha+\beta}{2}$.

$$\mathrm{Var}(X) = \mu_2 = \sigma^2 = E((X-\mu_1')^2) = \frac{(\beta-\alpha)^2}{12},$$

$$\text{Coefficient of skewness} = \beta_1 = \frac{\mu_3}{\mu_2^{3/2}} = 0,$$

$$\text{Kurtosis} = \nu_1 = \frac{\mu_4}{\mu_2^2} = \frac{9}{5} = 1.8,$$

$$\text{The d.f. of } X \sim U(\alpha,\beta) \text{ is given by } F(x) = \begin{cases} 0, & x < \alpha, \\ \frac{x-\alpha}{\beta-\alpha}, & \alpha \le x < \beta, \\ 1, & x \ge \beta. \end{cases}$$

Theorem 0.15. *Let $-\infty < \alpha < \beta < \infty$ and let X be a r.v. of continuous type with $P(\alpha \le X \le \beta) = 1$. Then $X \sim U(\alpha,\beta) \iff P(X \in I) = P(X \in J)$, for any pairs of intervals $I, J \subseteq (\alpha, \beta)$ having the same length.*

Proof. Suppose that $X \sim U(\alpha, \beta)$. Then, for $\alpha \le a < b \le \beta$, we have

$$\begin{aligned} P(X \in (a,b)) &= P(X \in [a,b)) = P(X \in (a,b]) = P(X \in [a,b]) \\ &= F(b|\alpha,\beta) - F(a|\alpha,\beta) \end{aligned}$$

$$= \frac{b-a}{\beta-\alpha} \to \text{dependes only on length } b-a \text{ of the interval } (a,b)/[a,b)/(a,b]/[a,b].$$

Conversely, suppose that $P(X \in I) = P(X \in J)$, for all pairs of intervals $I, J \subseteq (\alpha, \beta)$ having the same length. For $0 < s \leq 1$, let $G(s) = P(\alpha < X \leq \alpha + (\beta-\alpha)s) = F(\alpha + (\beta-\alpha)s|\alpha,\beta)$. Then for $0 < s_1, s_2 \leq 1$, $0 < s_1 + s_2 \leq 1$,

$$\begin{aligned} G(s_1+s_2) &= P(\alpha < X \leq \alpha + (\beta-\alpha)(s_1+s_2)) \\ &= P(\alpha < X \leq \alpha + (\beta-\alpha)s_1) + P\left(\underbrace{\alpha + (\beta-\alpha)s_1 < X \leq \alpha + (\beta-\alpha)(s_1+s_2)}_{\text{Depends only on the length } (\beta-\alpha)s_2 \text{ of } (\alpha+(\beta-\alpha)s_1, \alpha+(\beta-\alpha)(s_1+s_2))}\right) \\ &= G(s_1) + P(\alpha < X \leq \alpha + (\beta-\alpha)s_2) = G(s_1) + G(s_2). \end{aligned}$$

By induction, for $0 < s_i \leq 1$, $i = 1, 2, \ldots, n$, $0 < \sum_{i=1}^n s_i \leq 1$, we have $G(s_1 + s_2 + \cdots + s_n) = G(s_1) + G(s_2) + \cdots + G(s_n)$. This implies that

$$G(ms) = mG(s),\ \forall\, 0 < s \leq \frac{1}{m}, \tag{0.4}$$

$$G(s) = G\left(\underbrace{\frac{s}{n} + \frac{s}{n} + \cdots + \frac{s}{n}}_{n \text{ times}}\right) = nG\left(\frac{s}{n}\right). \tag{0.5}$$

For $m, n \in \{1, 2, \ldots\}$, $m < n$, we get

$$\begin{aligned} G\left(\frac{m}{n}\right) &= G\left(\underbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}_{m \text{ times}}\right) \\ &= mG\left(\frac{1}{n}\right), \quad \text{using (0.4)} \\ &= \frac{m}{n} G(1), \quad \text{using (0.5)} \\ &= \frac{m}{n} F(\beta|\alpha,\beta) = \frac{m}{n} \implies G(r) = r,\ \forall\, r \in IQ \cap (0,1), \end{aligned}$$

where IQ denotes the set of rational numbers. Now let $x \in (0,1)$. Then there exists a sequence $\{r_n\}_{n\geq 1}$ in $IQ \cap (0,1)$ such that $r_n \downarrow x$ (rationals are dense in $(0,1)$). Then, since G is continuous, we have

$$G(x) = \lim_{n\to\infty} G(r_n) = \lim_{n\to\infty} r_n = x\ \forall\, x \in (0,1).$$

This implies

$$\begin{aligned} & F(\alpha + (\beta-\alpha)x|\alpha,\beta) = x, \quad \forall\, x \in (0,1) \\ \implies & F(x|\alpha,\beta) = \frac{x-\alpha}{\beta-\alpha}, \quad x \in (\alpha,\beta) \\ \implies & F(x|\alpha,\beta) = \begin{cases} 0, & x < \alpha, \\ \dfrac{x-\alpha}{\beta-\alpha}, & \alpha \leq x < \beta, \\ 1, & x \geq \beta, \end{cases} \implies X \sim U(\alpha,\beta). \end{aligned}$$

This completes the proof. $\square$

$$\text{M.g.f. } M_X(t) = E(e^{tX}) = \int_{\alpha}^{\beta} \frac{e^{tx}}{\beta - \alpha} \mathrm{d}x = \begin{cases} \dfrac{e^{t\beta} - e^{t\alpha}}{t(\beta - \alpha)}, & t \neq 0, \\ 1, & t = 0. \end{cases}$$

Theorem 0.16. *Let $X \sim U(\alpha, \beta)$, $-\infty < \alpha < \beta < \infty$. 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 $b \in \mathbb{R}$, $Y = aX + b \sim U(a\beta + b, a\alpha + b)$,

(iii) $Z = \dfrac{X - \alpha}{\beta - \alpha} \sim U(0, 1)$.

Proof. Exercise. □

Recall that quantile function is defined by $Q_X(p) = \inf\{s \in \mathbb{R} : F_X(s) \geq p\}$, $0 < p < 1$.

Theorem 0.17. *Let X be a r.v. with d.f. F and quantile function $Q(\cdot)$. Then*

*(i) (**Probability Integral Transform**)*

$$X \text{ is of continuous type } \implies F(X) \sim U(0,1)$$

(ii) $U \sim U(0,1) \implies Q(U) \stackrel{d}{=} X$.

Proof. (i) Let G be the d.f. of $Y = F(X)$. Then $G(y) = P(F(X) \leq y)$, $y \in \mathbb{R}$. Clearly, for $y < 0$, $G(y) = 0$ and for $y \geq 1$, $G(y) = 1$. For $y \in [0, 1)$,

$$\begin{aligned} &\{s \in \mathbb{R} : F(s) \geq y\} = \{s \in \mathbb{R} : s \geq Q(y)\} \\ \implies &P(F(X) \geq y) = P(X \geq Q(y)) \\ \implies &P(F(X) < y) = P(X < Q(y)) \\ \implies &P(F(X) < y) = P(X \leq Q(y)) = F(Q(y)) = y, \quad \text{since } X \text{ is of continuous type.} \end{aligned}$$

Since X is of continuous type $P(F(X) = y) = P(x_1 \leq X \leq x_2) = 0$ for some x_1, x_2 with $F(x_1) = F(x_2)$. Thus,

$$\begin{aligned} &P(F(X) \leq y) = y, \ \forall\, y \in (0,1), \\ \implies G(y) = &\begin{cases} 0, & \text{if } y < 0, \\ y, & \text{if } 0 \leq y < 1, \\ 1, & \text{if } y \geq 1, \end{cases} \implies Y \sim U(0,1). \end{aligned}$$

(ii) Let $U \sim U(0,1)$ and let $Z = Q(U)$. Then the d.f. of Z is

$$H(z) = P(Z \leq z) = P(Q(U) \leq z) = P(Q(U) \leq z, 0 < U < 1).$$

Note that for $z \in (0,1)$, $\{p \in \mathbb{R} : Q(p) \leq z\} = \{p \in \mathbb{R} : F(z) \geq p\}$. Thus, for $z \in (0,1)$

$$H(z) = P(F(Z) \geq U, 0 < U < 1) = P(U \leq F(z)) = F(z) \implies Z = Q(U) \stackrel{d}{=} X.$$

This completes the proof. □

Remark 0.18. *The above theorem provides a method to generate observations from any arbitrary distributions using $U(0,1)$ observations. Suppose that we require an observation X from a distribution having d.f. F and quantile functions Q. To do so, the above theorem suggests that generate an observation U from $U(0,1)$ distribution and take $X = Q(U)$.*

0.2.2. Gamma and Related Distributions

Gamma Function: $\Gamma : (0,\infty) \to (0,\infty)$ is defined as

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} \mathrm{d}t, \quad \alpha > 0.$$

It converges for any $\alpha > 0$. Integration by parts yields $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, $\alpha > 0$ and $\Gamma(1) = 1$. For any $n \in \mathbb{N}$,

$$\Gamma(n) = (n-1)!, \quad \Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-t} t^{-1/2} \mathrm{d}t = 2\int_0^\infty e^{-x^2} \mathrm{d}x$$

This implies

$$\begin{aligned}\left(\Gamma\left(\frac{1}{2}\right)\right)^2 &= 4\int_0^\infty \int_0^\infty e^{-(x^2+y^2)} \mathrm{d}x\mathrm{d}y \\ &= 4\int_0^\infty \int_0^{\pi/2} re^{-r^2} \mathrm{d}\theta\mathrm{d}r, \quad (x = r\cos\theta,\ y = r\sin\theta) \\ &= \pi \implies \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.\end{aligned}$$

Also,

$$\begin{aligned}&\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2},\ \Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\cdot\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1\cdot 3\sqrt{\pi}}{2^2}, \\ &\Gamma\left(\frac{2n+1}{2}\right) = \frac{1\cdot 3\cdot 5\cdots(2n-1)\sqrt{\pi}}{2^n} = \frac{(2n)!}{n!4^n}\sqrt{\pi},\ \ n \in \mathbb{N}.\end{aligned}$$

Clearly,

$$\int_0^\infty e^{-x/\theta} x^{\alpha-1} \mathrm{d}x = \theta^\alpha \Gamma(\alpha),\ \ \alpha > 0,\ \theta > 0.$$

Definition 0.19. *A r.v. X is said to have a gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$ (written as $X \sim GAM(\alpha,\theta)$ if its p.d.f. is given by*

$$f(x|\alpha,\theta) = \begin{cases} \dfrac{e^{-x/\theta}x^{\alpha-1}}{\theta^\alpha\Gamma(\alpha)}, & x > 0, \\ 0, & \textit{otherwise}. \end{cases} \to \textit{family of distributions}\ \ \{GAM(\alpha,\theta), \alpha > 0, \theta > 0\}.$$

Let $X \sim GAM(\alpha,\theta) \implies \dfrac{X}{\theta} \sim GAM(\alpha,1)$ (θ is called scale parameter since the distribution of $\dfrac{X}{\theta}$ does not depend on θ). The p.d.f. of $Z \sim GAM(\alpha,1)$ is $f(z) = \begin{cases} \dfrac{e^{-z}z^{\alpha-1}}{\Gamma(\alpha)}, & z > 0, \\ 0, & \text{otherwise}. \end{cases}$

Also,

$$\begin{aligned}E(Z^r) = \frac{1}{\Gamma(\alpha)}\int_0^\infty z^{\alpha+r-1}e^{-z}\mathrm{d}z &= \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)},\ \ r > -\alpha,\ \alpha > 0, \\ &= \alpha(\alpha+1)\cdots(\alpha+r-1),\ \text{ if }\ r \in \mathbb{N}.\end{aligned}$$

$$\text{Mean} = \mu_1' = E(X) = \alpha\theta,\ \ \mu_2' = E(X^2) = \alpha(\alpha+1)\theta^2,\ \ \mu_2 = \sigma^2 = \mathrm{Var}(X) = \alpha\theta^2,$$

$$\mu_3 = E((X-\mu_1')^3) = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3 = 2\alpha\theta^3,$$
$$\mu_4 = E((X-\mu_1')^4) = \mu_4' - 4\mu_1'\mu_3' + 6(\mu_1')^2\mu_2' - 3(\mu_1')^4 = 3\alpha(\alpha+2)\theta^4,$$
$$\text{Coefficient of skewness} = \beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{2}{\sqrt{\alpha}}, \ \text{Kurtosis} = \nu_1 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{6}{\alpha}.$$

For $0 < \alpha \le 1$, $f(x|\alpha,\theta) \downarrow$ and for $\alpha > 1$, $f(x|\alpha,\theta) \uparrow$ in $(0,(\alpha-1)\theta)$ and $\downarrow$ in $((\alpha-1)\theta,\infty)$.

$$\text{m.g.f. } M_X(t) = E(e^{tX}) = E(e^{t\theta Z}), \quad (Z = X/\theta)$$
$$= \frac{1}{\Gamma(\alpha)}\int_0^\infty e^{t\theta z}e^{-z}z^{\alpha-1}\mathrm{d}z = \frac{1}{\Gamma(\alpha)}\int_0^\infty e^{-(1-t\theta)z}z^{\alpha-1}\mathrm{d}z = (1-t\theta)^{-\alpha}, \ t < \frac{1}{\theta}.$$

Theorem 0.20. *Let $X_1, X_2, \ldots, X_k$ be independent r.v.'s such that $X_i \sim GAM(\alpha_i,\theta)$, for some $\alpha_i > 0$, $\theta > 0$, $i = 1,2,\ldots,k$. Then $Y = \sum_{i=1}^k X_i \sim GAM(\sum_{i=1}^k \alpha_i,\theta)$.*

Proof. Note that

$$M_Y(t) = \prod_{i=1}^k M_{X_i}(t) = \prod_{i=1}^k (1-t\theta)^{-\alpha_i} = (1-t\theta)^{-\sum_{i=1}^k \alpha_i}, \ t < \frac{1}{\theta} = \text{m.g.f. of } GAM(\sum_{i=1}^k \alpha_i, \theta).$$

This completes the proof. □

Theorem 0.21 (Relationship beween Gamma and Poisson distribution). *For $n \in \mathbb{N}$, $\theta > 0$ and $t > 0$, let $X \sim GAM(n,\theta)$ and $Y \sim Po(t/\theta)$. Then $P(X > t) = P(Y \le n-1)$, i.e.*

$$\frac{1}{(n-1)!\theta^n}\int_t^\infty e^{-x/\theta}x^{n-1}\mathrm{d}x = \sum_{j=0}^{n-1}\frac{e^{-t/\theta}(t/\theta)^j}{j!}.$$

Proof. Use integration by parts. □

Remark 0.22. *For $n \in \mathbb{N}$ and $\theta > 0$, let $X \sim GAM(n,\theta)$. Then*

$$\sum_{j=n}^\infty \frac{e^{-X/\theta}(X/\theta)^j}{j!} \sim U(0,1) \ \text{ and } \ \sum_{j=0}^{n-1}\frac{e^{-X/\theta}(X/\theta)^j}{j!} \sim U(0,1) \quad (U \sim U(0,1) \implies 1-U \sim U(0,1)).$$

Definition 0.23. *For a $\theta > 0$, a $GAM(1,\theta)$ distribution is called exponential distribution with scale parameter θ (denoted by $Exp(\theta)$).*

The p.d.f. of $T \sim Exp(\theta)$ is given by

$$f_T(t) = \begin{cases} \frac{1}{\theta}e^{-t/\theta}, & t > 0, \\ 0, & \text{otherwise.} \end{cases}$$

and its d.f. is given by

$$F_T(t) = P(T \le t) = 1 - P(T > t) = \begin{cases} 0, & t \le 0, \\ 1 - e^{-t/\theta}, & t > 0. \end{cases}$$

Mean= $E(T) = \theta$, variance= θ^2, $\mu_r' = E(T^r) = r!\theta^r$, $r \in \mathbb{N}$, coefficient of skewness=$\beta_1 = 2$, Kurtosis=$\nu_1 = 9$. M.g.f.=$M_T(t) = (1-t\theta)^{-1}$, $t < 1/\theta$ and

$$P(T > t) = \begin{cases} 1, & t \le 0, \\ e^{-t/\theta}, & t > 0. \end{cases}$$

For $s > 0, t > 0$

$$P(T > s+t|T > s) = \frac{P(T > s+t)}{P(T > s)} = e^{-t/\theta} = P(T > t)$$
$$\implies P(T > s+t) = P(T > s)P(T > t),\ \forall\, s,t > 0 \to \text{Lack of Memory Property.}$$

Let T denote the lifetime of a system. Given that the system has survived $s(> 0)$ units of time the probability that it will survive t additional units of time is the same as the probability that a fresh system (of age 0) will survive t units of time. In other words, the system has no memory of its current age or it is not ageing with time.

Theorem 0.24. *Let Y be a r.v. of continuous type with d.f. F such that $F(0) = 0$. Then Y has LoM property (i.e. $\bar{F}(s+t) = \bar{F}(s)\bar{F}(t)\ \forall\, s,t > 0$, where $\bar{F} = 1 - F$) iff $Y \sim Exp(\theta)$, for some $\theta > 0$.*

Proof. Let $Y \sim Exp(\theta)$, $\theta > 0$. Then Y has LoM property (already discussed). Now suppose that $F(0) = 0$ and Y has LoM property. Then

$$\bar{F}(s+t) = \bar{F}(s)\bar{F}(t)\ \forall\, s,t > 0,$$
$$\implies \bar{F}(s_1 + s_2 + \cdots + s_m) = \bar{F}(s_1)\bar{F}(s_2)\cdots\bar{F}(s_m),\quad s_i > 0,\ \ i = 1,2,\ldots,m,$$
$$\implies \bar{F}\left(\frac{m}{n}\right) = \bar{F}\left(\underbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}_{m \text{ times}}\right) = \left[\bar{F}\left(\frac{1}{n}\right)\right]^m\ \forall\, m,n \in \mathbb{N}, \tag{0.6}$$
$$\implies \bar{F}(1) = \left[\bar{F}\left(\frac{1}{n}\right)\right]^n\ \forall\, n \in \mathbb{N}, \tag{0.7}$$
$$\implies \bar{F}\left(\frac{m}{n}\right) = \left[\bar{F}(1)\right]^{m/n}\ \forall\, m,n \in \mathbb{N}. \tag{0.8}$$

Let $\lambda = \bar{F}(1)$ so that $0 \le \lambda \le 1$.

$$\lambda = 0 \implies \bar{F}\left(\frac{1}{n}\right) = 0\ \forall\, n \in \mathbb{N}\ \text{(using 0.7)} \implies \bar{F}(0) = 0 \implies F(0) = 1\ \ \text{(contradiction, since } F(0) = 0)$$

$$\lambda = 1 \implies \bar{F}(m) = \left[\bar{F}(1)\right]^m = 1\ \forall\, m \in \mathbb{N} \implies \lim_{m\to\infty}\bar{F}(m) = 1 \implies \lim_{m\to\infty}F(m) = 0\ \ \to \text{contradiction.}$$

Thus $\lambda \in (0,1)$. Let $\lambda = e^{-1/\theta}$, $\theta > 0$ ($\theta = -1/\ln\lambda$). Then using (0.8), $\bar{F}(r) = e^{-r/\theta}\ \forall\, r \in IQ \cap (0,\infty)$. Let $x \in IQ \cap (0,\infty)$. Then there exists a sequence $\{r_n\}_{n\ge1}$ in $IQ \cap (0,\infty)$ such that $r_n \to x$. Then

$$\bar{F}(x) = \bar{F}\left(\lim_{n\to\infty} r_n\right) = \lim_{n\to\infty}\bar{F}(r_n) = \lim_{n\to\infty} e^{-r_n/\theta} = e^{-x/\theta},$$
$$\implies F(x) = \begin{cases} 0,\ x < 0, \\ 1 - e^{-x/\theta},\ x \ge 0, \end{cases} \implies Y \sim Exp(\theta).$$

This completes the proof. □

Example 0.25. X : *Waiting time for occurrence of an event E. Suppose that $X \sim Exp(3)$. Then the conditional probability that the waiting time for occurrences of E is atleast 5 hrs given that it has not occurred in first two hrs* $=P(X > 5|X > 2) = P(X > 3) = e^{-1}$.

Chi-squared Distribution: Let $n \in \mathbb{N}$. Then $GAM\left(\frac{n}{2}, 2\right)$ distribution is called Chi-squared distribution with n degrees of freedom (denoted by χ_n^2). Let $X \sim \chi_n^2$. The p.d.f. of X is

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

Mean = $E(X) = n$, $\text{Var}(X) = \mu_2 = \sigma^2 = 2n$, coefficient of skewness = $\beta_1 = 2\sqrt{\frac{2}{n}}$, Kurtosis = $\nu_1 = 3 + \frac{12}{n}$, m.g.f. $M_X(t) = (1-2t)^{-n/2}$, $t < \frac{1}{2}$.

Theorem 0.26. *Let $X_1, X_2, \ldots, X_k$ be independent with $X_i \sim \chi^2_{n_i}$, $n_i \in \mathbb{N}$, $i = 1, 2, \ldots, k$. Then $\sum_{i=1}^k X_i \sim \chi^2_n$, where $n = \sum_{i=1}^k n_i$.*

For various values of $n \in \mathbb{N}$ and $\alpha \in (0,1)$, tables for $(1-\alpha)$th quantile of χ^2_n distribution (*i.e.* $\tau_{n,\alpha}$ satisfying $P(\chi^2_n \le \tau_{n,\alpha}) = 1-\alpha$) are available in various textbook.

0.2.3. Beta Distribution

For $\alpha > 0$ and $\beta > 0$, we have

$$\begin{aligned}
\Gamma(\alpha)\Gamma(\beta) &= \int_0^\infty \int_0^\infty e^{-(s+t)} s^{\alpha-1} t^{\beta-1} \mathrm{d}s\mathrm{d}t \\
&= \int_0^1 \int_0^\infty e^{-v} (uv)^{\alpha-1} ((1-u)v)^{\beta-1} |v| \mathrm{d}v\mathrm{d}u, \\
&\qquad\qquad \text{making transformation: } s = uv,\ t = (1-u)v,\ \text{Jacobian : } J = v \\
&= \Gamma(\alpha+\beta) \int_0^1 u^{\alpha-1}(1-u)^{\beta-1} \mathrm{d}u \\
\implies \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} &= \int_0^1 u^{\alpha-1}(1-u)^{\beta-1} \mathrm{d}u \to \text{Beta function (function of } (\alpha,\beta),\ \alpha > 0,\ \beta > 0).
\end{aligned}$$

Note that $B(\alpha,\beta) = B(\beta,\alpha)$, $\forall\ \alpha,\ \beta > 0$.

Definition 0.27. *For given $\alpha > 0$ and $\beta > 0$, a r.v. X is said to have the beta distribution with parameter (α,β) (written as $X \sim Be(\alpha,\beta)$) if its p.d.f. is given by*

$$f(x|\alpha,\beta) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, & 0 < x < 1, \\ 0, & \textit{otherwise}. \end{cases}$$

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

$$\begin{aligned}
&E(X^r) = \frac{B(\alpha+r,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+r)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+r)},\ \ r > -\alpha, \\
&\text{Mean} = \mu_1' = E(X) = \frac{\alpha}{\alpha+\beta},\ \ \mu_2' = E(X^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}, \\
&\mu_2 = \sigma^2 = \text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}, \\
&\text{Mode} = M_0 = \frac{\alpha-1}{\alpha+\beta-2},\ \ \text{if } \alpha > 1 \text{ and } \alpha+\beta > 2, \\
&\text{Skewness} = \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{\sqrt{\alpha\beta}(\alpha+\beta+2)}, \\
&\text{Kurtosis} = \nu_1 = \frac{\mu_4}{\mu_2^2} = \frac{6[(\alpha-\beta)^2(\alpha+\beta+1) - \alpha\beta(\alpha+\beta+2)]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} + 3
\end{aligned}$$

$$= \frac{6[\alpha^3 + \alpha^2(1-2\beta) + \beta^2(1+\beta) - 2\alpha\beta(2+\beta)]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}.$$

Let $X \sim Be(\alpha, \alpha)$, $\alpha > 0$. Then

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

Clearly, $X \stackrel{d}{=} 1 - X \implies X - \frac{1}{2} \stackrel{d}{=} \frac{1}{2} - X$. Thus, if $X \sim Be(\alpha, \alpha)$. Then the distribuion of X is symmetric about $1/2$.

Theorem 0.28 (Relationship between Beta and Binomial Distribution)**.** *For $m, n \in \mathbb{N}$ and $x \in (0,1)$, let $X \sim Be(m,n)$ and $Y \sim Bin(m+n-1, x)$. Then $P(X \le x) = P(Y \ge m)$, i.e.*

$$\frac{1}{B(m,n)} \int_0^x t^{m-1}(1-t)^{n-1}\mathrm{d}t = \sum_{j=m}^{m+n-1} \binom{m+n-1}{j} x^j (1-x)^{m+n-1-j}.$$

Proof. Fix $m, n \in \mathbb{N}$ and $x \in (0,1)$. Let

$$\begin{aligned} I_{m,n} = LHS &= \frac{(m+n-1)!}{(m-1)!(n-1)!} \int_0^x t^{m-1}(1-t)^{n-1}\mathrm{d}t \\ &= \binom{m+n-1}{m} x^m (1-x)^{n-1} + \frac{(m+n-1)!}{m!(n-2)!} \int_0^x t^m (1-t)^{n-2}\mathrm{d}t \\ &= \binom{m+n-1}{m} x^m (1-x)^{n-1} + I_{m+1,n-1}. \end{aligned}$$

Proceeding recursively give the result. □

$$\begin{aligned} \text{m.g.f. } M_X(t) = E(e^{tX}) &= \frac{1}{B(\alpha,\beta)} \int_0^1 e^{tx} x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x \\ &= \frac{1}{B(\alpha,\beta)} \int_0^1 \left(\sum_{j=0}^{\infty} \frac{t^j x^j}{j!} \right) x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x \\ &= \frac{1}{B(\alpha,\beta)} \sum_{j=0}^{\infty} \frac{t^j}{j!} \int_0^1 x^{j+\alpha-1}(1-x)^{\beta-1}\mathrm{d}x = \frac{1}{B(\alpha,\beta)} \sum_{j=0}^{\infty} \frac{B(\alpha+j,\beta)t^j}{j!}, \quad t \in \mathbb{R}. \end{aligned}$$

Example 0.29. *Time (in hours) to finish a job follows beta distribution with mean $\frac{1}{3}$ hrs. and variance $\frac{2}{63}$ hrs. Find the probability that the job will be finished in 30 minutes.*

Solution: Define $X =$ time to finish job (in hours) $\sim Be(\alpha, \beta)$, say.

$E(X) = \frac{1}{3} \implies \frac{\alpha}{\alpha+\beta} = \frac{1}{3}$, $\text{Var}(X) = \frac{2}{63} \implies \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{2}{63}$. This implies $\alpha = 2$ and $\beta = 4$. Thus, $X \sim Be(2,4)$. Required probability

$$P(X < \frac{1}{2}) = \frac{1}{B(2,4)} \int_0^{1/2} x(1-x)^3\mathrm{d}x = \frac{13}{16}.$$

0.2.4. Normal Distribution

Recall that

$$\begin{aligned}\sqrt{\pi} = \Gamma\left(\frac{1}{2}\right) = \int_0^\infty e^{-t}t^{-1/2}\mathrm{d}t &= 2\int_0^\infty e^{-x^2}\mathrm{d}x \\ &= \int_{-\infty}^\infty e^{-x^2}\mathrm{d}x \\ &= \frac{1}{\sqrt{2}}\int_{-\infty}^\infty e^{-t^2/2}\mathrm{d}t \\ \implies \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-t^2/2}\mathrm{d}t &= 1 \\ \implies \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty e^{-(t-\mu)^2/2\sigma^2}\mathrm{d}t &= 1 \;\forall\; \mu\in\mathbb{R} \text{ and } \sigma>0.\end{aligned}$$

Definition 0.30. *Let $\mu \in \mathbb{R}$ and $\sigma > 0$ be given constants. An absolutely continuous type r.v. is said to follow a normal distribution with patrameters $\mu \in \mathbb{R}$ and $\sigma > 0$ (written as $X \sim N(\mu, \sigma^2)$) if its p.d.f. is given by*

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2},\;\; -\infty < x < \infty.$$

The $N(0,1)$ distribution is called standard normal distribution. The p.d.f. and d.f. of a standard normal distribution are denoted by $\phi(z)$ and $\Phi(z)$, respectively, so that

$$\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2},\;\; -\infty<z<\infty,\;\; \Phi(z) = \int_{-\infty}^z \phi(t)\mathrm{d}t = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^z e^{-t^2/2}\mathrm{d}t,\;\; z\in\mathbb{R}.$$

$$X\sim N(\mu,\sigma^2) \implies f(\mu-x|\mu,\sigma) = f(\mu+x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}},\;\; x\in\mathbb{R}.$$

This implies $X-\mu \overset{d}{=} \mu - X$ (distribution of X is symmetric about μ) $\implies$ $E(X) = \mu$ and $F(\mu|\mu,\sigma) = \frac{1}{2}$. Moreover,

$$P(X-\mu\le x) = P(\mu - X \le x) \implies F(\mu+x|\mu,\sigma) = 1 - F(\mu-x|\mu,\sigma)\;\forall\; x\in\mathbb{R}.$$

In particular,

$$\Phi(0) = \frac{1}{2} \text{ and } \Phi(-z)+\Phi(z) = 1\;\forall\; z\in\mathbb{R}.$$

The p.d.f. $f(x|\mu,\sigma)\uparrow$ in $(-\infty,\mu)$ and $\downarrow$ in (μ,∞) $\implies$ mode = $m_0 = \mu$. Thus mean = median = mode = μ.

Let $X\sim N(\mu,\sigma^2)$. Then m.g.f. of X is

$$\begin{aligned}M_X(t) = E(e^{tX}) &= \int_{-\infty}^\infty e^{tx}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\mathrm{d}x,\;\text{ take } \frac{x-\mu}{\sigma} = z,\;\; x = (\mu+\sigma z) \\ &= \int_{-\infty}^\infty e^{t(\mu+\sigma z)}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\mathrm{d}z \\ &= e^{\mu t+\frac{1}{2}\sigma^2t^2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z^2-2\sigma tz+\sigma^2t^2)}\mathrm{d}z \\ &= e^{\mu t+\frac{1}{2}\sigma^2t^2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-\sigma t)^2}\mathrm{d}z = e^{\mu t+\frac{1}{2}\sigma^2t^2},\quad t\in\mathbb{R}.\end{aligned}$$

· Let $X \sim N(\mu, \sigma^2)$. Then $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$.

$$M_Z(t) = E(e^{t\frac{(X-\mu)}{\sigma}}) = e^{-\mu t/\sigma} M_X(t/\sigma) = e^{-\mu t/\sigma} e^{\mu t/\sigma + t^2/2} = e^{t^2/2} \ \forall\, t \in \mathbb{R} \rightarrow \text{m.g.f. of } N(0,1)$$
$$\implies Z \sim N(0,1).$$

Theorem 0.31. *Let $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0,1)$.*

(a) For $a \neq 0$, $b \in \mathbb{R}$, $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$.

(b) $Z \stackrel{d}{=} \frac{X-\mu}{\sigma} \sim N(0,1)$.

(c)
$$E(Z^r) = \begin{cases} 0, & \text{if } r = 1,3,5,\dots, \\ \frac{r!}{2^{r/2}(r/2)!}, & \text{if } r = 2,4,6,\dots. \end{cases}$$

(d) Mean = $\mu_1' = E(X) = \mu$; Variance = $\mu_2 = \sigma^2$; coefficient of skewness = $\beta_1 = 0$; kurtosis = $\nu_1 = 3$.

(e) $Z^2 \sim \chi^2_1$.

Proof. (a) Note that

$$\begin{aligned} M_Y(t) = E(e^{tY}) &= E\left(e^{t(aX+b)}\right) \\ &= e^{bt} E\left(e^{(ta)X}\right) = e^{bt} M_X(at) \\ &= e^{bt} e^{\mu(at) + \frac{1}{2}\sigma^2 (at)^2} \\ &= e^{(a\mu+b)t + \frac{1}{2}(a^2\sigma^2)t^2}, \ \ t \in \mathbb{R} \implies Y \sim N(a\mu + b, a^2\sigma^2). \end{aligned}$$

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

(c) $M_Z(t) = e^{t^2/2} = \sum_{k=0}^{\infty} \frac{t^{2k}}{2^k k!}$, $t \in \mathbb{R}$.

$$E(Z^r) = \text{ Coefficient of } \frac{t^r}{r!} \text{ in the expansion of } M_Z(t) = \begin{cases} 0, & \text{if } r = 1,3,5,\dots, \\ \frac{r!}{2^{r/2}(r/2)!}, & \text{if } r = 2,4,6,\dots. \end{cases}$$

(d) $Z \stackrel{d}{=} \frac{X-\mu}{\sigma}$.

$$\begin{aligned} E\left(\frac{X-\mu}{\sigma}\right) &= E(Z) = 0 \implies \mu_1' = E(X) = \mu, \\ E\left(\left(\frac{X-\mu}{\sigma}\right)^2\right) &= E(Z^2) = 1 \implies \mu_2 = E((X-\mu)^2) = \sigma^2, \\ E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right) &= E(Z^3) = 0 \implies \mu_3 = E((X-\mu)^3) = 0, \\ E\left(\left(\frac{X-\mu}{\sigma}\right)^4\right) &= E(Z^4) = 3 \implies \mu_4 = 3\sigma^4 = 3, \end{aligned}$$
$$\text{Coefficient of skewness} = \beta_1 = \frac{\mu_3}{\mu_2^2} = 0, \ \ \text{kurtosis} = \frac{\mu_4}{\mu_2^2} = 3.$$

(e) Let $Y = Z^2$. Then

$$M_Y(t) = E(e^{tZ^2}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz^2} e^{-z^2/2} \mathrm{d}z = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(1-2t)}{2} z^2} \mathrm{d}z = (1-2t)^{-1/2},\ t < \frac{1}{2} \implies Z^2 \sim \chi_1^2.$$

This completes the proof. $\square$

Corollary 0.32. *Let $X_1, X_2, \dots, X_k$ be independent and let $X_i \sim N(\mu_i, \sigma_i^2)$, $-\infty < \mu_i < \infty$, $\sigma_i > 0$, $i = 1, 2, \dots, k$. Then $\sum_{i=1}^k \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_k^2$.*

Remark 0.33. *(i) In $N(\mu, \sigma^2)$ distribution the parameters $\mu \in (-\infty, \infty)$ and $\sigma^2 > 0$ are respectively, the mean and variance of the distribution.*

(ii) If $X \sim N(\mu, \sigma^2)$, then

$$P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right),\ \ x \in \mathbb{R}.$$

Let τ_α be the $(1-\alpha)$th quantile of Φ then $\Phi(-\tau_\alpha) = 1 - \Phi(\tau_\alpha) = \alpha$. Tables for values of $\Phi(x)$ for different values of x are available in various text books.

Example 0.34. *Let $X \sim N(2, 4)$. Find $P(X \leq 0)$, $P(|X| \geq 2)$, $P(1 < X \leq 3)$ and $P(X \leq 3 | X > 1)$.*

Solution: $P(X \leq 0) = \Phi\left(\frac{0-2}{2}\right) = \Phi(-1) = 1 - \Phi(1) = .1587$,

$$\begin{aligned} P(|X| \geq 2) = P(X \leq -2) + P(X \geq 2) &= \Phi\left(\frac{-2-2}{2}\right) + 1 - \Phi\left(\frac{2-2}{2}\right) \\ &= \Phi(-2) + 1 - \Phi(0) = 0.0228 + 0.5 = 0.5228, \end{aligned}$$

$$P(1 < X \leq 3) = P(X \leq 3) - P(X \geq 1) = \Phi\left(\frac{3-2}{2}\right) + 1 - \Phi\left(\frac{1-2}{2}\right) = 2\Phi(0.5) - 1 = 2 \times 0.6915 - 1 = 0.383,$$

$$P(X \leq 3 | X > 1) = \frac{P(1 < X \leq 3)}{P(X > 1)} = \frac{.383}{1 - \Phi\left(\frac{1-2}{2}\right)} = \frac{0.383}{\Phi(0.5)} = 0.55599.$$

Theorem 0.35. *Let $X_1, X_2, \dots, X_k$ be independent r.v.'s and let $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, k$. Let $a_1, a_2, \dots, a_k$ be real constants such that $\sum_{i=1}^k a_i^2 > 0$. Then $Y = \sum_{i=1}^k a_i X_i \sim N(\sum_{i=1}^k a_i \mu_i, \sum_{i=1}^k a_i^2 \sigma_i^2)$.*

Proof. Note that

$$\begin{aligned} M_Y(t) = E(e^{t\sum_{i=1}^k a_i X_i}) = E\left(\prod_{i=1}^k e^{ta_i X_i}\right) &= \prod_{i=1}^k E(e^{ta_i X_i}), \quad \text{(independent of } X_i\text{'s)} \\ &= \prod_{i=1}^k M_{X_i}(ta_i) = \prod_{i=1}^k e^{\mu_i t a_i + \frac{1}{2}\sigma_i^2 t^2 a_i^2} = e^{(\sum_{i=1}^k a_i \mu_i)t + \frac{(\sum_{i=1}^k a_i^2 \sigma_i^2)t}{2}} \\ &\to \text{m.g.f. of } N\left(\sum_{i=1}^k a_i \mu_i, \sum_{i=1}^k a_i^2 \sigma_i^2\right). \end{aligned}$$

By uniqueness of m.g.f.'s $Y \sim N(\sum_{i=1}^k a_i \mu_i, \sum_{i=1}^k a_i^2 \sigma_i^2)$. $\square$

0.3. Random Vectors and their Distribution Functions

Let $(\Omega, \mathcal{F}, P)$ be a given probability space. In many situations we may be interested in simultaneously studying two or more numerical characteristics of outcomes of a random experiment. This amounts to define a function

$$\underline{X} = (X_1, X_2, \ldots, X_p) : \Omega \to \mathbb{R}^p.$$

Example 0.36. *A fair coin is tossed three times independently. Then*

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \text{ and } P(\{\omega\}) = \frac{1}{8}\ \forall\, \omega \in \Omega.$$

Suppose that we are simultaneously interested in:

· number of heads in three tosses,

· number of heads in first two tosses.

Here we are interested in the function $(X, Y) : \Omega \to \mathbb{R}^2$ *defined by*

$$(X(\omega), Y(\omega)) = \begin{cases} (0,0) & \text{if } \omega = TTT, \\ (1,0) & \text{if } \omega = TTH, \\ (1,1) & \text{if } \omega = HTT, THT, \\ (2,1) & \text{if } \omega = HTH, THH, \\ (2,2) & \text{if } \omega = HHT, \\ (3,2) & \text{if } \omega = HHH. \end{cases}$$

The values assumed by (X, Y) *are random with*

$$\Pr\{(X,Y) = (x,y)\} = \begin{cases} \frac{1}{8}, & \text{if } (x,y) \in \{(0,0),(1,0),(2,2),(3,2)\}, \\ \frac{1}{4}, & \text{if } (x,y) \in \{(1,1),(2,1)\}, \\ 0, & \text{otherwise}. \end{cases}$$

Here, $\Pr((X,Y) \in \{(0,0),(1,0),(2,2),(3,2),(1,1),(2,1)\}) = 1.$

Definition 0.37. *Let* $(\Omega, \mathcal{F}, P)$ *be a given probability space. A function* $\underline{X} = (X_1, X_2, \ldots, X_p) : \Omega \to \mathbb{R}^p$ *(defined on the sample space* Ω*) is called a random vector (p-dimensional random vector). A one dimensional random vector is simply called a random variable.*

For any function $\underline{Y} = (Y_1, Y_2, \ldots, Y_p) : \Omega \to \mathbb{R}^p$ and $A \subseteq \mathbb{R}^p$, define $\underline{Y}^{-1} = \{\omega \in \Omega : \underline{Y}(\omega) \in A\}$. For probability space $(\Omega, \mathcal{F}, P)$ and a p-dimensional random vector $\underline{X} = (X_1, X_2, \ldots, X_p) : \Omega \to \mathbb{R}^p$, define $P_{\underline{X}}(B) = P(\underline{X}^{-1}(B))$, $B \in \mathscr{B}_p$ where for all practical purpose we take $\mathscr{B}_p$ to be power set of $\mathbb{R}^p$. We will simply write

$$P_{\underline{X}}(B) = P(\{\omega \in \Omega : \underline{X}(\omega) \in B\}) = \Pr(X \in B), \ \ B \in \mathscr{B}_p.$$

The following scenario has emerged: $(\Omega, \mathcal{F}, P) \xrightarrow{\underline{X}} (\mathbb{R}^p, \mathscr{B}_p, P_{\underline{X}})$.

Theorem 0.38. $(\mathbb{R}^p, \mathscr{B}_p, P_{\underline{X}})$ *defined above is a probability space, i.e.* $P_{\underline{X}}(\cdot)$ *is a probability function defined on* $\mathscr{B}_p$.

Proof. Similar to the proof of random variable case. $\square$

Definition 0.39. *The probability function $P_{\underline{X}}(\cdot)$ defined above is called the probability function / measure induced by random vector $\underline{X}$ and $(\mathbb{R}^p, \mathscr{B}_p, P_{\underline{X}})$ is called the probability space induced by random vector $\underline{X}$.*

The induced probability measure $P_{\underline{X}}(\cdot)$ describes the random behaviour of $\underline{X}$.

Example 0.40. *Consider the sample space defined in Example 0.36, where*

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\} \text{ and } P(\{\omega\}) = \frac{1}{8}\ \forall\, \omega \in \Omega$$

and $(X, Y) : \Omega \to \mathbb{R}^2$ is defined by

$$(X(\omega), Y(\omega)) = \begin{cases} (0,0) & \text{if } \omega = TTT, \\ (1,0) & \text{if } \omega = TTH, \\ (1,1) & \text{if } \omega = HTT, THT, \\ (2,1) & \text{if } \omega = HTH, THH, \\ (2,2) & \text{if } \omega = HHT, \\ (3,2) & \text{if } \omega = HHH. \end{cases}$$

Here, $(X, Y) : \Omega \to \mathbb{R}^2$ is a random vector with induced probability space $(\mathbb{R}^2, \mathscr{B}_2, P_{\underline{X}})$, where

$$P_{\underline{X}}(\{(i,j)\}) = \begin{cases} \frac{1}{8}, & \text{if } (i,j) \in \{(0,0), (1,0), (2,2), (3,2)\}, \\ \frac{1}{4}, & \text{if } (i,j) \in \{(1,1), (2,1)\}, \\ 0, & \text{otherwise}, \end{cases}$$

and for any $B \in \mathscr{B}_2$

$$P_X(B) = \sum_{(i,j) \in B \cap S} P_X(\{(i,j)\}), \text{ where } S = \{(0,0), (1,0), (2,2), (3,2), (1,1), (2,1)\}.$$

Definition 0.41. *(a) The joint distribution function of a p-dimensional random vector $\underline{X} = (X_1, X_2, \ldots, X_p)$ is defined as*

$$F_{\underline{X}}(x_1, x_2, \ldots, x_p) = \Pr(X_1 \le x_1, X_2 \le x_2, \ldots, X_p \le x_p),\ \ \underline{x} = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p.$$

(b) The joint d.f. of any subset of random vectors $(X_1, X_2, \ldots, X_p)$ is called a marginal distribution function of $F_{\underline{X}}(\cdot)$ (or $\underline{X} = (X_1, X_2, \ldots, X_p)$).

Example 0.42. *$F_{X_1,X_2}(x,y)$, $(x,y) \in \mathbb{R}^2$, $F_{X_2}(x)$, $x \in \mathbb{R}$ and $F_{X_1,X_2,X_3}(x,y,z)$, $(x,y,z) \in \mathbb{R}^3$ are marginal d.f.s of $F_{X_1,X_2,X_3,X_4}(x_1,x_2,x_3,x_4)$, $(x_1,x_2,x_3,x_4) \in \mathbb{R}^4$.*

In the sequel we will describe a notation for writing down all the vertices of a p-dimensional rectangle in a compact form.

For $-\infty \le a_i < b_i < \infty,\ \ i = 1, 2,\ \ \underline{a} = (a_1, a_2)$ and $\underline{b} = (b_1, b_2)$, the vertices of two dimensional rectangle

$$(\underline{a}, \underline{b}] = (a_1, b_1] \times (a_2, b_2] = \{(x,y) \in \mathbb{R}^2 : a_1 < x \le b_1,\ a_2 < y \le b_2\}$$

are

$$\{(b_1,b_2), (a_1,b_2), (b_1,a_2), (a_1,a_2)\} = \{(b_1,b_2)\} \cup \{(a_1,b_2), (b_1,a_2)\} \cup \{(a_1,a_2)\} = \Delta_{0,2} \cup \Delta_{1,2} \cup \Delta_{2,2},\ \text{say}.$$

Similarly, for $-\infty \le a_i < b_i < \infty$, $i = 1, 2, 3$, $\underline{a} = (a_1, a_2, a_3)$ *and* $\underline{b} = (b_1, b_2, b_3)$, *the vertices of three dimensional rectangle*

$$(\underline{a}, \underline{b}] = (a_1, b_1] \times (a_2, b_2] \times (a_3, b_3] = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : a_i < x_i \le b_i,\ i = 1, 2, 3\}$$

are

$$\begin{aligned}
&\{(b_1, b_2, b_3), (a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3), (a_1, a_2, b_3), (a_1, b_2, a_3), (b_1, a_2, a_3), (a_1, a_2, a_3)\} \\
&\quad = \{(b_1, b_2, b_3)\} \cup \{(a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3)\} \cup \{(a_1, a_2, b_3), (a_1, b_2, a_3), (b_1, a_2, a_3)\} \cup \{(a_1, a_2, a_3)\} \\
&\quad = \Delta_{0,3} \cup \Delta_{1,3} \cup \Delta_{2,3} \cup \Delta_{3,3},\ \textit{say.}
\end{aligned}$$

In general, for $-\infty \le a_i < b_i < \infty$, $i = 1, 2, \ldots, p$, $\underline{a} = (a_1, a_2, \ldots, a_p)$ *and* $\underline{b} = (b_1, b_2, \ldots, b_p)$ *define*

$$\Delta_{k,p} \equiv \Delta_{k,p}\left((\underline{a}, \underline{b}]\right) = \{\underline{z} \in \mathbb{R}^p : z_i \in \{a_i, b_i\},\ i = 1, 2, \ldots, p \text{ and exactly } k \text{ of } z_i's \text{ are } a_j's\} \ \left(\to \text{has } \binom{p}{k} \text{ elements}\right)$$

where $(\underline{a}, \underline{b}] = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_p, b_p]$.

Then $\bigcup_{k=0}^{p} \Delta_{k,p}$ *is the set of* $2^p \left(= \sum_{k=0}^{p} \binom{p}{k}\right)$ *vectors of* p*-dimensional rectangle* $(\underline{a}, \underline{b}]$.

Theorem 0.43. *For constants* $-\infty \le a_i < b_i < \infty$, $i = 1, 2, \ldots, p$

$$\Pr(a_i < X_i \le b_i,\ i = 1, 2, \ldots, p) = \sum_{k=0}^{p} (-1)^k \sum_{\underline{z} \in \Delta_{k,p}((\underline{a}, \underline{b}])} F_{\underline{X}}(\underline{z}).$$

Proof. **Special cases:**

<u>Case I: $p = 1$</u>

We have $\Delta_{0,1}((a_1, b_1]) = \{b_1\}$ and $\Delta_{1,1}((a_1, b_1]) = \{a_1\}$. Then

$$R.H.S. = F_{X_1}(b_1) - F_{X_1}(a_1) = \Pr(a_1 < X_1 \le b_1) = L.H.S.$$

<u>Case II: $p = 2$</u>

Here $\Delta_{0,2} = \{(b_1, b_2)\}$, $\Delta_{1,2} = \{(a_1, b_2), (b_1, a_2)\}$ and $\Delta_{2,2} = \{(a_1, a_2)\}$. Thus

$$\begin{aligned}
R.H.S. &= F_{\underline{X}}(b_1, b_2) - F_{\underline{X}}(a_1, b_2) - F_{\underline{X}}(b_1, a_2) + F_{\underline{X}}(a_1, a_2) \\
&= \Pr(X_1 \le b_1,\ X_2 \le b_2) - \Pr(X_1 \le a_1,\ X_2 \le b_2) - \Pr(X_1 \le b_1,\ X_2 \le a_2) + \Pr(X_1 \le a_1,\ X_2 \le a_2) \\
&= \Pr(a_1 < X_1 \le b_1,\ X_2 \le b_2) - \Pr(a_1 < X_1 \le b_1,\ X_2 \le a_2) \\
&= \Pr(a_1 < X_1 \le b_1,\ a_2 < X_2 \le b_2) = L.H.S.
\end{aligned}$$

<u>Case III: $p = 3$</u>

$$\begin{aligned}
&\Pr(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, a_3 < X_3 \le b_3) \\
&= \Pr(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, X_3 \le b_3) - \Pr(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, X_3 \le a_3) \\
&= \Pr(a_1 < X_1 \le b_1, X_2 \le b_2, X_3 \le b_3) - \Pr(a_1 < X_1 \le b_1, X_2 \le a_2, X_3 \le b_3) \\
&\quad - \{\Pr(a_1 < X_1 \le b_1, X_2 \le b_2, X_3 \le a_3) + \Pr(a_1 < X_1 \le b_1, X_2 \le a_2, X_3 \le a_3)\} \\
&= \Pr(X_1 \le b_1, X_2 \le b_2, X_3 \le b_3) - \Pr(X_1 \le a_1, X_2 \le b_2, X_3 \le b_3) - \Pr(X_1 \le b_1, X_2 \le a_2, X_3 \le b_3) \\
&\quad + \Pr(X_1 \le a_1, X_2 \le a_2, X_3 \le b_3) - \Pr(X_1 \le b_1, X_2 \le b_2, X_3 \le a_3) + \Pr(X_1 \le a_1, X_2 \le b_2, X_3 \le a_3) \\
&\quad + \Pr(X_1 \le b_1, X_2 \le a_2, X_3 \le a_3) - \Pr(X_1 \le a_1, X_2 \le a_2, X_3 \le a_3) \\
&= F_{\underline{X}}(b_1, b_2, b_3) - F_{\underline{X}}(a_1, b_2, b_3) - F_{\underline{X}}(b_1, a_2, b_3) + F_{\underline{X}}(a_1, a_2, b_3) - F_{\underline{X}}(b_1, b_2, a_3) + F_{\underline{X}}(a_1, b_2, a_3)
\end{aligned}$$

$$+ F_{\underline{X}}(b_1, a_2, a_3) - F_{\underline{X}}(a_1, a_2, a_3) = \sum_{k=0}^{3}(-1)^k \sum_{\underline{z} \in \Delta_{k,3}([\underline{a},\underline{b}])} F_{\underline{X}}(\underline{z}).$$

The proof can be completed using method of induction. □

The following theorem provides a technique to find marginal distributions.

Theorem 0.44. *Let $F(x_1, x_2, \ldots, x_p)$, $\underline{x} = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p$ be a d.f. of p-dimensional random vector $\underline{X} = (X_1, X_2, \ldots, X_p)$. Then the marginal distribution function of $\underline{Y} = (X_1, X_2, \ldots, X_{p-1})$ is*

$$G(x_1, x_2, \ldots, x_{p-1}) = \lim_{t \to \infty} F(x_1, x_2, \ldots, x_{p-1}, t), \;\; \underline{y} = (x_1, x_2, \ldots, x_{p-1}) \in \mathbb{R}^{p-1}.$$

Proof. For $\underline{y} = (x_1, x_2, \ldots, x_{p-1}) \in \mathbb{R}^{p-1}$

$$\begin{aligned} G(x_1, x_2, \ldots, x_{p-1}) &= \Pr(X_1 \le x_1, \ldots, X_{p-1} \le x_{p-1}) \\ &= \Pr(X_1 \le x_1, \ldots, X_{p-1} \le x_{p-1}, X_p < \infty) \\ &= \Pr\left(\bigcup_{t=1}^{\infty}\{X_1 \le x_1, \ldots, X_{p-1} \le x_{p-1}, X_p \le t\}\right) \\ &= \lim_{t \to \infty} \Pr(X_1 \le x_1, \ldots, X_{p-1} \le x_{p-1}, X_p \le t) = \lim_{t \to \infty} F(x_1, x_2, \ldots, x_{p-1}, t). \end{aligned}$$

This completes the proof. □

Theorem 0.45. *Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with d.f. $F(\cdot)$. Then*

(a) $\lim_{\substack{x_i \to \infty \\ i=1,2,\ldots,p}} F(x_1, x_2, \ldots, x_p) = 1,$

(b) for each $i = 1, 2, \ldots, p$, $\lim_{x_i \to -\infty} F(x_1, x_2, \ldots, x_p) = 0,$

(c) $F(\underline{x})$ is right continuous in each argument (keeping other arguments fixed),

(d) for each rectangle $(\underline{a}, \underline{b}] \subseteq \mathbb{R}^p$

$$\sum_{k=0}^{p}(-1)^k \sum_{\underline{z} \in \Delta_{k,p}((\underline{a},\underline{b}])} F(\underline{z}) \ge 0.$$

Conversely, any function $G : \mathbb{R}^p \to [0, 1]$ satisfying conditions $(a) - (d)$ above is a d.f. of some p-dimensional random vector.

Proof. For simplicity, we provide the proof for $p = 2$.

(a) Note that

$$\begin{aligned} \lim_{x_1 \to \infty, x_2 \to \infty} F(x_1, x_2) &= \lim_{x_1 \to \infty, x_2 \to \infty} \Pr(\{X_1 \le x_1, X_2 \le x_2\}) \\ &= \lim_{n \to \infty} \Pr(\{X_1 \le n, X_2 \le n\}), \quad \text{(since limit exists)} \\ &= \Pr(\bigcup_{n=1}^{\infty}\{X_1 \le n, X_2 \le n\}) = \Pr(\{X_1 < \infty, X_2 < \infty\}) = 1. \end{aligned}$$

(b) For fixed $x_2 \in \mathbb{R}$,

$$\lim_{x_1 \to -\infty} F(x_1, x_2) = \lim_{n \to \infty} \Pr(\{X_1 \le -n, X_2 \le x_2\})$$

$$= \Pr(\bigcap_{n=1}^{\infty}\{X_1 \leq -n, X_2 \leq x_2\}) = \Pr(\phi) = 0.$$

Similarly, $\lim_{x_2 \to -\infty} F(x_1, x_2) = 0$.

(c) Let $\{h_n\}_{n\geq 1}$ be a sequence in $\mathbb{R}$ such that $h_n \downarrow 0$. Then for $(x_1, x_2) \in \mathbb{R}^2$

$$\begin{aligned}\lim_{n\to\infty} F(x_1 + h_n, x_2) &= \lim_{n\to\infty} \Pr(\{X_1 \leq x_1 + h_n, X_2 \leq x_2\}) \\ &= \lim_{n\to\infty} \Pr\left(\left\{X_1 \leq x_1 + \frac{1}{n}, X_2 \leq x_2\right\}\right), \quad \text{(as limit exists)} \\ &= \Pr\left(\bigcap_{n=1}^{\infty}\left\{X_1 \leq x_1 + \frac{1}{n}, X_2 \leq x_2\right\}\right) = \Pr(\{X_1 \leq x_1, X_2 \leq x_2\}) = F(x_1, x_2),\end{aligned}$$

i.e. for every fixed $x_2 \in \mathbb{R}$, $F(x_1, x_2)$ is right continuous in $x_1 \in \mathbb{R}$. Similarly, it can be shown that for every fixed $x_1 \in \mathbb{R}$, $F(x_1, x_2)$ is right continuous in $x_2 \in \mathbb{R}$.

(d) For $-\infty < a_1 < b_1 < \infty$ and $-\infty < a_2 < b_2 < \infty$, we have

$$\begin{aligned}\sum_{k=0}^{2}(-1)^k \sum_{\underline{z}\in\Delta_{k,p}((\underline{a},\underline{b}])} F(\underline{z}) &= F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \\ &= P(a_1 < X_1 \leq b_1,\ a_2 < X_2 \leq b_2) \geq 0.\end{aligned}$$

This completes the proof. □

Remark 0.46. *(a) For $p = 1$, (d) of the above theorem reduces to $F(b) - F(a) \geq 0, \forall -\infty < a < b < \infty$, i.e., $F(\cdot)$ is monotone on $\mathbb{R}$.*

(b) $F(\cdot)$ is clearly non-decreasing in each argument.

0.3.1. Independent Random Variables

For an arbitary (countable or uncountable) set Δ, let $\{X_\lambda : \lambda \in \Delta\}$ be a family of random variables.

Definition 0.47. *The random variables $X_\lambda, \lambda \in \Delta$ are said to be mutually independent if for any finite subcollection $\{X_{\lambda_1}, X_{\lambda_2}, \ldots, X_{\lambda_p}\}$ in $\{X_\lambda : \lambda \in \Delta\}$*

$$F_{\lambda_1,\lambda_2,\ldots,\lambda_p}(x_1, x_2, \ldots, x_p) = \prod_{i=1}^{p} F_{\lambda_i}(x_i)\ \forall\ \underline{x} = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p,$$

where $F_{\lambda_1,\lambda_2,\ldots,\lambda_p}(\cdot)$ denotes the joint d.f. of $(X_{\lambda_1}, X_{\lambda_2}, \ldots, X_{\lambda_p})$ and $F_{\lambda_i}(\cdot),\ \ i = 1, 2, \ldots, p$ denotes the marginal d.f. of X_{λ_i}.

The random variables $X_\lambda, \lambda \in \Delta$ are said to be pairwise independent if for any $\lambda_1, \lambda_2 \in \Delta$ $(\lambda_1 \neq \lambda_2)$

$$F_{\lambda_1,\lambda_2}(x_1, x_2) = F_{\lambda_1}(x_1)F_{\lambda_2}(x_2)\ \forall\ \underline{x} = (x_1, x_2) \in \mathbb{R}^2.$$

Remark 0.48. *(a) Random variables $\{X_\lambda, \lambda \in \Delta\}$ are independent iff those in any finite subset of $\{X_\lambda : \lambda \in \Delta\}$ are independent.*

(b) Let $\Delta_1 \subseteq \Delta_2$. Then r.v.s $\{X_\lambda, \lambda \in \Delta_2\}$ are independent $\implies$ r.v.s $\{X_\lambda, \lambda \in \Delta_1\}$ are independent. In particular, if r.v.s in a collection are independent then they are pairwise independent. The converse may not be true.

Theorem 0.49. *For any positive integer p (≥ 2) the random variables $X_1, X_2, \ldots, X_p$ are independent iff*

$$F(x_1, x_2, \ldots, x_p) = \prod_{i=1}^{p} F_{X_i}(x_i) \;\forall\; \underline{x} = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p, \tag{0.9}$$

where $F(\cdot)$ is the joint d.f. of $\underline{X} = (X_1, X_2, \ldots, X_p)$.

Proof. Obviously, if $X_1, X_2, \ldots, X_p$ are independent then (0.9) holds. Conversely suppose that (0.9) holds. Consider a subset of $\{X_1, X_2, \ldots, X_p\}$. For simplicity let this subset be $\{X_1, X_2, \ldots, X_q\}$, for some $2 \leq q \leq p$. Thus for $\underline{x} = (x_1, x_2, \ldots, x_q) \in \mathbb{R}^q$ the joint (marginal) d.f. of $(X_1, X_2, \ldots, X_q)$ is

$$G(x_1, x_2, \ldots, x_q) = \lim_{\substack{x_i \to \infty \\ i=q+1,\ldots,p}} F(x_1, x_2, \ldots, x_q, x_{q+1}, \ldots, x_p) = \lim_{\substack{x_i \to \infty \\ i=q+1,\ldots,p}} \prod_{j=1}^{p} F_{X_j}(x_j) = \prod_{j=1}^{q} F_{X_j}(x_j),$$

$\forall\; \underline{x} = (x_1, x_2, \ldots, x_q) \in \mathbb{R}^q$. Here $F_{X_j}(\cdot)$ is the marginal d.f. of X_j, $j = 1, 2, \ldots, q$. $\square$

0.3.2. Discrete Random Vectors

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be p-dimensional random vector with d.f. $F(\cdot)$.

Definition 0.50. *(a) The random vector $\underline{X} = (X_1, X_2, \ldots, X_p)$ is said to be a discrete random vector if there exists a countable set S (finite or infinite) such that*

$$\Pr(\underline{X} = \underline{x}) > 0 \;\forall\; \underline{x} \in S, \;\; \textit{and} \;\; \Pr(\underline{X} \in S) = 1.$$

The set S is called support of random vector $\underline{X}$ (or of F).

(b) The joint p.m.f. of $\underline{X}$ having support S is defined by

$$f(\underline{x}) = \begin{cases} \Pr(\underline{X} = \underline{x}), \; \textit{if } \underline{x} \in S, \\ 0, \; \textit{otherwise}. \end{cases}$$

Remark 0.51. *(a) Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional discrete random vector with p.m.f. $f(\cdot)$ and d.f. $F(\cdot)$ and support S. Then, for any $A \subseteq \mathbb{R}^p$*

$$\Pr(\underline{X} \in A) = \Pr(\underline{X} \in A \cap S) = \sum_{\underline{x} \in A \cap S} f(\underline{x}), \quad (\Pr(\underline{X} \in S) = 1, \, A \cap S \subseteq S \textit{ and thus } A \cap S \textit{ is a countable set}).$$

Moreover, $F(\underline{x}) = \sum_{\underline{y} \in S \cap (-\infty, \underline{x}]} f(\underline{y})$, $\underline{x} \in \mathbb{R}^p$.

(b) Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional discrete random vector with p.m.f. $f(\cdot)$ and support S. Then the p.m.f. $f : \mathbb{R}^p \to \mathbb{R}$ satisfies:

$$(i) \quad f(\underline{x}) > 0 \;\forall\; x \in S \;\; \textit{and} \; f(\underline{x}) = 0 \;\forall\; x \in S^c, \quad (ii) \; \sum_{\underline{x} \in S} f(\underline{x}) = 1.$$

Conversely suppose that $g : \mathbb{R}^p \to \mathbb{R}$ is a function such that for some countable set T

$$(i) \quad g(\underline{x}) > 0 \;\forall\; x \in T \;\; \textit{and} \; g(\underline{x}) = 0 \;\forall\; x \in T^c, \quad (ii) \; \sum_{\underline{x} \in T} g(\underline{x}) = 1.$$

Then $g(\cdot)$ is a p.m.f. of some p-dimensional discrete random vector having support T.

(c) Marginal distributions of discrete random vector are discrete.

Theorem 0.52. *Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional discrete random vector with p.m.f. $f(\cdot)$ and support S. Then the marginal distribution of any subset of $\{X_1, X_2, \ldots, X_p\}$ (say that of $\underline{Y} = (X_1, X_2, \ldots, X_q),\ 1 \le q < p$) is again discrete with p.m.f.*

$$g(x_1, x_2, \ldots, x_q) = \begin{cases} \sum_{x_{q+1}} \sum_{x_{q+2}} \cdots \sum_{x_p} f(\underline{x}), \text{ if } \underline{x} \in T, \\ 0, \quad \text{otherwise} \end{cases}$$

and support

$$T = \{\underline{y} = (y_1, y_2, \ldots, y_q) \in \mathbb{R}^q : (y_1, y_2, \ldots, y_q, y_{q+1}, \ldots, y_p) \in S, \text{ for some } (y_{q+1}, y_{q+2}, \ldots, y_p) \in \mathbb{R}^{p-q}\}.$$

Proof. Follows using theorem of total probability. □

Conditional distribution of discrete random vectors

Let $\underline{Y} = (Y_1, Y_2, \ldots, Y_p),\ \ \underline{Z} = (Z_1, Z_2, \ldots, Z_q)$ and $\underline{X} = (\underline{Y}, \underline{Z}) = (Y_1, Y_2, \ldots, Y_p, Z_1, Z_2, \ldots, Z_q)$ be random vectors with p.m.f. $f_1,\ f_2$ and f, respectively. Suppose $\underline{X},\ \underline{Y}$ and $\underline{Z}$ have support $S,\ S_1,\ S_2$, respectively. For fixed $\underline{z} \in S_2$ define

$$T_{\underline{z}} = \{\underline{y} = (y_1, y_2, \ldots, y_p) \in \mathbb{R}^p : (\underline{y}, \underline{z}) \in S\}.$$

For fixed $\underline{z} \in S_2$, the conditional p.m.f. of $\underline{Y}$ given $\underline{Z} = \underline{z}$ is defined by

$$f(\underline{y}|\underline{z}) = \Pr(\underline{Y} = \underline{y}|\underline{Z} = \underline{z}) = \frac{\Pr(\underline{X} = (\underline{y}, \underline{z}))}{\Pr(\underline{Z} = \underline{z})} = \begin{cases} \dfrac{f(\underline{y}, \underline{z})}{f_2(\underline{z})}, \ \underline{y} \in T_{\underline{z}}, \\ 0, \ \text{otherwise.} \end{cases}$$

Clearly for each $\underline{z} \in S_2$, $f(\cdot|\underline{z})$ is a proper p.m.f. with support $T_{\underline{z}}$. Also for fix $\underline{z} \in S_2$

$$\Pr(Y_1 \le y_1, \ldots, Y_q \le y_q | \underline{Z} = \underline{z}) = \frac{\Pr(Y_1 \le y_1, \ldots, Y_q \le y_q, \underline{Z} = \underline{z})}{\Pr(\underline{Z} = \underline{z})} = \sum_{\underline{s} \in T_{\underline{z}}, \underline{s} \le \underline{y}} \frac{f(\underline{s}, \underline{z})}{f_{\underline{Z}}(\underline{z})} = \sum_{\underline{s} \in T_{\underline{z}}, \underline{s} \le \underline{y}} f(\underline{s}|\underline{z}).$$

Theorem 0.53. *Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with support S and p.m.f. $f(\cdot)$. Let $f_i(\cdot)$ denote the marginal p.m.f. of $X_i,\ \ i = 1, 2, \ldots, p$. Then $X_1, \ldots, X_p$ are independent iff*

$$f(x_1, x_2, \ldots, x_p) = \prod_{i=1}^{p} f_i(x_i) \ \forall \ \underline{x} \in S.$$

Proof. (For $p = 2$)

Suppose that $f(x_1, x_2) = f_1(x_1) f_2(x_2) \ \forall \ \underline{x} = (x_1, x_2) \in S$. Then the d.f. of $\underline{X} = (X_1, X_2)$ is

$$F(x_1, x_2) = \sum_{\substack{(y_1, y_2) \in S \\ y_1 \le x_1, y_2 \le x_2}} \sum f(y_1, y_2) = \sum_{\substack{(y_1, y_2) \in S \\ y_1 \le x_1, y_2 \le x_2}} \sum f_1(y_1) f_2(y_2), \ \ (x_1, x_2) \in \mathbb{R}^2.$$

Let S_1 and S_2 be supports of X_1 and X_2 respectively. Then

$$\begin{aligned} S &= \{(y_1, y_2) \in \mathbb{R}^2 : f(y_1, y_2) > 0\} \\ &= \{(y_1, y_2) \in \mathbb{R}^2 : f_1(y_1) f_2(y_2) > 0\} \\ &= \{(y_1, y_2) \in \mathbb{R}^2 : f_1(y_1) > 0 \ \text{ and } \ f_2(y_2) > 0\} \\ &= \{y_1 \in \mathbb{R} : f_1(y_1) > 0\} \times \{y_2 \in \mathbb{R} : f_2(y_2) > 0\} = S_1 \times S_2. \end{aligned}$$

Therefore, for $(x_1, x_2) \in \mathbb{R}^2$

$$F(x_1, x_2) = \sum_{\substack{y_1 \in S_1, y_2 \in S_2 \\ y_1 \le x_1, y_2 \le x_2}} \sum f_1(y_1) f_2(y_2) = \left(\sum_{\substack{y_1 \in S_1 \\ y_1 \le x_1}} f_1(y_1) \right) \left(\sum_{\substack{y_2 \in S_2 \\ y_2 \le x_2}} f_2(y_2) \right) = F_1(x_1) F_2(x_2)$$

where F_1 and F_2 are marginal d.f.s of X_1 and X_2 respectively. This imples X_1 and X_2 are independent. Conversely, suppose that X_1 and X_2 are independent. Then $F(y_1, y_2) = F_1(y_1)F_2(y_2) \; \forall \; (y_1, y_2) \in \mathbb{R}^2$. Then, for $(x_1, x_2) \in \mathbb{R}^2$

$$\begin{aligned}
f(x_1, x_2) &= \Pr(X_1 = x_1, X_2 = x_2) \\
&= \Pr\left(\bigcap_{n=1}^{\infty} \left\{ x_1 - \frac{1}{n} < X_1 \le x_1, x_2 - \frac{1}{n} < X_2 \le x_2 \right\} \right) \\
&= \lim_{n \to \infty} \Pr\left(x_1 - \frac{1}{n} < X_1 \le x_1, x_2 - \frac{1}{n} < X_2 \le x_2 \right) \\
&= \lim_{n \to \infty} \left[F(x_1, x_2) - F\left(x_1 - \frac{1}{n}, x_2 \right) - F\left(x_1, x_2 - \frac{1}{n} \right) + F\left(x_1 - \frac{1}{n}, x_2 - \frac{1}{n} \right) \right] \\
&= \lim_{n \to \infty} \left[F_1(x_1)F_2(x_2) - F_1\left(x_1 - \frac{1}{n} \right) F_2(x_2) - F_1(x_1)F_2\left(x_2 - \frac{1}{n} \right) + F_1\left(x_1 - \frac{1}{n} \right) F_2\left(x_2 - \frac{1}{n} \right) \right] \\
&= F_1(x_1)F_2(x_2) - F_1(x_1-)F_2(x_2) - F_1(x_1)F_2(x_2-) + F_1(x_1-)F_2(x_2-) \\
&= (F_1(x_1) - F_1(x_1-))F_2(x_2) - (F_1(x_1) - F_1(x_1-))F_2(x_2-) \\
&= (F_1(x_1) - F_1(x_1-))(F_2(x_2) - F_2(x_2-)) = f_1(x_1)f_2(x_2).
\end{aligned}$$

This completes the proof for $p = 2$ case. Similary, it can be proved for other cases. ☐

Remark 0.54. *(a) If $\underline{X} = (X_1, X_2, \dots, X_p)$ is a discrete r.v. with support S and X_i has support S_i, $i = 1, 2, \dots, p$ then $X_1, X_2, \dots, X_p$ are independent $\implies S = S_1 \times S_2 \times \cdots \times S_p$.*

(b) Let $\underline{X} = (X_1, X_2, \dots, X_p)$ be a discrete random vector with support S and p.m.f. $f(\cdot)$. Then $X_1, \dots, X_p$ are independent iff

$$f(x_1, x_2, \dots, x_p) = g_1(x_1)g_2(x_2) \cdots g_p(x_p), \;\; \underline{x} = (x_1, x_2, \dots, x_p) \in \mathbb{R}^p$$

and $S = A_1 \times A_2 \times \cdots A_p$ for some functions $A_1, \dots, A_p$ and $A_i = \{x \in \mathbb{R} : g_i(x) > 0\}$, $i = 1, 2, \dots, p$. In that case the marginal p.m.f. of X_i is $f_i(x) = c_i g_i(x)$, $x \in \mathbb{R}$ for some constant c_i such that $\sum_{x \in A_i} c_i g_i(x) = 1$, $i = 1, 2, \dots, p$.

(c) If $\underline{X} = (Y, Z)$ is a two-dimensional r.v. then Y and Z are independent iff $f(y|z) = f_1(y) \; \forall \; y \in \mathbb{R}$ and $z \in \mathbb{R}$ such that $f_2(z) > 0$, here $f(y|z)$ denotes the conditional p.m.f. of Y given $Z = z$ and $f_1(\cdot)$ denotes the marginal p.d.f. of Y.

(d) One can extend Definition 0.47 to define indepedence of a collection of random vectors. Then analogous of Theorem 0.49, Remark 0.51, Theorem 0.52, Theorem 0.53 and (c) above holds for random vectors.

Example 0.55. *Let $\underline{X} = (X_1, X_2, X_3)$ have the joint p.m.f.*

$$f(x_1, x_2, x_3) = \begin{cases} c x_1 x_2 x_3, \; x_1 = 1, 2, \;\; x_2 = 1, 2, 3, \;\; x_3 = 1, 3, \\ 0, \; \textit{otherwise}. \end{cases}$$

where c is a real constant.

(a) Find the value of c.

(b) Find the marginal p.m.f.s of X_1, X_2 and X_3.

(c) Are X_1, X_2 and X_3 independent.

(d) Find marginal p.m.f. of (X_1, X_3).

(e) Find conditional p.m.f. of X_1 given $(X_2, X_3) = (2, 1)$.

(f) Are X_1 and X_3 independent.

(g) Compute $\Pr(X_1 = X_2 = X_3)$.

Solution: Here the support of random vector $\underline{X}$ is $S_{\underline{X}} = \{1, 2\} \times \{1, 2, 3\} \times \{1, 3\}$.

(a) $\sum_{\underline{x} \in S_{\underline{X}}} f(\underline{x}) = 1 \implies c(1+3+2+6+3+9+2+6+4+12+6+18) = 1 \implies c = \frac{1}{72}$. Clearly $f(\underline{x}) \geq 0$ $\forall$ $\underline{x} \in \mathbb{R}^3$.

(b) For $x_1 \neq \{1, 2\}$, clearly $f_{X_1}(x_1) = 0$. For $x_1 \in \{1, 2\}$

$$f_{X_1}(x_1) = \sum_{(x_2,x_3) \in \{1,2,3\} \times \{1,3\}} \frac{x_1 x_2 x_3}{72} = \frac{x_1}{72} \left(\sum_{x_2=1}^{3} x_2 \right) \left(\sum_{x_1=1,3} x_3 \right) = \frac{x_1}{3}.$$

Thus,

$$f_{X_1}(x_1) = \begin{cases} \dfrac{x_1}{3}, \ x_1 \in \{1, 2\}, \\ 0, \ \text{otherwise.} \end{cases}$$

Similarly,

$$f_{X_2}(x_2) = \begin{cases} \dfrac{x_2}{6}, \ x_2 \in 1, 2, 3, \\ 0, \ \text{otherwise.} \end{cases} ; \quad f_{X_3}(x_3) = \begin{cases} \dfrac{x_3}{4}, \ x_3 \in 1, 3, \\ 0, \ \text{otherwise.} \end{cases}$$

(c) Clearly $f(x_1, x_2, x_3) = g_1(x_1) g_2(x_2) g_3(x_3)$, $(x_1, x_2, x_3) \in \mathbb{R}^3$ and $S_X = \Delta_1 \times \Delta_2 \times \Delta_3$ where $\Delta_1 = \{1, 2\}$, $\Delta_2 = \{1, 2, 3\}$ and $\Delta_3 = \{1, 3\}$.

$$g_1(x_1) = \begin{cases} c_1 x_1, \ x_1 \in \Delta_1, \\ 0, \ \text{otherwise.} \end{cases} ; \quad g_2(x_2) = \begin{cases} c_2 x_2, \ x_2 \in \Delta_2, \\ 0, \ \text{otherwise.} \end{cases} ; \quad g_3(x_3) = \begin{cases} c_3 x_3, \ x_3 \in \Delta_3, \\ 0, \quad \text{otherwise.} \end{cases}$$

Obviously $c_1 = \frac{1}{3}$, $c_2 = \frac{1}{6}$ and $c_3 = \frac{1}{4}$. Thus X_1, X_2 and X_3 are independent.

Alternatively, using (b), we have $f(x_1, x_2, x_3) = f_{X_1}(x_1) f_{X_2}(x_2) f_{X_3}(x_3)$ $\forall$ $(x_1, x_2, x_3) \in \mathbb{R}^3$.

(d) Marginal of (X_1, X_3) is $f_{X_1, X_3}(x_1, x_3) = \sum_{x_2} f_{\underline{X}}(x_1, x_2, x_3) = \dfrac{x_1 x_3}{72} \times 6 = \dfrac{x_1 x_3}{12}$. Thus,

$$f_{X_1, X_3}(x_1, x_3) = \begin{cases} \dfrac{x_1 x_3}{12}, \ (x_1, x_3) \in \{1, 2\} \times \{1, 3\}, \\ 0, \ \text{otherwise} \end{cases}$$

(e) For $x_1 \in \{1, 2\}$,

$$\Pr(X_1 = x_1 | X_2 = 2, X_3 = 1) = \frac{\Pr(X_1 = x_1, X_2 = 2, X_3 = 1)}{\Pr(X_2 = 2, X_3 = 1)} = \frac{x_1 2}{72} \Big/ \frac{1}{12} = \frac{x_1}{3}.$$

Thus

$$f_{X_1|(X_2,X_3)}(x_1|(2,1)) = \begin{cases} \dfrac{x_1}{3}, \ x_1 \in \{1, 2\}, \\ 0, \ \text{otherwise.} \end{cases}$$

Alternatively, since X_1, X_2 and X_3 are independent and X_1 and (X_2, X_3) are independent (why!), thus for fixed $(x_2, x_3) \in \mathbb{R}^2$ such that $f_{X_2,X_3}(x_2, x_3) > 0$,

$$f_{X_1|(X_2,X_3)}(x_1|(x_2,x_3)) = f_{X_1}(x_1) \; \forall \; x_1 \in \mathbb{R}$$

$$\implies f_{X_1|(X_2,X_3)}(x_1|(2,1)) = \begin{cases} \dfrac{x_1}{3}, \; x_1 \in \{1,2\}, \\ 0, \; \text{otherwise.} \end{cases}$$

(f) By (c), X_1 and X_3 are independent.

(g) $\Pr(X_1 = X_2 = X_3) = \sum\limits_{\substack{\underline{x} \in S_X \\ x_1 = x_2 = x_3}} \dfrac{x_1 x_2 x_3}{72} = P(X_1 = X_2 = X_3 = 1) = \dfrac{1}{72}.$

0.3.3. Continuous Random Vectors

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with d.f. F.

Definition 0.56. *The random vector $\underline{X}$ is called a continuous random vector if there exists a non-negative function $f : \mathbb{R}^p \to \mathbb{R}$ such that for any rectangale set A in $\mathbb{R}^p$*

$$P(\underline{X} \in A) = \underset{A}{\int \int \ldots \int} f(\underline{t}) \mathrm{d}\underline{t},$$

where $\underline{t} = (t_1, t_2 \ldots, t_p)$ and $\mathrm{d}\underline{t} = \mathrm{d}t_1 \mathrm{d}t_2 \ldots \mathrm{d}t_p$. The function $f(\cdot)$ is called probability density function of $\underline{X}$ and the set

$$S = \{\underline{x} \in \mathbb{R}^p : \Pr(x_i - h_i < x_i \le x_i + h_i, \; i = 1, 2, \ldots, p) > 0 \; \forall \; h_i > 0, \;\; i = 1, 2, \ldots, p\}$$

is called the support of F (or of $\underline{X}$).

Remark 0.57. *(a) In particular, if for fixed $\underline{x} = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p$ if $A = (-\infty, x_1] \times (-\infty, x_2] \times \cdots \times (-\infty, x_p]$, then*

$$F(x_1, x_2, \ldots, x_p) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_p} f(t_1, t_2, \ldots, t_p) \mathrm{d}t_1 \mathrm{d}t_2 \ldots \mathrm{d}t_p.$$

(b) If $\underline{X}$ is continuous random vector then its d.f. F is a continuous function.

(c) For a continuous random vector if its p.d.f. $f(\underline{x})$ is a piecewise continuous function then from the fundamental theorem of multivariable calculus

$$f(\underline{x}) = \frac{\partial^p}{\partial x_1 \partial x_2 \ldots \partial x_p} F(\underline{x}), \;\; \underline{x} \in \mathbb{R}^p,$$

whenever the derivative is defined.

** If $f(\cdot)$ is continuous at $\underline{x} \in \mathbb{R}^p$, then*

$$f(\underline{x}) = \lim_{\substack{h_i \to 0 \\ i=1,2,\ldots,p}} \frac{1}{h_1 \cdots h_p} \int_{x_1}^{x_1+h} \cdots \int_{x_p}^{x_p+h_p} f(\underline{t}) \mathrm{d}\underline{x} = \lim_{\substack{h_i \to 0 \\ i=1,2,\ldots,p}} \frac{1}{h_1 \cdots h_p} \Pr(x_i < X_i \le x_i + h_i, \; i = 1, 2, \ldots, p).$$

For small $\mathrm{d}x_1, \ldots, \mathrm{d}x_p$ if f is continuous at $\underline{x}$, then

$$\Pr(x_i < X_i \le x_i + \mathrm{d}x_i, \; i = 1, 2, \ldots, p) = \int_{x_1}^{x_1+\mathrm{d}x_1} \cdots \int_{x_p}^{x_p+\mathrm{d}x_p} f(t_1, t_2, \ldots, t_p) \mathrm{d}t_p \cdots \mathrm{d}t_1$$

$$\approx \mathrm{d}x_1 \cdots \mathrm{d}x_p f(x_1, x_2, \ldots, x_p).$$

Thus the probability that $\underline{X}$ is in a small neighborhood of $\underline{x} = (x_1, x_2, \ldots, x_p)$ is proportional to $f(x_1, x_2, \ldots, x_p)$.

(d) There are random vectors that are neither discrete nor continuous.

(e) If $\underline{X}$ is a continuous random vector with p.d.f. $f(\cdot)$ then $P(\underline{X} = \underline{a}) = \int \underset{\underline{x}=\underline{a}}{\int \ldots} \int f(\underline{t})\mathrm{d}\underline{t} = 0$.

(f) As in the univariate case the p.d.f. of a continuous random vector is not unique and it has different versions.

(g) It can be shown that if $\underline{X}$ is a p-dimensional random vector with continuous d.f. $F(\cdot)$ such that

$$\frac{\partial^p}{\partial x_1 \partial x_2 \ldots \partial x_p} F(x_1, x_2, \ldots, x_p)$$

exists everywhere except (possibly) on a set C comprising of countable number of curves (having 0 volume in $\mathbb{R}^p$) and

$$\int_{\mathbb{R}^p \setminus C} \frac{\partial^p}{\partial x_1 \partial x_2 \ldots \partial x_p} F(x_1, x_2, \ldots, x_p) \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_p = 1.$$

Then $\underline{X}$ is a continuous random vector with p.d.f.

$$f(\underline{x}) = \begin{cases} \dfrac{\partial^p}{\partial x_1 \partial x_2 \ldots \partial x_p} F(x_1, x_2, \ldots, x_p), \ \textit{if } \underline{x} \in \mathbb{R}^p \setminus C, \\ 0, \ \textit{if } \underline{x} \in C. \end{cases}$$

(h) Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a continuous random vector with joint p.d.f. $f_{\underline{X}}(\underline{x})$ and d.f. $F_{\underline{X}}(\underline{x})$. Then for $q \in \{1, 2, \ldots, p-1\}$ and $\underline{x} = (x_1, x_2, \ldots, x_q) \in \mathbb{R}^q$,

$$\begin{aligned}
&F_{X_1, X_2, \ldots, X_q}(x_1, x_2, \ldots, x_q) \\
&= \lim_{\substack{x_j \to \infty \\ j=q+1,\ldots,p}} F_{X_1, X_2, \ldots, X_q, X_{q+1}, \ldots, X_p}(x_1, x_2, \ldots, x_q, x_{q+1}, \ldots, x_p) \\
&= \lim_{\substack{x_j \to \infty \\ j=q+1,\ldots,p}} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_q} \int_{-\infty}^{x_{q+1}} \cdots \int_{-\infty}^{x_p} f_{X_1, X_2, \ldots, X_q, X_{q+1}, \ldots, X_p}(t_1, t_2, \ldots, t_q, t_{q+1}, \ldots, t_p) \mathrm{d}t_p \cdots \mathrm{d}t_1 \\
&= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_q} \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, X_2, \ldots, X_q, X_{q+1}, \ldots, X_p}(t_1, t_2, \ldots, t_q, t_{q+1}, \ldots, t_p) \mathrm{d}t_p \cdots \mathrm{d}t_{q+1} \right] t_q \cdots \mathrm{d}t_1
\end{aligned}$$

$\implies$ $(X_1, X_2, \ldots, X_q)$ *is a continuous random vector with p.d.f.*

$$f_{X_1, X_2, \ldots, X_q}(x_1, x_2, \ldots, x_q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, X_2, \ldots, X_q, X_{q+1}, \ldots, X_p}(x_1, x_2, \ldots, x_q, t_{q+1}, \ldots, t_p) \mathrm{d}t_p \cdots \mathrm{d}t_{q+1}.$$

Thus, marginal distribution of a continuous random vector $\underline{X}$ are continuous with p.d.f. of marginal distribution obtained by integrating out unwanted variables in the p.d.f. of $\underline{X}$.

Conditional Distributions of Continuous Random Vector

For simplicity consider $p = 2$ and let $\underline{X} = (X_1, X_2)$ be a random vector (discrete or continuous) with d.f. $F_{X_1, X_2}(x_1, x_2)$. Suppose that, for $x_1 \in S_{X_1}$ (the support of distribution of X_1) we want to define conditional d.f. of X_2 given $X_1 = x_1$. If X_1 is a continuous random variable then $\Pr(X_1 = x_1) = 0 \, \forall \, x_1 \in \mathbb{R}$ and therefore $\Pr(X_2 \le x | X_1 = x_1)$ is not defined for any $x_1 \in \mathbb{R}$; although it is defined for discrete random vector X_1 when $x_1 \in S_{X_1}$. Thus we define the conditional random vector X_1 when $x_1 \in S_{X_1}$. Thus we define the conditional d.f. of X_2 given $X_1 = x_1$, through the limiting argument

$$F_{X_2|X_1}(x|x_1) = \lim_{h \downarrow 0} \Pr(X_2 \le x | x_1 - h < X_1 \le x_1)$$

$$= \lim_{h \downarrow 0} \frac{\Pr(X_2 \le x, x_1 - h < X_1 \le x_1)}{\Pr(x_1 - h < X_1 \le x_1)} = \lim_{h \downarrow 0} \frac{F_{X_1,X_2}(x_1, x) - F_{X_1,X_2}(x_1 - h, x)}{F_{X_1}(x_1) - F_{X_1}(x_1 - h)}.$$

Clearly if $\underline{X} = (X_1, X_2)$ is discrete and $x_1 \in S_{X_1}$, then

$$F_{X_2|X_1}(x|x_1) = \lim_{h \downarrow 0} \frac{F_{X_1,X_2}(x_1, x) - F_{X_1,X_2}(x_1 - h, x)}{F_{X_1}(x_1) - F_{X_1}(x_1 - h)} = \frac{\Pr(X_1 = x_1, X_2 \le x)}{\Pr(X_1 = x_1)} = \Pr(X_2 \le x | X_1 = x_1).$$

Also, if $\underline{X} = (X_1, X_2)$ is continuous random vector with p.d.f. $f(x_1, x_2)$ then

$$F_{X_2|X_1}(x|x_1) = \lim_{h \downarrow 0} \frac{\frac{1}{h}\int_{-\infty}^{x}\int_{x_1-h}^{x_1} f_{X_1,X_2}(y_1, y_2)\mathrm{d}y_1\mathrm{d}y_2}{\frac{F_{X_1}(x_1) - F_{X_1}(x_1-h)}{h}} = \frac{\int_{-\infty}^{x} f_{X_1,X_2}(x_1, y_2)\mathrm{d}y_2}{f_{X_1}(x_1)}.$$

This implies conditional distribution of X_2 given $X_1 = x_1$ (provided $f_{X_1}(x_1) > 0$) is continuous with p.d.f.

$$f_{X_2|X_1}(x|x_1) = \frac{f_{X_1,X_2}(x_1, x)}{f_{X_1}(x_1)},\ x \in \mathbb{R}$$

provided $f_{X_1}(x_1) > 0$.

The above discussion easily extends to general $p \ge 2$ by defining conditional d.f. of $\underline{X}_2 = (X_{q+1}, \ldots, X_p)$ given $\underline{X}_1 = (X_1, X_2, \ldots, X_q) = (x_1, x_2, \ldots, x_q) = \underline{x}_1$ as

$$F_{\underline{X}_2|\underline{X}_1}(\underline{x}_2|\underline{x}_1) = \lim_{h \downarrow 0} \Pr(X_j \le x_j,\ j = q+1, \ldots, p | x_i - h_i < X_i \le x_i, i = 1, 2, \ldots, q),$$

where $\underline{x}_2 = (x_{q+1}, x_{q+2}, \ldots, x_p) \in S_{\underline{X}_2}$.

Definition 0.58. *Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with joint p.d.f. $f_{\underline{X}}(\cdot)$. Let $q \in \{1, 2, \ldots, p-1\}$, $\underline{X}_1 = (X_1, X_2, \ldots, X_q)$ and $\underline{X}_2 = (X_{q+1}, \ldots, X_p)$. Then the conditional p.d.f. of $\underline{X}_2$ given $\underline{X}_1 = \underline{x}_1$ is defined by*

$$f_{\underline{X}_2|\underline{X}_1}(\underline{x}_2|\underline{x}_1) = \frac{f_{\underline{X}_1,\underline{X}_2}(\underline{x}_1, \underline{x}_2)}{f_{\underline{X}_1}(\underline{x}_1)} = \frac{f_{\underline{X}}(\underline{x}_1, \underline{x}_2)}{f_{\underline{X}_1}(\underline{x}_1)},\ \underline{x}_2 \in \mathbb{R}^{p-q}.$$

Theorem 0.59. *Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a continuous random vector with joint p.d.f. $f_{\underline{X}}(\cdot)$ and marginal p.d.f.s $f_{X_i}(\cdot)$, $i = 1, 2, \ldots, p$. Then $X_1, X_2, \ldots, X_p$ are independent iff*

$$f_{X_1,X_2,\ldots,X_p}(x_1, x_2, \ldots, x_p) = \prod_{i=1}^{p} f_{X_i}(x_i),\ \ \underline{x} = (x_1, x_2, \ldots, x_p) \in \mathbb{R}^p.$$

Proof. Exercise. □

Remark 0.60. *(a) Let $S_{\underline{X}}$ be the support of distribution $\underline{X} = (X_1, X_2, \ldots, X_p)$ and let $S_{\underline{X_i}}$ be the support of distribution of X_i, $i = 1, 2, \ldots, p$. It can be shown that if $X_1, X_2, \ldots, X_p$ are independent then $S_{\underline{X}} = \prod_{i=1}^{p} S_{\underline{X_i}}$ (cartesian product).*

(b) Let $\underline{X} = (X_1, X_2)$ be a continuous random vector. Then X_1 and X_2 are independent iff $\forall\, x_1 \in S_{X_1}$,

$$f_{X_2|X_1}(x_2|x_1) = f_{X_2}(x_2)\ \forall\ x_2 \in \mathbb{R}.$$

Theorem 0.61. *Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a continuous random vector with joint p.d.f. $f_{\underline{X}}(\cdot)$ and marginal p.d.f.s $f_{X_i}(\cdot)$, $i = 1, 2, \ldots, p$. Then $X_1, X_2, \ldots, X_p$ are independent iff*

$$f_{X_1,X_2,\ldots,X_p}(x_1, \ldots, x_p) = \prod_{i=1}^{p} g_i(x_i),\ \ \underline{x} \in \mathbb{R}^p,$$

for some non-negative functions $g_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, p$. In that case $f_{X_i}(x) = c_i g_i(x)$, $x \in \mathbb{R}$, for some positive constants $c_i,\ i = 1, 2, \ldots, p$.

Example 0.62. *Let $\underline{X} = (X_1, X_2, X_3)$ have the joint p.d.f.*

$$f_{\underline{X}}(x_1, x_2, x_3) = \begin{cases} \dfrac{1}{x_1 x_2}, \; 0 < x_3 < x_2 < x_1 < 1, \\ 0, \; \textit{otherwise}. \end{cases}$$

(a) Show that $f_{\underline{X}}(\cdot)$ is a proper p.d.f.

(b) Find the marginal p.d.f. of (X_2, X_3).

(c) Find the marginal p.d.f. of X_1.

(d) Find the conditional p.d.f. of X_1 given $(X_2, X_3) = (x_2, x_3)$ where $0 < x_3 < x_2 < 1$.

(e) Are X_1, X_2 and X_3 independent.

(f) Find the conditional p.d.f. of (X_1, X_3) given $X_2 = x_2$, where $0 < x_2 < 1$.

(g) Are X_1 and X_3 independent given $X_2 = x_2$, where $0 < x_2 < 1$.

Solution: (a) Clearly, $f_{\underline{X}}(\underline{x}) \geq 0 \; \forall \; \underline{x} \in \mathbb{R}^3$. Also,

$$\int_{\mathbb{R}^3} f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} = \int_0^1 \int_0^{x_1} \int_0^{x_2} \frac{1}{x_1 x_2} \mathrm{d}x_3 \mathrm{d}x_2 \mathrm{d}x_1 = 1.$$

So, $f_{\underline{X}}(\underline{x})$ is a p.d.f.

(b) The marginal p.d.f. of (X_2, X_3) is obtained as,

$$f_{X_2,X_3}(x_2, x_3) = \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x}) \mathrm{d}x_1 = \int_{x_2}^{1} \frac{1}{x_1 x_2} \mathrm{d}x_1 = -\frac{\ln x_2}{x_2}, \quad 0 < x_3 < x_2 < 1.$$

So,

$$f_{X_2,X_3}(x, y) = \begin{cases} -\dfrac{\ln x}{x}, \; 0 < y < x < 1, \\ 0, \; \text{otherwise}. \end{cases}$$

(c) For $X_1 \in \mathbb{R}$, the marginal of X_1 is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\underline{X}}(\underline{x}) \mathrm{d}x_2 \mathrm{d}x_3.$$

Now $0 < x_1 < 1$,

$$f_{X_1}(x_1) = \int_0^{x_1} \int_0^{x_2} \frac{1}{x_1 x_2} \mathrm{d}x_3 \mathrm{d}x_2 = 1.$$

Thus,

$$f_{X_1}(x) = \begin{cases} 1, \; 0 < x < 1, \\ 0, \; \text{otherwise}. \end{cases}$$

(d) The conditional distribution of X_1 given $(X_2, X_3) = (x_2, x_3)$ is

$$f_{X_1|(X_2,X_3)}(x_1|x_2, x_3) = \frac{f_{\underline{X}}(x_1, x_2, x_3)}{f_{X_2,X_3}(x_2, x_3)} = \frac{(1/x_1 x_2)}{(-\ln(x_2)/x_2)} = -\frac{1}{x_1 \ln x_2}, \; x_2 < x_1 < 1.$$

So, the conditional distribution of X_1 given $(X_2, X_3) = (x_2, x_3)$ is

$$f_{X_1|(X_2,X_3)}(x_1|x_2,x_3) = \begin{cases} -\dfrac{1}{x_1 \ln x_2}, \ x_2 < x_1 < 1, \\ 0, \ \text{otherwise.} \end{cases}$$

(e) We have $S_{\underline{X}} = \{\underline{x} \in \mathbb{R}^3 : 0 \le x_3 \le x_2 \le x_1 \le 1\} \neq S_{X_1} \times S_{X_2} \times S_{X_3} = [0,1] \times [0,1] \times [0,1]$. So, X_1, X_2 and X_3 are not independent.

(f) For fixed $x_2 \in \mathbb{R}$, $f_{X_1,X_3|X_2}(x_1,x_3|x_2) \propto f_{X_1,X_2,X_3}(x_1,x_2,x_3)$. For fixed $0 < x_2 < 1$,

$$f_{X_1,X_3|X_2}(x_1,x_3|x_2) = \begin{cases} \dfrac{c(x_2)}{x_1}, \ 0 < x_3 < x_2, \ x_2 < x_1 < 1, \\ 0, \ \text{otherwise.} \end{cases}$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_3|X_2}(x_1,x_3|x_2)\mathrm{d}x_1\mathrm{d}x_3 = 1 \implies c(x_2) = -\frac{1}{x_2 \ln x_2}.$$

Thus, for fixed $0 < x_2 < 1$, $f_{X_1,X_3|X_2}(x_1,x_3|x_2) = g_{x_2}(x_1)h_{x_2}(x_3)$, $(x_1,x_3) \in \mathbb{R}^3$ where for fixed $x_2 \in (0,1)$

$$g_{x_2}(x) = \begin{cases} -\dfrac{1}{x x_2 \ln x_2}, \ x_2 < x < 1, \\ 0, \ \text{otherwise.} \end{cases} \quad ; \quad h_{x_2}(y) = \begin{cases} 1, \ 0 < y < x_2, \\ 0, \ \text{otherwise.} \end{cases}$$

$\implies$ given $X_2 = x_2$ $(0 < x_2 < 1)$ X_1 and X_3 are independently distributed.

0.3.4. Expectation and Moments

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with p.m.f. / p.d.f. $f(\cdot)$ and support S. Let $g : \mathbb{R}^p \to \mathbb{R}$ be a function.

Definition 0.63. *We say that the expected value of $g(\underline{X})$ (denoted by $E(g(\underline{X}))$) is finite and equals*

$$E(g(\underline{X})) = \begin{cases} \displaystyle\sum_{\underline{x}\in S} g(\underline{x})f(\underline{x}), \ \textit{if } \underline{X} \textit{ is discrete}, \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\underline{x})f(\underline{x})\mathrm{d}\underline{x}, \ \textit{if } \underline{X} \textit{ is continuous}, \end{cases}$$

provided $\displaystyle\sum_{\underline{x}\in S} |g(\underline{x})|f(\underline{x}) < \infty \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} |g(\underline{x})|f(\underline{x})\mathrm{d}\underline{x} < \infty\right)$.

Theorem 0.64. *Let $Y = g(\underline{X})$. Then Y has finite expectation iff* $\displaystyle\sum_{y\in S_Y} |y|f_Y(y) < \infty \left(\textit{or } \int_{-\infty}^{\infty} |y|f_Y(y)\mathrm{d}y < \infty\right)$ *and in that case*

$$E(g(\underline{X})) = \sum_{y\in S_Y} y f_Y(y) \quad \left(\int_{-\infty}^{\infty} y f_Y(y)\mathrm{d}y\right).$$

Here, S_Y denotes the support of Y and $f_Y(\cdot)$ denotes the p.m.f. / p.d.f. Y.

Some Special Expectations:

(a) For non-negative integers $k_1, k_2, \ldots, k_p$, $\mu'_{k_1,k_2,\ldots,k_p} = E\left(X_1^{k_1} X_2^{k_2} \ldots X_p^{k_p}\right)$, provided it is finite, is called a joint moment of order $k_1 + k_2 + \cdots + k_p$ of $\underline{X}$.

(b) For non-negative integers $k_1, k_2, \ldots, k_p$,

$$\mu_{k_1,k_2,\ldots,k_p} = E\left((X_1 - E(X_1))^{k_1} (X_2 - E(X_2))^{k_2} \ldots (X_p - E(X_p))^{k_p}\right),$$

provided it is finite, is called a joint central moment of order $k_1 + k_2 + \cdots + k_p$ of $\underline{X}$.

(c) The quantity $\mathrm{Cov}(X_1, X_2) = E\left((X_1 - E(X_1))(X_2 - E(X_2))\right)$, provided it is finite, is called covariance between X_1 and X_2.

Remark 0.65. *(a)*

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) &= E\left((X_1 - E(X_1))(X_2 - E(X_2))\right)\\ &= E\left((X_1 - \mu_1)(X_2 - \mu_2)\right)\\ &= E(X_1X_2 - X_1\mu_2 - \mu_2 X_1 + \mu_1\mu_2) = E(X_1X_2) - E(X_1)E(X_2).\end{aligned}$$

(b) $\mathrm{Cov}(X_1, X_1) = \mathrm{Var}(X_1)$.

(c) $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1)$

Theorem 0.66. *Let a_i, $i = 1, 2, \ldots, p$ and b_j, $j = 1, 2, \ldots, r$ are real constants and let X_i, $i = 1, 2, \ldots, p$, Y_j, $j = 1, 2, \ldots, r$ be random variables. Then*

(a) $E\left(\sum_{i=1}^{p} a_i X_i\right) = \sum_{i=1}^{p} a_i E(X_i)$, *provided the involved expectations are finite.*

(b) $\mathrm{Cov}\left(\sum_{i=1}^{p} a_i X_i, \sum_{j=1}^{r} b_j Y_j\right) = \sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j \,\mathrm{Cov}(X_i, Y_j)$, *provided the involved expectations are finite.*

(c) $$\begin{aligned}\mathrm{Var}\left(\sum_{i=1}^{p} a_i X_i\right) &= \sum_{i=1}^{p} a_i^2 \,\mathrm{Cov}(X_i, X_i) + \sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p} a_i a_j \,\mathrm{Cov}(X_i, X_j)\\ &= \textstyle\sum_{i=1}^{p} a_i^2 \,\mathrm{Var}(X_i) + 2 \sum\sum_{1\le i<j\le p} a_i a_j \,\mathrm{Cov}(X_i, X_j).\end{aligned}$$

Proof. (a) (*We will prove for continuous case*).

$$E\left(\sum_{i=1}^{p} a_i X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\sum_{i=1}^{p} a_i x_i\right) f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} = \sum_{i=1}^{p} a_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} = \sum_{i=1}^{p} a_i E(X_i).$$

(b) Note that

$$\begin{aligned}\mathrm{Cov}\left(\sum_{i=1}^{p} a_i X_i, \sum_{j=1}^{r} b_j Y_j\right) &= E\left[\left(\sum_{i=1}^{p} a_i X_i - E\left(\sum_{i=1}^{p} a_i X_i\right)\right)\left(\sum_{j=1}^{r} b_j Y_j - E\left(\sum_{j=1}^{r} b_j Y_j\right)\right)\right]\\ &= E\left[\left(\sum_{i=1}^{p} a_i X_i - \sum_{i=1}^{p} a_i E(X_i)\right)\left(\sum_{j=1}^{r} b_j Y_j - \sum_{j=1}^{r} b_j E(Y_j)\right)\right]\end{aligned}$$

$$
\begin{aligned}
&= E\left[\left(\sum_{i=1}^{p} a_i(X_i - E(X_i))\right)\left(\sum_{j=1}^{r} b_j(Y_j - E(Y_j))\right)\right] \\
&= E\left[\left(\sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j (X_i - E(X_i))(Y_j - E(Y_j))\right)\right] \\
&= \sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j E\left[(X_i - E(X_i))(Y_j - E(Y_j))\right] = \sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j \operatorname{Cov}(X_i, Y_j).
\end{aligned}
$$

(c) Note that

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{p} a_i X_i\right) &= \operatorname{Cov}\left(\sum_{i=1}^{p} a_i X_i, \sum_{j=1}^{p} a_j X_j\right) \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j \operatorname{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{p} a_i^2 \operatorname{Cov}(X_i, X_i) + \sum_{i=1}^{p}\sum_{\substack{j=1 \\ j\neq i}}^{p} a_i a_j \operatorname{Cov}(X_i, X_j) \\
&= \sum_{i=1}^{p} a_i^2 \operatorname{Var}(X_i) + \sum_{i=1}^{p}\sum_{\substack{j=1 \\ j\neq i}}^{p} a_i a_j \operatorname{Cov}(X_i, X_j) = \sum_{i=1}^{p} a_i^2 \operatorname{Var}(X_i) + 2\sum\sum_{1\leq i<j\leq p} a_i a_j \operatorname{Cov}(X_i, X_j).
\end{aligned}
$$

This completes the proof. □

Theorem 0.67. *Let $X_1, X_2, \ldots, X_p$ be independent random variables, let $\psi_i : \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, p$ be given functions. Then,*

(a) $E\left(\prod_{i=1}^{p} \psi_i(X_i)\right) = \prod_{i=1}^{p} E(\psi_i(X_i))$, provided the involved expectations are finite,

(b) for any $A_1, A_2, \ldots, A_p \in \mathscr{B}_p$,

$$
\Pr(X_1 \in A_1, X_2 \in A_2, \ldots, X_p \in A_p) = \prod_{i=1}^{p} \Pr(X_i \in A_i),
$$

(c) $\psi_1(X_1), \psi_2(X_2), \ldots, \psi_p(X_p)$ are independent random variables.

Proof. (*We will prove for $p = 2$ in continuous case*).

(a)

$$
\begin{aligned}
E(\psi_1(X_1)\psi_2(X_2)) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi_1(X_1)\psi_2(X_2) f_{X_1,X_2}(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi_1(X_1)\psi_2(X_2) f_{X_1}(x_1) f_{X_2}(x_2)\mathrm{d}x_1\mathrm{d}x_2 \quad (X_1 \perp\!\!\!\perp X_2) \\
&= \left(\int_{-\infty}^{\infty} \psi_1(X_1) f_{X_1}(x_1)\mathrm{d}x_1\right)\left(\int_{-\infty}^{\infty} \psi_2(X_2) f_{X_2}(x_2)\mathrm{d}x_2\right) = E(\psi_1(X_1))E(\psi_2(X_2)).
\end{aligned}
$$

(b) Take $\psi_i(X_i) = \begin{cases} 1, \text{ if } X_i \in A_i, \\ 0, \text{ otherwise,} \end{cases}$ in (a). Note that $\psi_1(X_1)\psi_2(X_2) = \begin{cases} 1, \text{ if } X_i \in A_i, \\ 0, \text{ otherwise.} \end{cases}$

$E(\psi_i(X_i)) = \Pr(X_i \in A_i)$, $i = 1, 2$ and $E(\psi_1(X_1)\psi_2(X_2)) = \Pr(X_1 \in A_1, X_2 \in A_2)$. Now the result follows from (a).

(c) Let $Y_i = \psi_i(X_i)$, $i = 1, 2$. For fixed $\underline{y} = (y_1, y_2) \in \mathbb{R}^2$, define

$$g_i(X_i) = \begin{cases} 1, \text{ if } Y_i = \psi_i(X_i) \le y_i, \ i = 1, 2, \\ 0, \text{ otherwise.} \end{cases}$$

Then by (a) $E(g_1(X_1)g_2(X_2)) = E(g_1(X_1))E(g_2(X_2))$. Also,

$$\begin{aligned} g_1(X_1)g_2(X_2) &= \begin{cases} 1, \text{ if } \psi_1(X_1) \le y_1, \ \psi_2(X_2) \le y_2, \\ 0, \text{ otherwise,} \end{cases} \\ &= \begin{cases} 1, \text{ if } Y_1 \le y_1, \ Y_2 \le y_2, \\ 0, \text{ otherwise.} \end{cases} \end{aligned}$$

So, $E(g_1(X_1)g_2(X_2)) = \Pr(Y_1 \le y_1, Y_2 \le y_2)$ and $E(g_i(X_i)) = \Pr(Y_i \le y_i), \ i = 1, 2$. Consequently, $\Pr(Y_1 \le y_1, Y_2 \le y_2) = \Pr(Y_1 \le y_1)\Pr(Y_2 \le y_2) \ \forall \ (y_1, y_2) \in \mathbb{R}^2 \implies Y_1 = \psi_1(X_1)$ and $Y_2 = \psi_2(X_2)$ are independent random variables. □

Corollary 0.68. *Let $X_1, X_2, \ldots, X_p$ are independent random variables. Then*

(a) $\mathrm{Cov}(X_i, X_j) = 0 \ \forall \ i \ne j$.

(b) For real constants $a_1, a_2, \ldots, a_p$, we have

$$\mathrm{Var}\left(\sum_{i=1}^{p} a_i X_i\right) = \sum_{i=1}^{p} a^2 \,\mathrm{Var}(X_i).$$

Proof. (a) For $i \ne j$, $\mathrm{Cov}(X_i, X_j) = E(X_iX_j) - E(X_i)E(X_j) = E(X_i)E(X_j) - E(X_i)E(X_j) = 0$.

(b) $\mathrm{Var}\left(\sum_{i=1}^{p} a_i X_i\right) = \sum_{i=1}^{p} a^2 \,\mathrm{Var}(X_i) + \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \ne i}}^{p} a_i a_j \,\mathrm{Cov}\,(X_i, X_j) = \sum_{i=1}^{p} a^2 \,\mathrm{Var}(X_i)$, (using (a)). □

Definition 0.69. *(a) The correlation between random variables X_1 and X_2 is defined by*

$$\rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}},$$

provided $0 < \mathrm{Var}(X_i) < \infty$, $i = 1, 2$.

(b) Random variables X_1 and X_2 are said to be uncorrelated if $\rho(X_1, X_2) = 0$ (or equivalently $\mathrm{Cov}(X_1, X_2) = 0$).

Remark 0.70. *If X_1 and X_2 are independent random variables $\implies$ X_1 and X_2 are uncorrelated. converse may not be true.*

Example 0.71 (Uncorrelated random variables may not be independent)**.** *Let (X, Y) have joint p.m.f.*

$$f(x, y) = \begin{cases} \frac{1}{2}, \textit{ if } (x, y) = (0, 0), \\ \frac{1}{4}, \textit{ if } (x, y) = (1, -1), (1, 1), \\ 0, \textit{ otherwise}. \end{cases}$$

Then

$$f_X(x) = \begin{cases} \frac{1}{2}, & \text{if } x = 0, 1, \\ 0, & \text{otherwise}, \end{cases} \quad f_Y(y) = \begin{cases} \frac{1}{4}, & \text{if } y = -1, 1, \\ \frac{1}{2}, & \text{if } y = 0, \\ 0, & \text{otherwise}. \end{cases}$$

Clearly, there exists $(x, y) \in \mathbb{R}^2$ *such that* $f_{X,Y}(x, y) \neq f_X(x)f_Y(y) \implies X$ *and* Y *are not independent (in fact* $\Pr(X = Y^2) = 1$*).*

However, $E(XY) = E(Y) = 0$ *and* $E(X) = \frac{1}{2} \implies \operatorname{Cov}(X, Y) = 0 \implies \rho(X, Y) = 0$.

Theorem 0.72 (Cauchy-Schwarz Inequality)**.** *For random variables* X *and* Y

$$(E(XY))^2 \leq E(X^2)E(Y^2) \tag{0.10}$$

provided involved expectations are finite. The equality is attained iff $\Pr(Y = cX) = 1$ *or* $\Pr(X = cY) = 1$*, for some real constant* c*.*

Proof. Case I: $E(X^2) = 0$. In this case $\Pr(X = 0) = 1$. Therefore $\Pr(XY = 0) = 1$ and $E(XY) = 0$. We have inequality in (0.10).

Case II: $E(X^2) > 0$. Then

$$\begin{aligned} &E((Y - cX)^2) \geq 0 \;\forall\, c \in \mathbb{R} \\ &\implies c^2E(X^2) - 2cE(XY) + E(Y^2) \geq 0 \;\forall\, c \in \mathbb{R} \\ &\implies \text{Discriminant} \leq 0 \implies (2E(XY))^2 - 4(E(X^2))E(Y^2) \leq 0 \implies (E(XY))^2 \leq E(X^2)E(Y^2). \end{aligned}$$

Clearly, equality is attained iff $E((Y - cX)^2) = 0$ for some $c \in \mathbb{R} \implies \Pr(Y = cX) = 1$ for some $c \in \mathbb{R}$. By symmetry $\Pr(X = cY) = 1$ for some $c \in \mathbb{R}$. ☐

Corollary 0.73. *Let* X_1 *and* X_2 *be random variables with* $E(X_i) = \mu_i \in (-\infty, \infty)$ *and* $\operatorname{Var}(X_i) = \sigma_i^2 \in (0, \infty)$, $i = 1, 2$. *Then*

(a) $|\rho(X_1, X_2)| \leq 1$.

(b) $|\rho(X_1, X_2)| = 1$ *iff* $\Pr\left(\frac{X_1 - \mu_1}{\sigma_1} = c\frac{X_2 - \mu_2}{\sigma_2}\right) = 1$ *or* $\Pr\left(\frac{X_2 - \mu_2}{\sigma_2} = c\frac{X_1 - \mu_1}{\sigma_1}\right) = 1$*, for some real constant* c*.*

Proof. Let $X = \frac{X_1 - \mu_1}{\sigma_1}$ and $Y = \frac{X_2 - \mu_2}{\sigma_2}$. Using Cauchy-Schwarz inequality $(E(XY))^2 \leq E(X^2)E(Y^2)$ but

$$E(X^2) = \frac{E(X_1 - \mu_1)^2}{\sigma_1^2} = 1 \text{ and } E(Y^2) = \frac{E(X_2 - \mu_2)^2}{\sigma_2^2} = 1.$$

Thus

$$\left(\frac{E((X_1 - \mu_1)(X_2 - \mu_2))}{\sigma_1\sigma_2}\right)^2 \leq 1 \implies \rho^2(X_1, X_2) \leq 1 \implies |\rho(X_1, X_2)| \leq 1$$

and equality is attained iff $\Pr(X = cY) = 1$, for some real constants $c \implies \Pr\left(\frac{X_1 - \mu_1}{\sigma_1} = c\frac{X_2 - \mu_2}{\sigma_2}\right) = 1$ for some real constants c. ☐

0.3.5. Conditional Expectation, Conditional Variance and Conditional Covariance

Definition 0.74. *(a) Let $\underline{X}$ be a p-dimensional random vector and $\underline{Y}$ be a q-dimensional random vector. Let $\underline{y} \in \mathbb{R}^q$ be such that $f_{\underline{Y}}(\underline{y}) > 0$ and let $\psi : \mathbb{R}^p \to \mathbb{R}$ be a given function. Here $f_{\underline{Y}}(\cdot)$ is the p.d.f. / p.m.f. of random vector $\underline{Y}$. Then*

(i) The conditional expectation of $\psi(\underline{X})$ given $\underline{Y} = \underline{y}$ (denoted by $E(\psi(\underline{X})|\underline{Y} = \underline{y})$) is the expectation of $\psi(\underline{X})$ under the conditional distribution of $\underline{X}$ given $\underline{Y} = \underline{y}$.

(ii) The conditional variance of $\psi(\underline{X})$ given $\underline{Y} = \underline{y}$ (denoted by $\mathrm{Var}(\psi(\underline{X})|\underline{Y} = \underline{y})$) is the variance of $\psi(\underline{X})$ under the conditional distribution of $\underline{X}$ given $\underline{Y} = \underline{y}$.

(b) Let X_1 and X_2 be two random variables and $\underline{Y}$ be a q-dimensional random vector. Then the conditional covariance between X_1 and X_2 given $\underline{Y} = \underline{y}$, (denoted by $\mathrm{Cov}(X_1, X_2|\underline{Y} = \underline{y})$) is the covariance between X_1 and X_2 under the conditional distribution of (X_1, X_2) given $\underline{Y} = \underline{y}$.

Notation Let for $\underline{y} \in \{\underline{t} \in \mathbb{R}^q : f_{\underline{Y}}(\underline{t}) > 0\}$, $\psi_1(\underline{y}) = E(\psi(\underline{X})|\underline{Y} = \underline{y})$ and $\psi_2(\underline{y}) = \mathrm{Var}(\psi(\underline{X})|\underline{Y} = \underline{y})$ and $\psi_3(\underline{y}) = \mathrm{Cov}(X_1, X_2|\underline{Y} = \underline{y})$. We denote $\psi_1(\underline{Y}) = E(\psi(\underline{X})|\underline{Y})$ and $\psi_2(\underline{Y}) = \mathrm{Var}(\psi(\underline{X})|\underline{Y})$ and $\psi_3(\underline{Y}) = \mathrm{Cov}(X_1, X_2|\underline{Y})$.

Theorem 0.75. *Under the above notation*

(a) $E(\psi(\underline{X})) = E(E(\psi(\underline{X})|\underline{Y}))$,

(b) $\mathrm{Var}(\psi(\underline{X})) = \mathrm{Var}(E(\psi(\underline{X})|\underline{Y})) + E(\mathrm{Var}(\psi(\underline{X})|\underline{Y}))$,

(c) $\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(E(X_1|\underline{Y}), E(X_2|\underline{Y})) + E(\mathrm{Cov}(X_1, X_2|\underline{Y}))$.

Proof. (*We will prove for $p = q = 1$ continuous case*).
(a)

$$\begin{aligned}
E(E(\psi(\underline{X})|\underline{Y})) &= \int_{-\infty}^{\infty} E(\psi(\underline{X})|\underline{Y} = \underline{y}) f_{\underline{Y}}(\underline{y}) \mathrm{d}\underline{y} \\
&= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \psi(\underline{x}) f_{X|Y}(x|y) \mathrm{d}x \right] f_{\underline{Y}}(\underline{y}) \mathrm{d}\underline{y} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x) f_{X|Y}(x|y) f_Y(y) \mathrm{d}x \mathrm{d}y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x) f_{X,Y}(x,y) \mathrm{d}x \mathrm{d}y = E(\psi(X)).
\end{aligned}$$

(b) Follows from (c).

(c) From (a), we have

$$\mathrm{Cov}(X_1, X_2) = E((X_1 - E(X_1))(X_2 - E(X_2))) = E\left[E\left[(X_1 - E(X_1))(X_2 - E(X_2))|\underline{Y}\right]\right].$$

Now,

$$\begin{aligned}
&E\left[(X_1 - E(X_1))(X_2 - E(X_2))|\underline{Y}\right] \\
&\quad = E\left[(X_1 - E(X_1|\underline{Y}) + E(X_1|\underline{Y}) - E(X_1))(X_2 - E(X_2|Y) + E(X_2|Y) - E(X_2))|\underline{Y}\right] \\
&\quad = E\left[(X_1 - E(X_1|Y))(X_2 - E(X_2|\underline{Y}))|\underline{Y}\right] + (E(X_1|\underline{Y}) - E(X_1))(E(X_2|\underline{Y}) - E(X_2)) \\
&\quad = \mathrm{Cov}(X_1, X_2|\underline{Y}) + (E(X_1|\underline{Y}) - E(X_1))(E(X_2|Y) - E(X_2)). \\
\Longrightarrow \mathrm{Cov}(X_1, X_2) &= E(\mathrm{Cov}(X_1, X_2|\underline{Y})) + E\left[(E(X_1|\underline{Y}) - E(X_1))(E(X_2|Y) - E(X_2))\right] \\
&= \mathrm{Cov}(E(X_1|\underline{Y}), E(X_2|\underline{Y})) + E(\mathrm{Cov}(X_1, X_2|\underline{Y})).
\end{aligned}$$

This completes the proof. □

0.3.6. Joint Moment Generating Function

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with p.d.f. /p.m.f. $f_{\underline{Y}}(\cdot)$. $A = \{\underline{t} = (t_1, t_2, \ldots, t_p) \in \mathbb{R}^p : E\left(e^{\sum_{i=1}^p t_i X_i}\right) < \infty\}$.

Definition 0.76. *(a) The function $M_{\underline{X}} : A \to \mathbb{R}$ defined by*

$$M_{\underline{X}}(\underline{t}) = E\left(e^{\sum_{i=1}^p t_i X_i}\right), \;\; \underline{t} = (t_1, t_2, \ldots, t_p) \in A$$

is called the joint moment generating function (m.g.f.) of random vector $\underline{X} = (X_1, X_2, \ldots, X_p)$.

Notation: For $\underline{a} = (a_1, a_2, \ldots, a_p) \in \mathbb{R}^p$, $-\underline{a} = (-a_1, -a_2, \ldots, -a_p)$ and $(-\underline{a}, \underline{a}) = (-a_1, a_1) \times \cdots \times (-a_p, a_p)$, $\underline{a} = (a_1, a_2, \ldots, a_p) > 0 \iff a_i > 0, i = 1, 2, \ldots, p$.

Remark 0.77. *(i) As $M_{\underline{X}}(\underline{0}) = 1$, we have $A \neq \phi$. Moreover $M_{\underline{X}}(\underline{t}) > 0 \, \forall \, \underline{t} \in A$.*

(ii) If $X_1, X_2, \ldots, X_p$ are independent then

$$M_{\underline{X}}(\underline{t}) = E\left(e^{\sum_{i=1}^p t_i X_i}\right) = E\left(\prod_{i=1}^p e^{t_i X_i}\right) = \prod_{i=1}^p E\left(e^{t_i X_i}\right) = \prod_{i=1}^p M_{X_i}(t_i) \; \forall \; \underline{t} \in A.$$

Conversely, suppose that $A \subseteq (-\underline{a}, \underline{a})$ for some $\underline{a} > 0$ and $M_{\underline{X}}(\underline{t}) = \prod_{i=1}^p M_{X_i}(t_i) \; \forall \; \underline{t} \in A$, then it can be shown that $X_1, X_2, \ldots, X_p$ are independent.

(iii) Let $X_1, X_2, \ldots, X_p$ be independent random variables and let $Y = \sum_{i=1}^p X_i$, then

$$M_Y(t) = E\left(e^{t \sum_{i=1}^p X_i}\right) = E\left(\prod_{i=1}^p e^{t X_i}\right) = \prod_{i=1}^p E\left(e^{t X_i}\right) = \prod_{i=1}^p M_{X_i}(t), \; t \in A.$$

In particular, if $X_1, X_2, \ldots, X_p$ are independent and identically distributed (iid) with common m.g.f. $M(t)$, then $M_Y(t) = (M(t))^p$, $t \in A$.

Theorem 0.78. *Suppose that the joint m.g.f. $M_{\underline{X}}(\underline{t})$ is finite on a rectangle $(-\underline{a}, \underline{a}) \in \mathbb{R}^p$, $\underline{a} > 0$. Then $M_{\underline{X}}(\underline{t})$ posseses partial derivatives of all order in $(-\underline{a}, \underline{a})$. Furthermore, for non-negative integers $k_1, k_2, \ldots, k_p$*

$$E\left(X_1^{k_1} X_2^{k_2} \ldots X_p^{k_p}\right) = \left[\frac{\partial^{k_1 + k_2 + \cdots + k_p}}{\partial t_1^{k_1} \partial t_2^{k_2} \ldots \partial t_p^{k_p}} M_{\underline{X}}(\underline{t})\right]_{\underline{t} = \underline{0}}.$$

Proof. (*We give an outline of the proof*).

$$\begin{aligned}
M_{\underline{X}}(t_1, t_2, \ldots, t_p) &= E\left(e^{\sum_{i=1}^p t_i X_i}\right) = \int_{\mathbb{R}^p} e^{\sum_{i=1}^p t_i x_i} f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x}, \\
\frac{\partial^{k_1 + \cdots + k_p}}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} M_{\underline{X}}(\underline{t}) &= \int_{\mathbb{R}^p} x_1^{k_1} \cdots x_p^{k_p} e^{\sum_{i=1}^p t_i X_i} f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x}, \\
\left[\frac{\partial^{k_1 + \cdots + k_p}}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} M_{\underline{X}}(\underline{t})\right]_{\underline{t} = \underline{0}} &= \int_{\mathbb{R}^p} x_1^{k_1} \cdots x_p^{k_p} f_{\underline{X}}(\underline{x}) \mathrm{d}\underline{x} = E\left(X_1^{k_1} X_2^{k_2} \ldots X_p^{k_p}\right).
\end{aligned}$$

This completes the proof. ☐

Let $\psi_{\underline{X}}(\underline{t}) = \ln M_{\underline{X}}(\underline{t})$, $\underline{t} \in (-\underline{a}, \underline{a})$. Then

$$E(X_i) = \left[\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} = \left[\frac{\partial}{\partial t_i} \psi_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}},$$
$$E(X_i^m) = \left[\frac{\partial^m}{\partial t_i^m} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}},\ m = 1, 2, \ldots,\ i = 1, 2, \ldots, p,$$
$$\mathrm{Var}(X_i) = \left[\frac{\partial^2}{\partial t_i^2} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} - \left(\left[\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}}\right)^2 = \left[\frac{\partial^2}{\partial t_i^2} \psi_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}},\ i = 1, 2, \ldots, p,$$

provided $M_{\underline{X}}(\underline{t})$ is finite on $(-\underline{a}, \underline{a})$, for some $\underline{a} > 0$. For $i \neq j$, if $M_{\underline{X}}(\underline{t})$ is finite on $(-\underline{a}, \underline{a})$, for some $\underline{a} > 0$,

$$\begin{aligned}\mathrm{Cov}(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) = E((X_i - E(X_i))(X_j - E(X_j))) \\ &= \left[\frac{\partial^2}{\partial t_i \partial t_j} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} - \left[\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} \left[\frac{\partial}{\partial t_j} M_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}} = \left[\frac{\partial^2}{\partial t_i \partial t_j} \psi_{\underline{X}}(\underline{t})\right]_{\underline{t}=\underline{0}}.\end{aligned}$$

Moreover,

$$M_{\underline{X}}(0, \ldots, 0, t_i, 0, \ldots, 0) = E(e^{t_i X_i}) = M_{X_i}(t_i),\ i = 1, 2, \ldots, p,$$
$$M_{\underline{X}}(0, \ldots, 0, t_i, 0, \ldots, 0, t_j, 0, \ldots, 0) = E(e^{t_i X_i + t_j X_j}) = M_{X_i, X_j}(t_i, t_j),$$

provided the m.g.f. is finite.

0.3.7. Equality in Distribution

Definition 0.79. *Two p-dimensional random vectors $\underline{X}$ and $\underline{Y}$ are said to have the same distribution (written as $\underline{X} \stackrel{d}{=} \underline{Y}$) if $F_{\underline{X}}(\underline{x}) = F_{\underline{Y}}(\underline{x})\ \forall\ \underline{x} \in \mathbb{R}^p$.*

Theorem 0.80. *(a) Let $\underline{X}$ and $\underline{Y}$ be discrete random vectors with p.m.f.s $f_{\underline{X}}(\cdot)$ and $f_{\underline{Y}}(\cdot)$, respectively. Then*

$$\underline{X} \stackrel{d}{=} \underline{Y} \iff f_{\underline{X}}(\underline{x}) = f_{\underline{Y}}(\underline{x})\ \forall\ \underline{x} \in \mathbb{R}^p.$$

(b) Let $\underline{X}$ and $\underline{Y}$ be continuous random vectors. Then

$$\underline{X} \stackrel{d}{=} \underline{Y} \iff f_{\underline{X}}(\underline{x}) = f_{\underline{Y}}(\underline{x})\ \forall\ \underline{x} \in \mathbb{R}^p,$$

for some versions $f_{\underline{X}}(\cdot)$ and $f_{\underline{Y}}(\cdot)$ of p.d.f.s of $\underline{X}$ and $\underline{Y}$, respectively.

(c) Let $\underline{X}$ and $\underline{Y}$ be p-dimensional random vectors and let $\psi : \mathbb{R}^p \to \mathbb{R}^q$ be a given function. Then

$$\underline{X} \stackrel{d}{=} \underline{Y} \iff \psi(\underline{X}) \stackrel{d}{=} \psi(\underline{Y})$$

(d) Let $\underline{X}$ and $\underline{Y}$ be p-dimensional random vectors with finite m.g.f.s $M_{\underline{X}}(\underline{t})$ and $M_{\underline{Y}}(\underline{t})$ on a rectangle $(-\underline{a}, \underline{a})$, for some $\underline{a} > 0$. Then

$$M_{\underline{X}}(\underline{t}) = M_{\underline{Y}}(\underline{t})\ \forall\ (-\underline{a}, \underline{a}) \implies \underline{X} \stackrel{d}{=} \underline{Y}.$$

0.3.8. Some Generalizations

Let $\underline{X}_i$: a p_i- dimensional random vector, $i = 1, 2, \ldots, m$. $F_{\underline{X}_i}$: d.f. of $\underline{X}_i$, $i = 1, 2, \ldots, m$, $f_{\underline{X}_i}$: p.m.f. / p.d.f. of $\underline{X}_i$, $i = 1, 2, \ldots, m$, $\sum_{i=1}^{p} p_i = p$, $\underline{X} = (\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_m)$: p-dimensional random vector with d.f. $F_{\underline{X}}(\cdot)$ and p.m.f. / p.d.f. $f_{\underline{X}}(\cdot)$.

Definition 0.81. *The random vectors $\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}$ are said to be independent if for any subcollection $\{\underline{X_{i_1}}, \underline{X_{i_2}}, \ldots, \underline{X_{i_q}}\}$ of $\{\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}\}$ ($2 \leq q \leq m$)*

$$F_{\underline{X_{i_1}}, \underline{X_{i_2}}, \ldots, \underline{X_{i_q}}}(\underline{x_1}, \underline{x_2}, \ldots, \underline{x_q}) = \prod_{j=1}^{q} F_{\underline{X_{i_j}}}(\underline{x_j}) \; \forall \; \underline{x} = (\underline{x_1}, \underline{x_2}, \ldots, \underline{x_q}) \in \mathbb{R}^{\sum_{j=1}^{q} p_{i_j}}.$$

Remark 0.82. *$\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}$ are independent $\implies$ random variables in any subset of $\{\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}\}$ are independent.*

Theorem 0.83. *(a) The following statements are equivalent:*

(i) $\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}$ are independent random vectors.

(ii) $F_{\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}}(\underline{x_1}, \underline{x_2}, \ldots, \underline{x_m}) = \prod_{i=1}^{m} F_{\underline{X_i}}(x_i) \; \forall \; \underline{x} = (\underline{x_1}, \underline{x_2}, \ldots, \underline{x_m}) \in \mathbb{R}^p$.

(iii) $f_{\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}}(\underline{x_1}, \underline{x_2}, \ldots, \underline{x_m}) = \prod_{i=1}^{m} f_{\underline{X_i}}(x_i) \; \forall \; \underline{x} = (\underline{x_1}, \underline{x_2}, \ldots, \underline{x_m}) \in \mathbb{R}^p$.

(iv) $f_{\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}}(\underline{x_1}, \underline{x_2}, \ldots, \underline{x_m}) = \prod_{i=1}^{m} g_i(x_i) \; \forall \; \underline{x} = (\underline{x_1}, \underline{x_2}, \ldots, \underline{x_m}) \in \mathbb{R}^p$ for some non-negative real valued function $g_i : \mathbb{R}^p \to \mathbb{R}$, $i = 1, 2, \ldots, m$.

(v) $\Pr(\underline{X_i} \in A_i, \; i =, 2, \ldots, m) = \prod_{i=1}^{m} \Pr(\underline{X_i} \in A_i) \; \forall \; A_i \in \mathscr{B}_{p_i}$, $i = 1, 2, \ldots, m$.

(b) If $\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}$ are independent random vectors, then

(i) $E\left(\prod_{i=1}^{m} \psi_i(\underline{X_i})\right) = \prod_{i=1}^{m} E\left(\psi_i(\underline{X_i})\right)$ for any functions ψ_i, $i = 1, 2, \ldots, m$.

(ii) $\psi_1(\underline{X_1}), \psi_2(\underline{X_2}), \ldots, \psi_m(\underline{X_m})$ are independent random vectors for any functions $\psi_1, \psi_2, \ldots, \psi_m$.

Definition 0.84. *Let Δ be an arbitrary index set. The random vectors $\{\underline{X}_\lambda : \lambda \in \Delta\}$ are said to be independent if random variables in any finite subcollection of $\{\underline{X}_\lambda : \lambda \in \Delta\}$ are independent.*

Theorem 0.85. *Under the notation of Theorem 0.83, $\underline{X_1}, \underline{X_2}, \ldots, \underline{X_m}$ are independent random vectors $\iff$ for some $\underline{a} > 0$ and $\forall \; \underline{t} = (\underline{t_1}, \underline{t_2}, \ldots, \underline{t_m}) \in (-\underline{a}, \underline{a})$,*

$$M_{\underline{X}}(\underline{t_1}, \underline{t_2}, \ldots, \underline{t_m}) = \prod_{i=1}^{m} M_{\underline{X_i}}(\underline{t_i}).$$

0.3.9. Functions of Random Vector

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with p.m.f. / p.d.f. $f(\cdot)$. Let $g : \mathbb{R}^p \to \mathbb{R}^q$, where $1 \leq q \leq p$ be a function defined on $\mathbb{R}^p$ and taking values in $\mathbb{R}^q$. Sometimes it may be of interest to derive the probability distribution of $\underline{Y} = g(\underline{X})$.

Definition 0.86. *(a) Let $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ be a collection of iid random vectors each having the (joint) d.f. F and the same p.m.f. / p.d.f. $f(\cdot)$. We call $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ a random sample (r.s.) of size n from a distribution having d.f. F (p.m.f. / p.d.f. $f(\cdot)$). In other words a random sample is a collection of iid random vectors.*

(b) A function of one or more random vectors that does not depend on any unknown parameter is called a statistic.

Example 0.87. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having p.d.f.*

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, \; x > 0, \\ 0, \; \textit{otherwise}, \end{cases}$$

where $\theta \in \mathfrak{H} = (0, \infty)$ is unknown. Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a statistic (called sample mean) but $X_1 - \theta$ is not a statistic. Some other statistic are:

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ or } S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \longrightarrow \textit{Sample Variance},$$

$$X_{r:n} = r\text{-th smallest of } X_1, X_2, \ldots, X_n,\ r = 1, 2, \ldots, n \text{ so that}$$
$$X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n} \longrightarrow r\text{-th order statistic},\ r = 1, \ldots, n,$$

$$X_{[np]:n}, 0 < p < 1;\ [x] = \text{largest integer} \leq x \longrightarrow p\text{-th sample quantile},$$

$$X_{[n/4]:n} \longrightarrow \text{sample lower quantile},\quad X_{[3n/4]:n} \longrightarrow \text{sample upper quantile},$$

$$M = \begin{cases} X_{\frac{n+1}{2}:n}, & \text{if } n \text{ is odd}, \\ \frac{X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}}{2}, & \text{if } n \text{ is even}, \end{cases} \longrightarrow \text{sample median},$$

$$S_n = \sqrt{S_n^2} \text{ or } S_{n-1} = \sqrt{S_{n-1}^2} \longrightarrow \text{sample standard deviation},$$

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right)}} \longrightarrow \text{sample correlation coefficient}.$$

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having d.f. F and p.m.f. / p.d.f. $f(\cdot)$. Then the joint d.f. of $\underline{X} = (X_1, X_2, \ldots, X_n)$ is

$$F_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} F(x_i),\ \ \underline{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

and the joint p.m.f. / p.d.f. of $\underline{X}$ is

$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{n} f(x_i),\ \ \underline{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n.$$

Theorem 0.88. *If $X_1, X_2, \ldots, X_n$ is a random sample, then*

$$(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_{\beta_1}, X_{\beta_2}, \ldots, X_{\beta_n})$$

for any permutation $(\beta_1, \beta_2, \ldots, \beta_n)$ of $(1, 2, \ldots, n)$.

Example 0.89. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a given distribution.*

(a) If X_1 is a continuous r.v. then $\Pr(X_1 < X_2 < \cdots < X_n) = \Pr(X_{\beta_1} < X_{\beta_2} < \cdots < X_{\beta_n}) = \frac{1}{n!}$, for any permutation $(\beta_1, \beta_2, \ldots, \beta_n)$ of $(1, 2, \ldots, n)$.

(b) If X_1 is a continuous r.v. then for any $r \in \{1, 2, \ldots, n\}$, $\Pr(X_i = X_{r:n}) = \frac{1}{n}$, $i = 1, 2, \ldots, n$.

(c) $E\left(\frac{X_i}{X_1 + X_2 + \cdots + X_n}\right) = \frac{1}{n}$, $i = 1, 2, \ldots, n$.

(d) $E\left(X_i | \sum_{j=1}^n X_j = t\right) = \frac{t}{n}$, $i = 1, 2, \ldots, n$.

Solution (a)

$$
\begin{aligned}
& X_1 \text{ is a continuous r.v. } \implies \underline{X} = (X_1, X_2, \ldots, X_n) \text{ is a continuous random vector. (Why?)} \\
\implies & (X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_{\beta_1}, X_{\beta_2}, \ldots, X_{\beta_n}) \text{ for any permutation } (\beta_1, \beta_2, \ldots, \beta_n) \text{ of } (1, 2, \ldots, n) \text{ and} \\
& \quad \Pr(\text{all } X_i\text{'s are distinct}) = 1 \\
\implies & (X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_{\beta_1}, X_{\beta_2}, \ldots, X_{\beta_n}) \text{ for any permutation } (\beta_1, \beta_2, \ldots, \beta_n) \text{ of } (1, 2, \ldots, n) \text{ and} \\
& \quad \sum_{\beta \in S_n} \Pr(X_{\beta_1} < X_{\beta_2} < \cdots < X_{\beta_n}) = 1, \text{ where } S_n \text{ is the set of all permutation of } (1, 2, \ldots, n) \\
\implies & \Pr(X_{\beta_1} < X_{\beta_2} < \cdots < X_{\beta_n}) = \Pr(X_1 < X_2 < \cdots < X_n) = \frac{1}{n!}.
\end{aligned}
$$

(b)

$$
\begin{aligned}
& \text{For any } i = 1, 2, \ldots, n, (X_1, X_2, \ldots, X_i, \ldots, X_n) \stackrel{d}{=} (X_i, X_2, \ldots, X_1, \ldots, X_n) \\
\implies & X_{r:n}, \ r\text{-th smallest of } (X_1, X_2, \ldots, X_i, \ldots, X_n) = r\text{-th smallest of } (X_i, X_2, \ldots, X_1, \ldots, X_n) \text{ and} \\
& \Pr(X_1 = r\text{-th smallest of } (X_1, X_2, \ldots, X_i, \ldots, X_n)) = \Pr(X_i = r\text{-th smallest of } (X_i, X_2, \ldots, X_1, \ldots, X_n)) \\
\implies & \Pr(X_1 = X_{r:n}) = \Pr(X_i = X_{r:n}), \ i = 1, 2, \ldots, n \\
& \quad \text{since } \Pr(X_{1:n} < X_{2:n} < \cdots < X_{n:n}) = 1, \ \text{(by (a)), we have } \sum_{i=1}^n \Pr(X_i = X_{r:n}) = 1 \\
\implies & \Pr(X_1 = X_{r:n}) = \Pr(X_i = X_{r:n}) = \frac{1}{n}.
\end{aligned}
$$

(c)

$$
\begin{aligned}
& (X_1, X_2, \ldots, X_i, \ldots, X_n) \stackrel{d}{=} (X_i, X_2, \ldots, X_1, \ldots, X_n) \\
\implies & E\left(\frac{X_1}{X_1 + X_2 + \cdots + X_i + \cdots + X_n}\right) = E\left(\frac{X_i}{X_i + X_2 + \cdots + X_1 + \cdots + X_n}\right) \\
\implies & E\left(\frac{X_1}{\sum_{j=1}^n X_j}\right) = E\left(\frac{X_i}{\sum_{j=1}^n X_j}\right) \text{ but } \sum_{i=1}^n E\left(\frac{X_i}{\sum_{j=1}^n X_j}\right) = E\left(\frac{\sum_{i=1}^n X_i}{\sum_{j=1}^n X_j}\right) = 1 \\
\implies & E\left(\frac{X_i}{\sum_{j=1}^n X_j}\right) = E\left(\frac{X_1}{\sum_{j=1}^n X_j}\right) = \frac{1}{n}, \ \ i = 1, 2, \ldots, n.
\end{aligned}
$$

(d)

$$
\begin{aligned}
& (X_1, X_2, \ldots, X_i, \ldots, X_n) \stackrel{d}{=} (X_i, X_2, \ldots, X_1, \ldots, X_n) \\
\implies & E\left(X_1 | X_1 + X_2 + \cdots + X_i + \cdots + X_n = t\right) = E\left(X_i | X_i + X_2 + \cdots + X_1 + \cdots + X_n = t\right) \\
\implies & E\left(X_1 \middle| \sum_{j=1}^n X_j = t\right) = E\left(X_i \middle| \sum_{j=1}^n X_j = t\right) \text{ but } \sum_{i=1}^n E\left(X_i \middle| \sum_{j=1}^n X_j = t\right) = E\left(\sum_{i=1}^n X_i \middle| \sum_{j=1}^n X_j = t\right) = t.
\end{aligned}
$$

Therefore

$$
E\left(X_i \middle| \sum_{j=1}^n X_j = t\right) = E\left(X_1 \middle| \sum_{j=1}^n X_j\right) = \frac{t}{n}, \ \ i = 1, 2, \ldots, n.
$$

0.3.10. Distribution Function Technique

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional random vector with d.f. F and p.m.f. / p.d.f. $f(\cdot)$. Also, let $\underline{g} : \mathbb{R}^p \to \mathbb{R}^q : \underline{g} = (g_1, g_2, \ldots, g_q)$, $\underline{Y} = (Y_1, Y_2, \ldots, Y_q) = (g_1(\underline{X}), g_2(\underline{X}), \ldots, g_q(\underline{X}))$. We are interested in the distribution of random vector $\underline{Y}$.

One can first find the d.f. of $\underline{Y} = (Y_1, Y_2, \ldots, Y_q)$

$$F_{\underline{Y}}(y_1, y_2, \ldots, y_q) = \Pr(g_1(\underline{X}) \le y_1, g_2(\underline{X}) \le y_2, \ldots, g_q(\underline{X}) \le y_q), \;\; \underline{y} = (y_1, y_2, \ldots, y_q) \in \mathbb{R}^q,$$

and then find the p.m.f. / p.d.f. of $\underline{Y} = (Y_1, Y_2, \ldots, Y_q)$.

Example 0.90. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having d.f. F, p.m.f. / p.d.f. f and support S. Let $Y_1 = \min\{X_1, X_2, \ldots, X_n\}$ and $Y_2 = \max\{X_1, X_2, \ldots, X_n\}$.*

(a) Find the joint d.f. of $\underline{Y} = (Y_1, Y_2)$.

(b) Find the marginal d.f.s of Y_1 and Y_2 using findings of (a).

(c) Find the marginal d.f.s of Y_1 and Y_2 directly (that is, without using (a)).

(d) Find the marginal p.m.f. / p.d.f. $\underline{Y} = (Y_1, Y_2)$ using findings in (b).

Solution: (a) For $(y_1, y_2) \in \mathbb{R}^2$,

$$\begin{aligned}
F_{\underline{Y}}(y_1, y_2) &= \Pr(Y_1 \le y_1, Y_2 \le y_2) \\
&= \Pr(\min\{X_1, X_2, \ldots, X_n\} \le y_1, \max\{X_1, X_2, \ldots, X_n\} \le y_2) \\
&= \Pr(\max\{X_1, X_2, \ldots, X_n\} \le y_2) - \Pr(\min\{X_1, X_2, \ldots, X_n\} > y_1, \max\{X_1, X_2, \ldots, X_n\} \le y_2) \\
&= \Pr(X_i \le y_2,\; i = 1, 2, \ldots, n) - \Pr(X_i > y_1,\; i = 1, 2, \ldots, n, X_i \le y_2, i = 1, 2, \ldots, n) \\
&= \prod_{i=1}^{n} \Pr(X_i \le y_2) - \Pr(y_1 < X_i \le y_2, i = 1, 2, \ldots, n) \\
&= \prod_{i=1}^{n} \Pr(X_i \le y_2) - \prod_{i=1}^{n} \Pr(y_1 < X_i \le y_2) = \begin{cases} [F(y_2)]^n - [F(y_2) - F(y_1)]^n, & -\infty < y_1 < y_2 < \infty, \\ [F(y_2)]^n, & -\infty < y_2 < y_1 < \infty. \end{cases}
\end{aligned}$$

(b)

$$\begin{aligned}
F_{Y_1}(y_1) &= \lim_{y_2 \to \infty} F_{\underline{Y}}(y_1, y_2) = 1 - [1 - F(y_1)]^n, -\infty < y_1 < \infty, \\
F_{Y_2}(y_2) &= \lim_{y_1 \to \infty} F_{\underline{Y}}(y_1, y_2) = [F(y_2)]^n, -\infty < y_2 < \infty.
\end{aligned}$$

(c)

$$\begin{aligned}
F_{Y_1}(y_1) &= \Pr(Y_1 \le y_1) \\
&= \Pr(\min\{X_1, X_2, \ldots, X_n\} \le y_1) \\
&= 1 - \Pr(\min\{X_1, X_2, \ldots, X_n\} > y_1) \\
&= 1 - \Pr(X_i > y_1, i = 1, 2, \ldots, n) = 1 - \prod_{i=1}^{n} \Pr(X_i > y_1) = 1 - [1 - F(y_1)]^n, -\infty < y_1 < \infty.
\end{aligned}$$

$$\begin{aligned}
F_{Y_2}(y_2) &= \Pr(Y_2 \le y_2) \\
&= \Pr(\max\{X_1, X_2, \ldots, X_n\} \le y_2)
\end{aligned}$$

$$= \Pr(X_i \leq y_2, i = 1, 2, \ldots, n) = \prod_{i=1}^{n} \Pr(X_i \leq y_2) = [F(y_2)]^n, -\infty < y_2 < \infty.$$

(d) **Case I:** X_1 is a discrete r.v. Then $S_{X_1} = S_{Y_1} = S_{Y_2}$. For $y_1 \in S_{X_1}$

$$f_{Y_1}(y_1) = \Pr(Y_1 = y_1) = F_{Y_1}(y_1) - F_{Y_1}(y_1-) = [1 - F(y_1-)]^n - [1 - F(y_1)]^n.$$

Thus,

$$f_{Y_1}(y_1) = \begin{cases} [1 - F(y_1-)]^n - [1 - F(y_1)]^n, \text{ if } y_1 \in S_{X_1}, \\ 0, \text{ otherwise.} \end{cases}$$

Similarly,

$$f_{Y_2}(y_2) = F_{Y_2}(y_2) - F_{Y_2}(y_2-) = \begin{cases} [F(y_2)]^n - [F(y_2-)]^n, \ \text{ if } y_2 \in S_{X_1} \\ 0, \ \text{ otherwise} \end{cases}$$

Case II: X_1 is a continuous r.v.

Let $F(\cdot)$ be differentiable everywhere (except possibly on a set having length zero (that is, it does not contain any open interval)

$$f_{Y_1}(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left(1 - [1 - F(y)]^n\right) = n\,[1 - F(y)]^{n-1} f(y), -\infty < y < \infty,$$
$$f_{Y_2}(y) = \frac{\mathrm{d}}{\mathrm{d}y}[F(y)]^n = n\,[F(y)]^{n-1} f(y), -\infty < y < \infty.$$

Example 0.91. *Let X_1 and X_2 be iid r.v.s with common p.d.f.*

$$f(x) = \begin{cases} 2x, \ 0 < x < 1, \\ 0, \ \textit{otherwise}. \end{cases}$$

Find the d.f. of $Y = X_1 + X_2$. Hence find the p.d.f. of Y.

Solution: The joint p.d.f. of $\underline{X} = (X_1, X_2)$ is

$$f_{\underline{X}}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = f(x_1) f(x_2) = \begin{cases} 4x_1x_2, \ 0 < x_1 < 1, \ 0 < x_2 < 1, \\ 0, \ \text{otherwise.} \end{cases}$$

For $y \in \mathbb{R}$,

$$F_Y(y) = \Pr(Y \leq y) = \Pr(X_1 + X_2 \leq y) = \underset{x_1+x_2 \leq y}{\int_0^1 \int_0^1} 4x_1x_2 \mathrm{d}x_1 \mathrm{d}x_2.$$

Clearly for $y < 0$, $F_Y(y) = 0$ and for $y \geq 2$, $F_Y(y) = 1$. Now consider $y \in [0, 1)$,

$$F_Y(y) = \int_0^y \int_0^{y-x_1} 4x_1x_2 \mathrm{d}x_2 \mathrm{d}x_1 = \frac{y^4}{6}.$$

For $y \in [1, 2)$,

$$F_Y(y) = \int_0^{y-1} \int_0^1 4x_1x_2 \mathrm{d}x_2 \mathrm{d}x_1 + \int_{y-1}^1 \int_0^{y-x_1} 4x_1x_2 \mathrm{d}x_2 \mathrm{d}x_1 = (y-1)^2 + \frac{(4y-3) - (y+3)(y-1)^3}{6}.$$

Thus,

$$F_Y(y) = \begin{cases} 0,\ y < 0, \\ \frac{y^4}{6},\ 0 \le y < 1, \\ (y-1)^2 + \frac{(4y-3)-(y+3)(y-1)^3}{6},\ 1 \le y < 2, \\ 1,\ y \ge 2. \end{cases}$$

Clearly, Y is continuous r.v. with p.d.f.

$$f_Y(y) = \begin{cases} \frac{2}{3}y^3,\ 0 < y < 1, \\ 2(y-1) + \frac{2}{3[1-(y+2)(y-1)^2]},\ 1 < y < 2, \\ 0,\ \text{otherwise.} \end{cases}$$

0.3.11. Transformation of Variable Technique

Let $\underline{X} = (X_1, X_2, \ldots, X_p)$ be a p-dimensional discrete random vector with support S and p.m.f. $f(\cdot)$. Let $g_i : \mathbb{R}^p \to \mathbb{R}$, $i = 1, 2, \ldots, k$ and $Y_i = g_i(\underline{X})$, $i = 1, 2, \ldots, k$ where $1 \le k \le p$ is an integer. Then $\underline{Y} = (Y_1, Y_2, \ldots, Y_k)$ is discrete random vector with support

$$T = \{(y_1, y_2, \ldots, y_k) : y_i = g_i(x_1, x_2, \ldots, x_p), i = 1, 2, \ldots, k \ \text{ for some } \underline{x} = (x_1, x_2, \ldots, x_p) \in S\},$$

d.f. $G(\underline{y}) = G(y_1, y_2, \ldots, y_k) = \sum_{\underline{x} \in A_{\underline{y}}} f(\underline{x})$, $\underline{y} \in \mathbb{R}^k$ and p.m.f.

$$g(\underline{y}) = \begin{cases} \sum_{\underline{x} \in B_{\underline{y}}} f(\underline{x}),\ \text{if } \underline{y} \in T, \\ 0,\ \text{otherwise,} \end{cases}$$

where $A_{\underline{y}} = \{\underline{x} : (x_1, x_2, \ldots, x_p) \in S : g_i(\underline{x}) \le y_i,\ i = 1, 2, \ldots, k\}$ and $B_{\underline{y}} = \{\underline{x} : (x_1, x_2, \ldots, x_p) \in S : g_i(\underline{x}) = y_i,\ i = 1, 2, \ldots, k\}$.

Example 0.92. *Let $X_1, X_2, \ldots, X_p$ be independent r.v.s with X_i having the p.m.f. (Binomial distribution), that is,*

$$f_i(x) = \begin{cases} \binom{n_i}{x}\theta^x(1-\theta)^{n_i - x},\ x \in \{0, 1, \ldots, n_i\}, \\ 0,\ \textit{otherwise,} \end{cases}$$

$i = 1, 2, \ldots, k$, where $\theta \in (0,1)$ and $n_i \in \{1, 2, \ldots\}$, $i = 1, 2, \ldots, k$ are fixed real constants. Let $Y = X_1 + X_2 + \cdots + X_p$. Find the p.m.f. of Y.

Solutions: The joint p.m.f. of $\underline{X} = (X_1, X_2, \ldots, X_p)$ is

$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{p} f_i(x_i) = \begin{cases} \left(\prod_{i=1}^{p} \binom{n_i}{x_i}\right) \theta^{\sum_{i=1}^p x_i}(1-\theta)^{n - \sum_{i=1}^p x_i},\ \ \underline{x} \in \prod_{i=1}^p \{0, 1, \ldots, n_i\}, \\ 0,\ \text{otherwise,} \end{cases}$$

where $n = \sum_{i=1}^p n_i$. Clearly, $f_Y(y) = \Pr(X_1 + \cdots + X_p = y) = 0$, if $y \neq \{0, 1, \ldots, n\}$. For $y \in \{0, 1, \ldots, n\}$

$$\begin{aligned} f_Y(y) &= \Pr(Y = y) \\ &= \Pr(X_1 + \cdots + X_p = y) \\ &= \sum_{\substack{x_1=0 \\ \\ x_1 + \cdots + x_p = y}}^{n_1} \sum_{x_2=0}^{n_2} \cdots \sum_{x_p=0}^{n_p} \left(\prod_{i=1}^{p} \binom{n_i}{x_i}\right) \theta^{\sum_{i=1}^p x_i}(1-\theta)^{n - \sum_{i=1}^p x_i} \end{aligned}$$

$$= \theta^y(1-\theta)^{n-y}\sum_{\substack{x_1=0\\ x_1+\cdots+x_p=y}}^{n_1}\sum_{x_2=0}^{n_2}\cdots\sum_{x_p=0}^{n_p}\left(\prod_{i=1}^{p}\binom{n_i}{x_i}\right) = \binom{n}{y}\theta^y(1-\theta)^{n-y}.$$

Thus,

$$f_Y(y) = \begin{cases}\binom{n}{y}\theta^y(1-\theta)^{n-y}, \ y \in \{0,1,\dots,n\},\\ 0, \ \text{otherwise.}\end{cases}$$

Exercise: Let $X_1, X_2, \dots, X_p$ be independent r.v.s with X_i having the p.m.f. (Poisson distribution)

$$f_i(x) = \begin{cases}\frac{e^{-\lambda_i}\lambda_i^x}{x!}, \ x \in \{0,1,\dots\},\\ 0, \ \text{otherwise,}\end{cases}$$

where $\lambda_i > 0$, $i = 1,2,\dots,p$ are fixed real constants. Show that the p.m.f. of $Y = X_1 + X_2 + \cdots + X_p$ is

$$f_Y(y) = \begin{cases}\frac{e^{-\lambda}\lambda^y}{y!}, \ y \in \{0,1,\dots\},\\ 0, \ \text{otherwise,}\end{cases}$$

where $\lambda = \sum_{i=1}^{p}\lambda_i$.

Theorem 0.93. *Let $\underline{X} = (X_1, X_2, \dots, X_p)$ be a continuous random vector with support S and joint p.d.f. $f(\cdot)$. Let $S_i \subseteq \mathbb{R}^p$, $i \in \Delta$ be a countable partition of S ($S_i \cap S_j \ \forall \ i \neq j$ and $\cup_{i\in\Delta}S_i = S$). Suppose that $h_j : \mathbb{R}^p \to \mathbb{R}$, $j = 1,2,\dots,p$ are functions such that in each S_i^0, $\underline{h} = (h_1, h_2, \dots, h_p) : S_i^0 \to \mathbb{R}$ is one-to-one with inverse transformation $\underline{h}_i^{-1}(\underline{t}) = (h_{1,i}^{-1}(\underline{t}), h_{2,i}^{-1}(\underline{t}), \dots, h_{p,i}^{-1})(\underline{t})$, $i \in \Delta$, here S_1^0 denotes the interior of S_i, $i \in \Delta$. Further suppose that $h_{j,i}^{-1}(\underline{t})$, $j = 1,2,\dots,p$, $i \in \Delta$ have continuous partial derivatives and the Jacobian determinants*

$$J_i = \begin{vmatrix}\frac{\partial h_{1,i}^{-1}(\underline{t})}{\partial t_1} & \cdots & \frac{\partial h_{1,i}^{-1}(\underline{t})}{\partial t_p}\\ \vdots & \vdots & \vdots\\ \frac{\partial h_{p,i}^{-1}(\underline{t})}{\partial t_1} & \cdots & \frac{\partial h_{p,i}^{-1}(\underline{t})}{\partial t_p}\end{vmatrix} \neq 0, \ \ i \in \Delta.$$

Define $\underline{h}(S_j^0) = \{\underline{h}(\underline{x}) = (h_1(\underline{x}), h_2(\underline{x}), \dots, h_p(\underline{x})) \in \mathbb{R}^p : \underline{x} \in S_j\}$, $j \in \Delta$ and $T_l = h_l(X_1, X_2, \dots, X_p)$, $l = 1,2,\dots,p$. Then the random vector $\underline{T} = (T_1, T_2, \dots, T_p)$ is a continuous random vector with p.d.f.

$$f_{\underline{T}}(\underline{t}) = \sum_{j\in\Delta} f(h_{1,j}^{-1}(\underline{t}), h_{2,j}^{-1}(\underline{t}), \dots, h_{p,j}^{-1}(\underline{t}))|J_j|I_{\underline{h}(S_j^0)}(\underline{t}).$$

Corollary 0.94. *Under the notation and assumption of the above theorem suppose that $\underline{h} = (h_1, h_2, \dots, h_p) : S^0 \to \mathbb{R}^p$ is one-to-one with inverse transformation $\underline{h}^{-1}(\underline{t}) = (h_1^{-1}(\underline{t}), h_2^{-1}(\underline{t}), \dots, h_p^{-1})(\underline{t})$ (say), here S^0 denotes the interior of S. Furthermore suppose that $h_i^{-1}(\underline{t})$, $i = 1,2,\dots,p$ have continuous partial derivatives and the jacobian determinant*

$$J = \begin{vmatrix}\frac{\partial h_1^{-1}(\underline{t})}{\partial t_1} & \cdots & \frac{\partial h_1^{-1}(\underline{t})}{\partial t_p}\\ \vdots & \vdots & \vdots\\ \frac{\partial h_p^{-1}(\underline{t})}{\partial t_1} & \cdots & \frac{\partial h_p^{-1}(\underline{t})}{\partial t_p}\end{vmatrix} \neq 0.$$

Define $\underline{h}(S^0) = \{h(\underline{x}) : \underline{x} \in S\}$ and $T_j = h_j(X_1, X_2, \dots, X_p)$, $j = 1,2,\dots,p$. Then the random vector $\underline{T} = (T_1, T_2, \dots, T_p)$ is a continuous random vector with joint p.d.f.

$$f_{\underline{T}}(\underline{t}) = f(h_1^{-1}(\underline{t}), h_2^{-1}(\underline{t}), \dots, h_p^{-1}(\underline{t}))|J|I_{\underline{h}(S^0)}(\underline{t}).$$

Example 0.95. *Let X_1 and X_2 be iid r.v.s with common p.d.f.* $f(x) = \begin{cases} e^{-x}, \ x > 0, \\ 0, \ \textit{otherwise}. \end{cases}$ *Find the p.d.f. of* $Y = \frac{X_1}{X_1+X_2}$.

Solution: The joint p.d.f. of $\underline{X} = (X_1, X_2)$ is

$$f_{\underline{X}}(x_1, x_2) = f(x_1)f(x_2) = \begin{cases} e^{-(x_1+x_2)}, \ x_1 > 0, \ x_2 > 0, \\ 0, \ \text{otherwise.} \end{cases}$$

Here $S = [0, \infty) \times [0, \infty)$, $S^0 = (0, \infty) \times (0, \infty)$. Define $Z = X_1 + X_2$, $h_1(x_1, x_2) = \frac{x_1}{x_1+x_2}$ and $h_2(x_1, x_2) = x_1 + x_2$. Then $\underline{h} : S^0 \to \mathbb{R}^2$ as 1-1; have $\underline{h} = (h_1, h_2)$. We have

$$h_1(x_1, x_2) = \frac{x_1}{x_1 + x_2} = y, \ h_2(x_1, x_2) = x_1 + x_2 = z \implies x_1 = h_1^{-1}(y, z) = yz \text{ and } x_2 = h_2^{-1}(y, z) = z(1 - y).$$

$\underline{x} \in S^0 \iff x_1 > 0, x_2 > 0 \iff yz > 0, \ z(1-y) > 0 \iff 0 < y < 1, \ z > 0$. Thus, $\underline{h}(S^0) = (0, 1) \times (0, \infty)$,

$$J = \begin{vmatrix} \frac{\partial h_1^{-1}(y,z)}{\partial y} & \frac{\partial h_1^{-1}(y,z)}{\partial z} \\ \frac{\partial h_p^{-1}(y,z)}{\partial y} & \frac{\partial h_p^{-1}(y,z)}{\partial z} \end{vmatrix} = \begin{vmatrix} z & y \\ -z & 1 - y \end{vmatrix} = z.$$

Thus the joint p.d.f. of (Y, Z) is

$$f_{Y,Z}(y, z) = f_{\underline{X}}(yz, z(1 - y))|z|I_{(0,1)\times(0,\infty)}(y, z) = \begin{cases} ze^{-z}, \ \ 0 < y < 1, \ z > 0, \\ 0, \ \ \text{otherwise} \end{cases} = f_Y(y)f_Z(z),$$

where

$$f_Y(y) = \begin{cases} 1, \ 0 < y < 1, \\ 0, \ \text{otherwise,} \end{cases} \ \text{ and } \ f_Z(z) = \begin{cases} ze^{-z}, \ z > 0, \\ 0, \ \text{otherwise.} \end{cases}$$

Thus, Y and Z are independent r.v.s with p.d.f.s given above. In particular the p.d.f. of Y is $f_Y(y) = \begin{cases} 1, \ 0 < y < 1, \\ 0, \ \text{otherwise.} \end{cases}$

Exercise: Let X_1 and X_2 be iid r.v.s with common p.d.f.

$$f(x) = \begin{cases} \frac{1}{2}, \ -2 < x < -1, \\ \frac{1}{6}, \ 0 < x < 3, \\ 0, \ \text{otherwise.} \end{cases}$$

Find the p.d.f. of $Y_1 = |X_1| + |X_2|$.

Hint: Define auxiliary variable $Y_2 = |X_1|$. Here $S = ([-2, -1] \cup [0, 3]) \times ([-2, -1] \cup [0, 3])$ and

$$S^0 = (((-2, -1) \cup (0, 3)) \times ((-2, -1) \cup (0, 3))) = S_1^0 \cup S_2^0 \cup S_3^0 \cup S_4^0,$$

where $S_1^0 = (-2, -1) \times (-2, -1)$, $S_2^0 = (-2, -1) \times (0, 3)$, $S_3^0 = (0, 3) \times (-2, -1)$ and $S_4^0 = (0, 3) \times (0, 3)$. On each S_1^0, $\underline{h}(\underline{x}) = (h_1(x_1, x_2), h_2(x_1, x_2)) = (y_1, y_2) = (|x_1| + |x_2|, |x_1|)$ is 1-1. Now proceed.

0.3.12. Moment Generating Function Technique

Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ be a random vector with p.m.f. / p.d.f. $f_{\underline{X}}(\cdot)$ and let $g : \mathbb{R}^p \to \mathbb{R}^q$ be a given function. Suppose that we seek probability distribution (p.m.f. / p.d.f.) of $\underline{Y} = g(\underline{X})$. Under the m.g.f. technique, we try to identify the m.g.f. $M_{\underline{Y}}(t)$ of random vector $\underline{Y}$ with the m.g.f. of some known distribution on a rectangle containing origin. Then the uniqueness of m.g.f. as stated in the following theorem, ascertains that $\underline{Y}$ has that known distribution.

Theorem 0.96. *Let $\underline{X}$ and $\underline{Y}$ be 1-dimensional random vectors. Suppose that there exists an $h > 0$ such that*

$$M_{\underline{X}}(\underline{t}) = M_{\underline{Y}}(\underline{t}) \; \forall \; \underline{t} \in (-h,h) \times (-h,h) \times \cdots \times (-h,h).$$

Then $\underline{X} \stackrel{d}{=} \underline{Y}$.

0.3.13. Order Statistics

Let $X_1, X_2, \ldots, X_n$ be a random sample (of continuous r.v.s) from a distribution having d.f. F, p.d.f. f and support S. Let $Y_r = r$-th smallest of $X_1, X_2, \ldots, X_n$, $r = 1, 2, \ldots, n$. The Y_r is called the r-th order statistic based on random sample $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are called order statistics based on random sample $X_1, X_2, \ldots, X_n$.

Note that if $X_1, X_2, \ldots, X_n$ are continuous r.v.s then $\Pr(Y_1 < Y_2 < \cdots < Y_n) = 1$ and thus $Y_1, Y_2, \ldots, Y_n$ are uniquely defined with probability one.

Theorem 0.97. *Under the above notation,*

(a) the joint p.d.f. of $\underline{Y} = (Y_1, Y_2, \ldots, Y_n)$ is

$$g(y_1, y_2, \ldots, y_n) = \begin{cases} n! \prod_{i=1}^{n} f(y_i), \; -\infty < y_1 < y_2 < \cdots < y_n < \infty, \\ 0, \; \textit{otherwise}, \end{cases}$$

(b) the marginal p.d.f. of Y_r, $r = 1, 2, \ldots, n$ is

$$g_r(y) = \frac{n!}{(r-1)!(n-r)!} [F(y)]^{r-1} [1 - F(y)]^{n-r} f(y), \; -\infty < y < \infty.$$

Proof. Since $\underline{X} = (X_1, X_2, \ldots, X_n)$ is a continuous random vector $\Pr(Y_1 < Y_2 < \cdots < Y_n) = 1$ (why?). Define $S_n = S \times S \times \cdots \times S$, so that support of $\underline{X} = (X_1, X_2, \ldots, X_n)$ is S_n. Define

$$\begin{aligned} S_1^0 &= \{\underline{x} \in S_n : x_1 < x_2 < \cdots < x_n\}, \\ S_2^0 &= \{\underline{x} \in S_n : x_1 < x_2 < \ldots x_n < x_{n-1}\}, \\ &\vdots \\ S_{n!}^0 &= \{\underline{x} \in S_n : x_n < x_{n-1} < \cdots < x_1\}. \end{aligned}$$

On each S_i^0, $\underline{Y} = (Y_1, Y_2, \ldots, Y_n) = (h_{1,i}(\underline{X}), h_{2,i}(\underline{X}), \ldots, h_{n,i}(\underline{X}))$ is 1-1 with inverse transformation $\underline{h}_i^{-1} = (h_{1,i}^{-1}, h_{2,i}^{-1}, \ldots, h_{n,i}^{-1})$, $i = 1, 2, \ldots, n!$. Note that as a set

$$\{h_{1,i}^{-1}, h_{2,i}^{-1}, \ldots, h_{n,i}^{-1}\} = \{y_1, y_2, \ldots, y_n\}, \;\; i = 1, 2, \ldots, n!.$$

Therefore the Jacobian of inverse transformation in each S_i is ± 1.

$$\underline{h}(S_1^0) = \{\underline{y} \in S_n : y_1 < y_2 < \cdots < y_n\} = T_1, \text{ say.}$$

Then the joint p.d.f. of $\underline{Y} = (Y_1, Y_2, \ldots, Y_n)$ is

$$\begin{aligned} g(\underline{y}) &= \sum_{j=1}^{n!} f_{\underline{X}}(h_{1,j}^{-1}(\underline{y}), \ldots, h_{n,j}^{-1}(\underline{y})) |J_j| I_{\underline{h}(S_j^0)}(\underline{y}) \\ &= \sum_{j=1}^{n!} \left(\prod_{i=1}^{n} f(h_{ij}^{-1}(\underline{y})) \right) |\pm 1| I_{T_j}(\underline{y}) = n! \prod_{i=1}^{n} f(y_i), \; -\infty < y_1 < y_2 < \cdots < y_n < \infty. \end{aligned}$$

(b) The marginal p.d.f. of Y_r is

$$g_r(y) = \int_{-\infty}^{y}\int_{-\infty}^{y_{r-1}}\cdots\int_{-\infty}^{y_2}\int_{y}^{\infty}\int_{y_{r+1}}^{\infty}\cdots\int_{y_{n-1}}^{\infty} n!f(y_1)\cdots f(y_{r-1})f(y)f(y_{r+1})\cdots f(y_n)\mathrm{d}y_n\cdots\mathrm{d}y_{r+1}\mathrm{d}y_1\cdots\mathrm{d}y_r$$
$$= \frac{n!}{(r-1)!(n-r)!}[F(y)]^{r-1}\,[1-F(y)]^{n-r}\,f(y),\ \ -\infty<y<\infty.$$

Similarly, for $1\le r<s\le n$, the joint p.d.f. of (Y_r,Y_s) is

$$f_{Y_r}f_{Y_s}(y,z) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}[F(y)]^{r-1}\,[F(z)-F(y)]^{s-r-1}\,[1-F(y)]^{n-s}\,f(x)f(y),$$

$-\infty<x<y<\infty$. This completes the proof. $\square$

Theorem 0.98. *(a) Let X_1 and X_2 be independent r.v.'s with $X_i\sim GAM(\alpha_i,\theta)$, $\alpha_i>0$, $\theta>0$, $i=1,2$. 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 $Y_1\sim GAM(\alpha_1+\alpha_2,\theta)$ and $Y_2\sim Be(\alpha_1,\alpha_2)$.*

(b) Let X_1 and X_2 be iid $Exp(\theta)$ r.v.'s. Then $Y=\frac{X_1}{X_1+X_2}\sim U(0,1)$.

Proof. (a) The joint p.d.f. of $\underline{X}=(X_1,X_2)$ is

$$f_{\underline{X}}(x_1,x_2)=\prod_{i=1}^{2}f_{X_i}(x_i)=\prod_{i=1}^{2}\left\{\frac{e^{-x_i/\theta}x_i^{\alpha_i-1}}{\theta^{\alpha_i}\Gamma(\alpha_i)}I_{(0,\infty)}(x_i)\right\}=\begin{cases}\dfrac{e^{-(x_1+x_2)/\theta}x_1^{\alpha_1-1}x_2^{\alpha_2-1}}{\theta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)}, & \text{if } x_1>0,\ x_2>0,\\ 0,\ \text{otherwise.}\end{cases}$$

Here, $S_{\underline{X}}=(0,\infty)^2$. Let $h_1(X_1,X_2)=Y_1=X_1+X_2$ and $h_2(X_1,X_2)=Y_2=\frac{X_1}{X_1+X_2}$. Thus $\underline{h}=(h_1,h_2): S_{\underline{X}}\to\mathbb{R}^2$ is $1-1$ with inverse image (h_1^{-1},h_2^{-1}), where

$$h_1^{-1}(y_1,y_2)=y_1y_2,\ h_2^{-1}(y_1,y_2)=y_1(1-y_2),\ J=\begin{vmatrix}y_2 & y_1\\ 1-y_2 & -y_1\end{vmatrix}=-y_1.$$

$\underline{h}^{-1}(\underline{y})\in S_{\underline{X}}\iff y_1y_2>0,\ y_1(1-y_2)>0\iff y_1>0,\ 0<y_2<1\implies h(S_{\underline{X}})=(0,\infty)\times(0,1)$. Thus the joint p.d.f. of $\underline{Y}=(Y_1,Y_2)$ is

$$f_{\underline{Y}}(y_1,y_2)=\frac{e^{-(y_1y_2+y_1(1-y_2))/\theta}(y_1y_2)^{\alpha_1-1}(y_1(1-y_2))^{\alpha_2-1}}{\theta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)}|-y_1|I_{(0,\infty)\times(0,1)}(y_1,y_2)$$
$$=\left\{\frac{e^{-y_1/\theta}y_1^{\alpha_1+\alpha_2-1}}{\Gamma(\alpha_1+\alpha_2)}I_{(0,\infty)}(y_1)\right\}\left\{\frac{1}{B(\alpha_1,\alpha_2)}y_2^{\alpha_1-1}(1-y_2)^{\alpha_2-1}I_{(0,1)}(y_2)\right\}=f_{Y_1}(y_1)f_{Y_2}(y_2),$$

where $Y_1\sim GAM(\alpha_1+\alpha_2,\theta)$ and $Y_2\sim Be(\alpha_1,\alpha_2)$. Clearly Y_1 and Y_2 are independent. Part (b) can similarly be proved. $\square$

0.4. Special Multivariate Distribution

0.4.1. Multinomial Distribution (A generalization of binomial distribution)

$\mathscr{E}$: a random experiment whose each trial results in one (and only one) of $p+1$ possible outcomes $E_1,E_2,\ldots,E_{p+1}$ where $E_i\cap E_j=\phi$ and $\sum_{i=1}^{p+1}E_i=\Omega$. Let $P(E_i)=\theta_i\in(0,1)$, $i=1,2\ldots,p$ and $\sum_{i=1}^{p}\theta_i<1$ so that $P(E_{p+1})=1-\sum_{i=1}^{p}\theta_i\in(0,1)$.

Consider n independent trials of $\mathscr{E}$. Define X_i = the number of times E_i occurs in n trials, $i = 1, 2 \ldots, p + 1$. Then $\sum_{i=1}^{p+1} X_i = n$, that is, $X_{p+1} = n - \sum_{i=1}^{n} X_i$. One may interested in probability distribution of $\underline{X} = (X_1, X_2, \ldots, X_p)$. We have

$$S_{\underline{X}} = \{\underline{x} = (x_1, x_2, \ldots, x_p) : x_i \in \{0, 1, \ldots, n\}, i = 1, 2, \ldots, p, \ \sum_{i=1}^{n} x_i \leq n\}$$

and

$$\begin{aligned} f_{\underline{X}}(x_1, x_2, \ldots, x_p) &= P(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) \\ &= \begin{cases} \dfrac{n!}{x_1! x_2! \cdots x_p! (n - \sum_{i=1}^{p} x_i)!} \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_p^{x_p} (1 - \sum_{i=1}^{p} \theta_i)^{n - \sum_{i=1}^{p} x_i}, & \underline{x} \in S_{\underline{X}}, \\ 0, \text{ otherwise.} \end{cases} \end{aligned}$$

$\longrightarrow$ Multinomial distribution with n trials and cell probabilities $\theta_1, \ldots, \theta_p$ (denoted by $Mult(n, \theta_1, \theta_2, \ldots, \theta_p)$)$\longrightarrow$ a family of distribution with varying $n \in \mathbb{N}$ and $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_p) \in \Theta = \{(t_1, t_2, \ldots, t_p) : 0 < t_i < 1, \ i = 1, 2, \ldots, p$ and $\sum_{i=1}^{p} t_i < 1\}$.

Remark 0.99. *For $p = 1$, $Mult(n, \theta_1)$ distribution is the same as $Bin(n, \theta_1)$ distribution.*

Theorem 0.100. *Suppose that $\underline{X} = (X_1, X_2, \ldots, X_p) \sim Mult(n, \theta_1, \theta_2, \ldots, \theta_p)$, where $n \in \mathbb{N}$ and $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_p) \in \Theta$. Then*

(a) $X_i \sim Bin(n, \theta_i)$, $i = 1, 2, \ldots, p$,

(b) $X_i + X_j \sim Bin(n, \theta_i + \theta_j)$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p$, $i \neq j$,

(c) $E(X_i) = n\theta_i$ and $\operatorname{Var}(X_i) = n\theta_i(1 - \theta_i)$, $i = 1, 2, \ldots, p$,

(d) $\operatorname{Cov}(X_i, X_j) = -n\theta_i\theta_j$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, p$, $i \neq j$.

Proof. (a) Fix $i \in \{1, 2, \ldots, p\}$. A given trial of the experiment treat the occurrence of E_i as success and its non-occurrence (that is, occurrence of any other $E_j, j \neq i$) as failure. Then we have a sequence of independent Bernoulli trials with probability of success in each trial as $P(E_i) = \theta_i$. Thus

$$X_i = \text{the number of times } E_i \text{ occurs in } n \text{ Bernoulli trials} \sim Bin(n, \theta_i), i = 1, 2, \ldots, p.$$

(b) Fix $i, j \in \{1, 2, \ldots, p\}$ $i \neq j$. In any given trial of $\mathscr{E}$ consider occurrence of E_i or E_j as success and occurrence of any other E_l $(l \neq i, j)$ as failure. Then we have a sequence of n Bernoulli trials with success probability in each trials as $P(E_i \cup E_j) = \theta_i + \theta_j$,

$$X_i + X_j = \text{the number of success occurs in } n \text{ Bernoulli trials} \sim Bin(n, \theta_i + \theta_j).$$

(c) Obvious.

(d)

$$\begin{aligned} &\operatorname{Var}(X_i + X_j) = n(\theta_i + \theta_j)(1 - \theta_i - \theta_j) \\ \Longrightarrow\ &\operatorname{Var}(X_i) + \operatorname{Var}(X_j) + 2\operatorname{Cov}(X_i, X_j) = n(\theta_i + \theta_j)(1 - \theta_i - \theta_j) \\ \Longrightarrow\ &n\theta_i(1 - \theta_i) + n\theta_j(1 - \theta_i) + 2\operatorname{Cov}(X_i, X_j) = n(\theta_i + \theta_j)(1 - \theta_i - \theta_j) \implies \operatorname{Cov}(X_i, X_j) = -n\theta_i\theta_j. \end{aligned}$$

This completes the proof. $\square$

The m.g.f. of $\underline{X} = (X_1, X_2, \ldots, X_p)$ is given by

$$\begin{aligned}
M_{\underline{X}}(t_1, t_2, \ldots, t_p) &= E(e^{t_1X_1+t_2X_2+\cdots+t_pX_p}) \\
&= \sum_{\substack{x_1=0 \\ x_1+x_2+\cdots+x_p\leq n}}^{n} \sum_{x_2=0}^{n} \cdots \sum_{x_p=0}^{n} e^{t_1x_1+t_2x_2+\cdots+t_px_p} \frac{n!\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_p^{x_p}}{x_1!x_2!\cdots x_p!(n-\sum_{i=1}^{p} x_i)!}(1-\sum_{i=1}^{p}\theta_i)^{n-\sum_{i=1}^{p} x_i} \\
&= \sum_{\substack{x_1=0 \\ x_1+x_2+\cdots+x_p\leq n}}^{n} \sum_{x_2=0}^{n} \cdots \sum_{x_p=0}^{n} \frac{n!(\theta_1e^{t_1})^{x_1}(\theta_2e^{t_2})^{x_2}\cdots(\theta_pe^{t_p})^{x_p}}{x_1!x_2!\cdots x_p!(n-\sum_{i=1}^{p} x_i)!}(1-\sum_{i=1}^{p}\theta_i)^{n-\sum_{i=1}^{p} x_i} \\
&= \left(\theta_1e^{t_1}+\theta_2e^{t_2}+\cdots+\theta_pe^{t_p}+1-\sum_{i=1}^{p}\theta_i\right)^n, \ \underline{t} \in \mathbb{R}^p.
\end{aligned}$$

Remark 0.101. *The last theorem can also be proved using m.g.f. For example (for $i, j \in \{1, 2, \ldots, p\}, i \neq j$)*

$$M_{X_i+X_j}(\underline{t}) = M_{\underline{X}}(0, \ldots, 0, \underset{\textit{ith position}}{t}, 0, \ldots, 0, \underset{\textit{jth position}}{t}, 0, \ldots, 0) = ((\theta_i+\theta_j)e^t + 1 - \theta_i - \theta_j), \ \ \underline{t} \in \mathbb{R}^p.$$

0.5. Bivariate Normal Distribution

Definition 0.102. *A bivariate r.v. $\underline{X} = (X_1, X_2)$ is said to follow bivariate normal distribution $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ if for some $-\infty < \mu_i < \infty$, $\sigma_i > 0$, $i = 1, 2$ and $-1 < \rho < 1$, the joint p.d.f. of $\underline{X} = (X_1, X_2)$ is given by*

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]}, \ \ \infty < x_i < \infty, \ i = 1, 2.$$

Clearly, $f_{X_1,X_2}(x_1, x_2) \geq 0 \ \forall \ \underline{x} \in \mathbb{R}^2$ and on making the transformation $\frac{x_1-\mu_1}{\sigma_1} = z_1$ and $\frac{x_2-\mu_2}{\sigma_2} = z_2$ (so that $J = \sigma_1\sigma_2$) we have

$$\begin{aligned}
\int_{\mathbb{R}^2} f_{X_1,X_2}(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}(z_1^2-2\rho z_1z_2+z_2^2)}\mathrm{d}z_1\mathrm{d}z_2 \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{z_2^2}{2}} \left\{\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}(z_1-\rho z_2)^2}\mathrm{d}z_2\right\}\mathrm{d}z_1 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z_2^2}{2}}\mathrm{d}z_2 = 1 \implies f_{X_1,X_2}(x_1, x_2) \text{ is a p.d.f.}
\end{aligned}$$

Note that for $\underline{x} = (x_1, x_2) \in \mathbb{R}^2$

$$\begin{aligned}
f_{X_1,X_2}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1} - \rho\frac{x_2-\mu_2}{\sigma_2}\right)^2 + (1-\rho^2)\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x_1-\left(\mu_1+\frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2)\right)\right]^2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2} \\
&= f_{X_1|X_2}(x_1|x_2)f_{X_2}(x_2)
\end{aligned}$$

$$\implies X_1|X_2 = x_2 \sim N\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2), \sigma_1^2(1-\rho^2)\right), \ X_2 \sim N(\mu_2, \sigma_2^2).$$

By symmetry

$$X_2|X_1 = x_1 \sim N\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x_1-\mu_1), \sigma_2^2(1-\rho^2)\right), \ X_1 \sim N(\mu_1, \sigma_1^2).$$

Clearly, $\mu_1 = E(X_1),\ \mu_2 = E(X_2),\ \sigma_1^2 = \text{Var}(X_1)$ and $\sigma_2^2 = \text{Var}(X_2)$.

$$\begin{aligned}
\text{m.g.f. } M_{X_1,X_2}(t_1,t_2) &= E\left(e^{t_1X_1+t_2X_2}\right)\\
&= E\left(E\left(e^{t_1X_1+t_2X_2}|X_2\right)\right) = E\left(e^{t_2X_2}E\left(e^{t_1X_1}|X_2\right)\right),\ \underline{t}=(t_1,t_2)\in\mathbb{R}^2,\\
E\left(e^{t_1X_1}|X_2\right) &= \text{m.g.f. of conditional distribution } X_1|X_2 \text{ at point } t_2\\
&= e^{\{\mu_1+\frac{\rho\sigma_1}{\sigma_2}(X_2-\mu_2)\}t_1+\frac{\sigma_1^2(1-\rho^2)t_1^2}{2}}\\
M_{X_1,X_2}(t_1,t_2) &= e^{\{\mu_1-\frac{\rho\sigma_1}{\sigma_2}\mu_2\}t_1+\frac{\sigma_1^2(1-\rho^2)t_1^2}{2}}E\left[e^{t_2X_2}e^{\frac{\rho\sigma_1}{\sigma_2}t_1X_2}\right]\\
&= e^{\{\mu_1-\frac{\rho\sigma_1}{\sigma_2}\mu_2\}t_1+\frac{\sigma_1^2(1-\rho^2)t_1^2}{2}}M_{X_2}\left[t_2+\frac{\rho\sigma_1}{\sigma_2}t_1\right]\\
&= e^{\{\mu_1-\frac{\rho\sigma_1}{\sigma_2}\mu_2\}t_1+\frac{\sigma_1^2(1-\rho^2)t_1^2}{2}}e^{\mu_2\{t_2+\frac{\rho\sigma_1}{\sigma_2}t_1\}+\frac{\sigma_2^2}{2}(t_2+\frac{\rho\sigma_1}{\sigma_2}t_1)^2}\\
&= e^{\mu_1t_1+\mu_2t_2+\frac{\sigma_1^2t_1^2}{2}+\frac{\sigma_2^2t_2^2}{2}+\rho\sigma_1\sigma_2t_1t_2},\ \underline{t}=(t_1,t_2)\in\mathbb{R}^2.
\end{aligned}$$

Thus we have the following theorem.

Theorem 0.103. *Suppose that* $\underline{X} = (X_1,X_2) \sim N_2(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, $-\infty < \mu_i < \infty$, $\sigma_i > 0$, $i = 1,2$ *and* $-1 < \rho < 1$. *Then*

(a) $X_1 \sim N(\mu_1,\sigma_1^2)$ *and* $X_2 \sim N(\mu_2,\sigma_2^2)$*;*

(b) For fixed $x_2 \in \mathbb{R}$, $X_1|X_2 = x_2 \sim N\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2-\mu_2), \sigma_1^2(1-\rho^2)\right)$ *and for fixed* $x_1 \in \mathbb{R}$, $X_2|X_1 = x_1 \sim N\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x_1-\mu_1), \sigma_2^2(1-\rho^2)\right)$*;*

(c) The m.g.f. of $\underline{X} = (X_1,X_2)$ *is*

$$M_{X_1,X_2}(t_1,t_2) = e^{\mu_1t_1+\mu_2t_2+\frac{\sigma_1^2t_1^2}{2}+\frac{\sigma_2^2t_2^2}{2}+\rho\sigma_1\sigma_2t_1t_2},\ \underline{t}=(t_1,t_2)\in\mathbb{R}^2;$$

(d) $\rho(X_1,X_2) = \text{Corr}(X_1,X_2) = \rho$*;*

(e) X_1 *and* X_2 *are independent iff* $\rho = 0$*;*

(f) For real constants C_1 *and* C_2 *such that* $(C_1,C_2) \neq (0,0)$

$$C_1X_1 + C_2X_2 \sim N(C_1\mu_1 + C_2\mu_2, C_1^2\sigma_1^2 + C_2^2\sigma_2^2 + 2\rho C_1C_2\sigma_1\sigma_2)$$

Proof. (a)-(c) Already done.

(d) For $\underline{t} = (t_1,t_2) \in \mathbb{R}^2$

$$\begin{aligned}
\psi_{X_1,X_2}(t_1,t_2) &= \ln M_{X_1,X_2}(t_1,t_2) = \mu_1t_1+\mu_2t_2+\frac{\sigma_1^2t_1^2}{2}+\frac{\sigma_2^2t_2^2}{2}+\rho\sigma_1\sigma_2t_1t_2\\
\frac{\partial}{\partial t_1}\psi_{X_1,X_2}(t_1,t_2) &= \mu_1 + 2\sigma_1^2t_1 + \rho\sigma_1\sigma_2t_2\\
\frac{\partial^2}{\partial t_2\partial t_1}\psi_{X_1,X_2}(t_1,t_2) &= \rho\sigma_1\sigma_2\\
\implies \text{Cov}(X_1,X_2) &= \left[\frac{\partial^2}{\partial t_2\partial t_1}\psi_{X_1,X_2}(t_1,t_2)\right]_{\underline{t}=\underline{0}} = \rho\sigma_1\sigma_2\\
\implies \rho(X_1,X_2) &= \text{Corr}(X_1,X_2) = \frac{\text{Cov}(X_1,X_2)}{\sqrt{\text{Var}(X_1)\,\text{Var}(X_2)}} = \rho.
\end{aligned}$$

(e) Obviously, if X_1 and X_2 are independent then $\rho = \mathrm{Corr}(X_1, X_2) = 0$. Now suppose that $\rho = 0$. Then

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}[x_1-\mu_1]^2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x_2-\mu_2)^2} \\ &= f_{X_1}(x_1) f_{X_2}(x_2) \;\forall\; \underline{x} = (x_1,x_2) \in \mathbb{R}^2 \implies X_1 \text{ and } X_2 \text{ are independent.} \end{aligned}$$

(f) Let $Y = C_1X_1 + C_2X_2$. Then

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E(e^{t(C_1X_1+C_2X_2)}) = M_{X_1,X_2}(tC_1, tC_2) \\ &= \exp\left\{C_1t\mu_1 + C_2t\mu_2 + \frac{C_1^2t^2\sigma_1^2}{2} + \frac{C_2^2t^2\sigma_2^2}{2} + \rho t^2 C_1C_2\sigma_1\sigma_2\right\} \\ &= \exp\left\{(C_1\mu_1 + C_2\mu_2)t + (C_1^2\sigma_1^2 + C_2^2\sigma_2^2 + 2\rho C_1C_2\sigma_1\sigma_2)\frac{t^2}{2}\right\} \\ &\longrightarrow \text{m.g.f. of } N(C_1\mu_1 + C_2\mu_2, C_1^2\sigma_1^2 + C_2^2\sigma_2^2 + 2\rho C_1C_2\sigma_1\sigma_2). \end{aligned}$$

This completes the proof. □

Theorem 0.104. *Let* $\underline{X} = (X_1, X_2)$ *be a bivariate r.v. with* $E(X_i) = \mu_i \in (-\infty, \infty)$, $\mathrm{Var}(X_i) = \sigma_i^2$, *(*$\sigma_i > 0$*),* $i = 1, 2$ *and* $\mathrm{Corr}(X_1, X_2) = \rho \in (-1, 1)$. *Then* $\underline{X} \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ *iff for any* $\underline{t} = (t_1, t_2) \in \mathbb{R}^2 - \{0\}$, $Y = t_1X_1 + t_2X_2 \sim N(t_1\mu_1 + t_2\mu_2, t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2\rho t_1t_2\sigma_1\sigma_2)$.

Proof. Let $\underline{X} \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Then by (f) of last theorem

$$Y = t_1X_1 + t_2X_2 \sim N(t_1\mu_1 + t_2\mu_2, t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2\rho t_1t_2\sigma_1\sigma_2) \;\forall\; \underline{t} \in \mathbb{R}^2 - \{0\}.$$

Conversely, suppose that for all $\underline{t} = (t_1, t_2) \in \mathbb{R}^2 - \{0\}$, $Y = t_1X_1 + t_2X_2 \sim N(t_1\mu_1 + t_2\mu_2, t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2\rho t_1t_2\sigma_1\sigma_2)$. Then for $\underline{t} = (t_1, t_2) \in \mathbb{R}^2 - \{0\}$

$$\begin{aligned} M_{X_1,X_2}(t_1,t_2) &= E(e^{t_1X_1+t_2X_2}) \\ &= M_Y(1) = \exp\left\{(t_1\mu_1 + t_2\mu_2) + (t_1^2\sigma_1^2 + t_2^2\sigma_2^2 + 2\rho t_1t_2\sigma_1\sigma_2)\frac{1}{2}\right\} \\ &\longrightarrow \text{m.g.f. of } \; N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \implies \underline{X} = (X_1, X_2) \sim N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho). \end{aligned}$$

This completes the proof. □

Theorem 0.105. *Let* $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ *be a random sample from* $N(\mu, \sigma^2)$ *distribution, where* $-\infty < \mu < \infty$ *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) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$;

(ii) $\bar{X}$ *and* S^2 *are independent r.v.'s;*

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

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

Proof. (i) Follows from last theorem by taking $k = n$, $a_i = \frac{1}{n}$, $\mu_i = \mu$, $\sigma_i^2 = \sigma$, $i = 1, 2, \ldots, n$.

(ii) Let $Y_i = X_i - \bar{X}$, $i = 1, 2, \ldots, n$ and let $\underline{Y} = (Y_1, Y_2, \ldots, Y_n)$. Then

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

$$(n-1)S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} Y_i^2 \quad \text{(a function of } \underline{Y}\text{)}$$

The joint m.g.f. of $(\underline{Y}, \bar{X})$ is given by

$$M_{\underline{Y},\bar{X}}(\underline{t}) = E\left(e^{\sum_{i=1}^{n} t_i Y_i + t_{n+1}\bar{X}}\right), \quad \underline{t} = (t_1, t_2, \ldots, t_n, t_{n+1}) \in \mathbb{R}^{n+1}$$

$$\begin{aligned}
\sum_{i=1}^{n} t_i Y_i + t_{n+1}\bar{X} &= \sum_{i=1}^{n} t_i(X_i - \bar{X}) + t_{n+1}\bar{X} \\
&= \sum_{i=1}^{n} t_i X_i + \frac{(t_{n+1} - \sum_{i=1}^{n} t_i)}{n} \sum_{i=1}^{n} X_i \\
&= \sum_{i=1}^{n}\left(t_i - \bar{t} + \frac{t_{n+1}}{n}\right) X_i, \text{ where } \bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i \\
&= \sum_{i=1}^{n} u_i X_i, \text{ where } u_i = t_i - \bar{t} + \frac{t_{n+1}}{n}, \ i = 1(1)n.
\end{aligned}$$

Then $\sum_{i=1}^{n} u_i = t_{n+1}$ and $\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n}(t_i - \bar{t})^2 + \frac{t_{n+1}^2}{n}$.

$$\begin{aligned}
M_{\underline{Y},\bar{X}}(\underline{t}) &= E\left(e^{\sum_{i=1}^{n} u_i X_i}\right) \\
&= \prod_{i=1}^{n} M_{X_i}(u_i) \\
&= \prod_{i=1}^{n} e^{\mu u_i + \frac{1}{2}\sigma^2 u_i^2} \\
&= e^{\mu \sum_{i=1}^{n} u_i + \frac{\sigma^2}{2}\sum_{i=1}^{n} u_i^2} \\
&= e^{\mu t_{n+1} + \frac{\sigma^2}{2}\{\sum_{i=1}^{n}(t_i - \bar{t})^2 + \frac{t_{n+1}^2}{n}\}} = \left\{e^{\mu t_{n+1} + \frac{\sigma^2 t_{n+1}^2}{2n}}\right\}\left\{e^{\frac{\sigma^2}{2}\sum_{i=1}^{n}(t_i - \bar{t})^2}\right\}
\end{aligned}$$

$$\begin{aligned}
& M_{\underline{Y}}(t_1, t_2, \ldots, t_n) = M_{\underline{Y},\bar{X}}(t_1, t_2, \ldots, t_n, 0) = e^{\frac{\sigma^2}{2}\sum_{i=1}^{n}(t_i - \bar{t})^2}, (t_1, t_2, \ldots, t_n) \in \mathbb{R}^n \\
& M_{\bar{X}}(t_{n+1}) = M_{\underline{Y},\bar{X}}(0, \ldots, 0, t_{n+1}) = e^{\mu t_{n+1} + \frac{\sigma^2 t_{n+1}^2}{2n}}, \ t_{n+1} \in \mathbb{R} \\
\Longrightarrow\ & M_{\underline{Y},\bar{X}}(\underline{t}) = M_{\underline{Y}}(t_1, t_2, \ldots, t_n) M_{\bar{X}}(t_{n+1}), \forall\, \underline{t} = (t_1, t_2, \ldots, t_n, t_{n+1}) \in \mathbb{R}^{n+1} \\
\Longrightarrow\ & \underline{Y} \text{ and } \bar{X} \text{are independent} \\
\Longrightarrow\ & \sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ and } \bar{X} \text{ are independent.}
\end{aligned}$$

(iii) Let $Z_i = \frac{X_i - \mu}{\sigma}$, $i = 1, 2, \ldots, n$. Then $Z_1, Z_2, \ldots, Z_n$ are iid $N(0, 1)$ r.v.'s. Also let $Z = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$ (using (i)). Let

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

Then by (ii), W and T are independent r.v.s. Also $W \sim \chi^2_1$ and $V = \sum_{i=1}^n Z_i^2 \sim \chi^2_n$.

$$\begin{aligned} V = \sum_{i=1}^n Z_i^2 &= \frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\mu)^2 \\ &= \frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\bar{X}+\bar{X}-\mu)^2 \\ &= \frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\bar{X})^2 + \frac{n(\bar{X}-\mu)^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2} = T+W. \end{aligned}$$

This implies

$$\begin{aligned} & M_V(t) = M_T(t)M_W(t) \\ \implies & M_T(t) = \frac{M_V(t)}{M_W(t)} = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-(n-1)/2},\ \ t<\frac{1}{2} \to \text{m.g.f. of } \chi^2_{n-1} \\ \implies & T = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}. \end{aligned}$$

(iv) $T = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_v$, where $v = n-1$. Thus

$$\begin{aligned} E(T^s) &= \int_0^\infty t^s \frac{1}{2^{v/2}\Gamma(v/2)} e^{-t/2} t^{v/2-1}\mathrm{d}t \\ &= \frac{1}{2^{v/2}\Gamma(v/2)}\int_0^\infty e^{-t/2} t^{\frac{v+2s}{2}-1}\mathrm{d}t = \frac{2^{\frac{v+2s}{2}}\Gamma\left(\frac{v+2s}{2}\right)}{2^{v/2}\Gamma(v/2)} = \frac{2^s\Gamma\left(\frac{v+2s}{2}\right)}{\Gamma(v/2)},\ s > -\frac{v}{2}. \end{aligned}$$

This implies

$$\begin{aligned} & \frac{(n-1)^s}{\sigma^{2s}}E(S^{2s}) = \frac{2^s\Gamma\left(\frac{v+2s}{2}\right)}{\Gamma(v/2)} \\ \implies & E(S^r) = \left(\frac{2}{n-1}\right)^{r/2}\frac{\Gamma\left(\frac{v+r}{2}\right)}{\Gamma(v/2)}\sigma^r, \\ \implies & E(S^r) = \left(\frac{2}{n-1}\right)^{r/2}\frac{\Gamma\left(\frac{n-1+r}{2}\right)}{\Gamma((n-1)/2)}\sigma^r,\ \ r>0 \\ \implies & E(S) = \left(\frac{2}{n-1}\right)^{1/2}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma((n-1)/2)}\sigma \\ \implies & E(S^2) = \left(\frac{2}{n-1}\right)\frac{\Gamma\left(\frac{n-1}{2}+1\right)}{\Gamma((n-1)/2)}\sigma^2 = \sigma^2 \\ \implies & E(S^4) = \left(\frac{2}{n-1}\right)^2\frac{\Gamma\left(\frac{n-1}{2}+2\right)}{\Gamma((n-1)/2)}\sigma^4 = \frac{n+1}{n-1}\sigma^4 \\ & \mathrm{Var}(S^2) = E(S^4) - (E(S^2))^2 = \frac{2\sigma^4}{n-1}. \end{aligned}$$

This completes the proof. □

Remark 0.106. *Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having p.m.f. / p.d.f. f. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\bar{X})^2$. Let $E(X_1)=\mu$ and $\mathrm{Var}(X_1)=\sigma^2$. Then $E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^n X_i\right) =$*

μ, $\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}$.

$$\begin{aligned}
E\left[(n-1)S^2\right] &= E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] \\
\implies (n-1)E(S^2) &= E\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right] \\
&= \sum_{i=1}^{n} E(X_i^2) - nE(\bar{X}^2) \\
&= n\left[E(X_1^2) - E(\bar{X}^2)\right] \\
&= n\left[\text{Var}(X_1) + (E(X_1))^2 - \text{Var}(\bar{X}) - (E(\bar{X}))^2\right] \\
&= n(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2) = (n-1)\sigma^2 \implies E(S^2) = \sigma^2.
\end{aligned}$$

For this reason $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ *is called sample variance and not* $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$. *Note that* $E(S_1^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right) = \frac{n-1}{n}\sigma^2 < \sigma^2$, *i.e.,* S_1^2 *underestimates* σ^2.

0.6. Distributions Based on Sampling from Normal Distribution

Definition 0.107. *(a) For a positive integer* m, *a random variable* 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*

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

(b) For positive integers n_1 *and* n_2 *a random variable* X *is said to have the Snedecor* F *distribution with* (n_1, n_2) *degrees of freedom (written as* $X \sim F_{n_1,n_2}$*) if its p.d.f. is given by*

$$f(x|n_1,n_2) = \begin{cases} \dfrac{\left(\dfrac{n_1}{n_2}\right)\left(\dfrac{n_1 x}{n_2}\right)^{\frac{n_1}{2}-1}}{B(n_1/n_2, n_2/2)}\left(1+\frac{n_1 x}{n_2}\right)^{-(n_1+n_2)}, & x > 0, \\ 0, \; \textit{otherwise}. \end{cases}$$

Remark 0.108. *(a) Note that*

$$\begin{aligned}
X \sim t_m &\implies f(x|m) = f(-x|m), \;\forall\, x \\
&\implies X \overset{d}{=} -X \\
&\implies \textit{distribution of } X \textit{ is symmetric about } 0 \implies m_e = 0 \textit{ and } E(X) = 0, \textit{ provided it exists}.
\end{aligned}$$

(b) $X \sim t_m \implies f(x|m) \uparrow m(-\infty, 0), \downarrow (0,\infty) \implies m_0 = 0.$

(c) t_1 *distribution is nothing but Cauchy distribution with p.d.f.*

$$f(x|1) = \frac{1}{\pi}\frac{1}{1+x^2}, \; -\infty < x < \infty \implies E(X) \textit{ does not exist}.$$

(d) Let $X \sim F_{n_1,n_2}$. *Then,*

$$f(x|n_1,n_2) = \begin{cases} \dfrac{\frac{n_1}{n_2}}{B(n_1/2, n_2/2)}\left(\dfrac{\frac{n_1 x}{n_2}}{1+\frac{n_1 x}{n_2}}\right)^{\frac{n_1}{2}-1}\left(\dfrac{1-\frac{n_1 x}{n_2}}{1+\frac{n_1 x}{n_2}}\right)^{n_2-1}\left(1+\dfrac{n_1 x}{n_2}\right)^{-2}, & x > 0, \\ 0, \; \textit{otherwise}. \end{cases}$$

$$\implies Y = \frac{\frac{n_1 X}{n_2}}{1 + \frac{n_1 X}{n_2}} \sim Be\,(n_1/2, n_2/2)\,.$$

Theorem 0.109. *(a) Let $Z \sim N(0,1)$ and let $Y \sim \chi^2_m$, $m \in \{1, 2, \dots\}$ be independent random variables. Then*

$$T = \frac{Z}{\sqrt{Y/m}} \sim t_m.$$

(b) For positive integers n_1 and n_2, let $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ be independent random variables, then

$$U = \frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}.$$

(c) Let $X \sim t_m$. Then $E(X^2)$ is not finite if $r \in \{m, m+1, \dots\}$. For $r \in \{1, 2, \dots, m-1\}$ $(m \geq r+1)$

$$E(X^r) = \begin{cases} 0, \text{ if } r \text{ is odd}, \\ \dfrac{m^{r/2} r! \Gamma((m-r)/2)}{2^r (r/2)! \Gamma(m/2)}, \text{ if } r \text{ is even}. \end{cases}$$

(d) If $X \sim t_m$ then

$$\begin{aligned} \textit{Mean} = \mu_1' = E(X) &= 0, \;\; m = 2, 3, \dots, \\ \mathrm{Var}(X) = \mu_2 = E((X - \mu_1')^2) &= \frac{m}{m-2}, \;\; m \in \{3, 4, \dots\}, \\ \textit{Coefficient of skewness} = \beta_1 &= 0, \; m = 4, 5, 6, \dots, \\ \textit{Kurtosis} = \nu_1 &= \frac{3(m-2)}{m-4}, \;\; m \in \{5, 6, \dots\}. \end{aligned}$$

(e) Let n_1, n_2 and r be positive integers, and let $X \sim F_{n_1,n_2}$. Then, for $n_2 \in \{1, 2, \dots, 2r\}$ and $r \geq \frac{n_2}{2}$, it follows that $E(X^r)$ is not finite. For $n_2 \in \{2r+1, 2r+2, \dots\}$ and $r \geq \frac{n_2 - 1}{2}$, we have

$$E(X^r) = \left(\frac{n_2}{n_1}\right)^r \prod_{i=1}^{r} \left(\frac{n_1 + 2(i-1)}{n_2 - 2i}\right).$$

(f) If $X \sim F_{n_1,n_2}$ then

$$\begin{aligned} \textit{Mean} = \mu_1' = E(X) &= \frac{n_2}{n_2 - 2}, \text{ if } n_2 \in \{3, 4, \dots\}, \\ \mathrm{Var}(X) = \mu_2 = E((X - \mu_1')^2) &= \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_1-2)^2(n_2-4)}, \;\; \text{if } n_2 \in \{5, 6, \dots\}, \\ \textit{Coefficient of skewness} = \beta_1 &= \frac{2(2n_1 + n_2 - 2)}{n_2 - 6} \sqrt{\frac{2(n_2-4)}{n_1(n_1+n_2-2)}}, \;\; n_2 \in \{7, 8, \dots\}, \\ \textit{Kurtosis} = \nu_1 &= \frac{12\left[(n_2-2)^2(n_2-4) + n_1(n_1+n_2-2)(5n_2-22)\right]}{n_1(n_2-6)(n_2-8)(n_1+n_2-2)}. \end{aligned}$$

Proof. (a) The joint p.d.f. of (Y, Z) is given by

$$f_{Y,Z}(y,z) = f_Y(y) f_Z(z) = \begin{cases} \dfrac{1}{2^{(m+1)/2}\Gamma(m/2)\sqrt{\pi}} e^{-\frac{y+z^2}{2}} y^{\frac{m}{2}-1}, \; y > 0, \; -\infty < z < \infty, \\ 0, \text{ otherwise}. \end{cases}$$

Let $U = \sqrt{\frac{Y}{m}}$. $S_{Y,Z} = (0,\infty)\times\mathbb{R}$. Let $\underline{h} = (h_1,h_2) : (0,\infty)\times\mathbb{R}\to\mathbb{R}^2$ where $h_1(y,z) = \frac{z}{\sqrt{y/m}}$ and $h_2(y,z) = \sqrt{y/m}$. The transformation $\underline{h} : S_{Y,Z}\to\mathbb{R}$ is 1-1 with inverse transformation $\underline{h}^{-1} = (h_1^{-1}, h_2^{-1})$, where $h_1^{-1}(t,u) = mu^2$, $h_2^{-1}(t,u) = tu$, $J = \begin{vmatrix} 0 & 2mu \\ u & t \end{vmatrix} = -2mu^2$.

$$\underline{h}(S_{Y,Z}) = \{(t,u) : mu^2 > 0,\ -\infty < tu < \infty\} = \{(t,u) : u > 0,\ t\in\mathbb{R}\} = \mathbb{R}\times(0,\infty).$$

The joint p.d.f. of (T,U) is given by

$$\begin{aligned} f_{T,U}(t,u) &= f_{Y,Z}(h_1^{-1}(t,u), h_2^{-1}(t,u))|J| I_{h(S_{Y,Z})}(t,u) \\ &= \frac{1}{2^{(m+1)/2}\Gamma(m/2)\sqrt{\pi}} e^{-\frac{mu^2+t^2u^2}{2}} (mu^2)^{\frac{m}{2}-1} |2mu^2| I_{\mathbb{R}\times(0,\infty)}(t,u) \\ &= \frac{m^{m/2}u^m}{2^{(m-1)/2}\Gamma(m/2)\sqrt{\pi}} e^{-\frac{(m+t^2)u^2}{2}} I_{\mathbb{R}}(t) I_{(0,\infty)}(u). \end{aligned}$$

The marginal p.d.f. of T is

$$\begin{aligned} f_T(t) &= \int_{-\infty}^{\infty} f_{T,U}(t,u)\mathrm{d}u \\ &= \frac{m^{m/2}}{2^{(m-1)/2}\Gamma(m/2)\sqrt{\pi}} \int_0^{\infty} u^m e^{-\frac{(m+t^2)u^2}{2}}\mathrm{d}u \\ &= \frac{1}{\Gamma(m/2)\sqrt{m\pi}} \frac{1}{(1+t^2/m)^{\frac{m+1}{2}}} \int_0^{\infty} y^{\frac{m-1}{2}} e^{-y}\mathrm{d}y \ \ (u^2 = y) \\ &= \frac{\Gamma(\frac{m+1}{2})}{\Gamma(m/2)\sqrt{m\pi}} \frac{1}{(1+t^2/m)^{\frac{m+1}{2}}}, \quad t\in\mathbb{R} \longrightarrow \text{p.d.f. of } t_m. \end{aligned}$$

(b) The joint p.d.f. of $\underline{X} = (X_1, X_2)$ is given by

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2) = \frac{1}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} e^{-\frac{x_1+x_2}{2}} x_1^{\frac{n_1}{2}-1} x_2^{\frac{n_2}{2}-1} I_{(0,\infty)\times(0,\infty)}(x_1,x_2).$$

Let $V = \frac{X_2}{n_2}$. $S_{\underline{X}} = (0,\infty)\times(0,\infty)$. Consider the transformation: $\underline{h} = (h_1,h_2) : (0,\infty)\times(0,\infty)\to\mathbb{R}$ defined by $h_1(x_1,x_2) = \frac{x_1/n_1}{x_2/n_2}$ and $h_2(x_1,x_2) = \frac{x_2}{n_2}$ so that $U = h_1(X_1,X_2)$ and $V = h_2(X_1,X_2)$.

The transformation $\underline{h} : (0,\infty)\times(0,\infty)\to\mathbb{R}^2$ is 1-1 with inverse transformation $\underline{h}^{-1} = (h_1^{-1}, h_2^{-1})$, where

$$h_1^{-1}(u,v) = n_1uv, h_2^{-1}(u,v) = n_2v,\ \ J = \begin{vmatrix} n_1v & n_1u \\ 0 & n_2 \end{vmatrix} = n_1n_2v,$$

$$\underline{h}(S_{\underline{X}}) = \{(u,v) : n_1uv > 0,\ n_2v > 0\} = \{(u,v) : u > 0,\ v > 0\} = (0,\infty)\times(0,\infty).$$

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

$$\begin{aligned} f_{U,V}(u,v) &= f_{X_1,X_2}(h_1^{-1}(u,v), h_2^{-1}(u,v))|J| I_{\underline{h}(S_{\underline{X}})}(u,v) \\ &= \frac{n_1^{n_1/2} n_2^{n_2/2}}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} e^{-\frac{(n_2+n_1u)v}{2}} u^{\frac{n_1}{2}-1} v^{\frac{n_1+n_2}{2}-1} I_{(0,\infty)}(u) I_{(0,\infty)}(v) \end{aligned}$$

The marginal p.d.f. of U is given by

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\mathrm{d}v$$

$$
\begin{aligned}
&= \frac{n_1^{n_1/2} n_2^{n_2/2}}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} u^{\frac{n_1}{2}-1} \int_0^\infty e^{-\frac{(n_2+n_1u)v}{2}} v^{\frac{n_1+n_2}{2}-1} \mathrm{d}v \\
&= \frac{\Gamma(\frac{n_1+n_2}{2})}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} \frac{(n_1u/n_2)^{\frac{n_1}{2}-1}}{(1+n_1u/n_2)^{\frac{n_1+n_2}{2}}} I_{(0,\infty)} \longrightarrow \text{p.d.f. of } F_{n_1,n_2}.
\end{aligned}
$$

(c) Fix $m \in \{1,2,\dots\}$. Then $X \stackrel{d}{=} \frac{Z}{\sqrt{Y/m}}$ where $Z \sim N(0,1)$ and $Y \sim \chi^2_m$ are independent. This implies that

$$
\begin{aligned}
E(X^r) &= E\left(\frac{Z}{\sqrt{Y/m}}\right)^r = m^{r/2}E(Z^rY^{-r/2}) = m^{r/2}E(Z^r)E(Y^{-r/2}) \ \ (Y \text{ and } Z \text{ are independent}) \\
E(Z^r) &= \begin{cases} 0, \ \text{if } r = 1,3,5,\dots, \\ \frac{r!}{2^{r/2}(r/2)!}, \ \text{if } r = 2,4,6\dots. \end{cases} \\
E(Y^{-r/2}) &= \frac{1}{2^{m/2}(m/2)!} \int_0^\infty y^{\frac{m-r}{2}-1} e^{-y/2} \mathrm{d}y = \infty, \ \text{if } r \geq m.
\end{aligned}
$$

For $r < m$, we have

$$
\begin{aligned}
E(Y^{-r/2}) &= \frac{2^{\frac{m-r}{2}}\Gamma(\frac{m-r}{2})}{2^{m/2}\Gamma(m/2)} = \frac{\Gamma(\frac{m-r}{2})}{2^{r/2}\Gamma(m/2)} \\
\implies E(X^r) &= \begin{cases} 0, \ \text{if } r \text{ is odd and } r < m, \\ \dfrac{m^{r/2} r! \Gamma(m-r)/2}{2^r (r/2)! \Gamma(m/2)}, \ \text{if } r \text{ is even and } \ r < n. \end{cases}
\end{aligned}
$$

(d) Exercise.

(e) Fix $n_1, n_2 \in \mathbb{N}$. Then

$$
X \stackrel{d}{=} \frac{X_1/n_1}{X_2/n_2} = \frac{n_2}{n_1}\frac{X_1}{X_2},
$$

where $X_1 \sim \chi^2_{n_1}$ and $X_2 \sim \chi^2_{n_2}$ are independent. For $r \in \mathbb{N}$,

$$
\begin{aligned}
E(X^r) &= \left(\frac{n_2}{n_1}\right)^r E\left(\frac{X_1^r}{X_2^r}\right) = \left(\frac{n_2}{n_1}\right)^r E(X_1^r) E\left(\frac{1}{X_2^r}\right), \\
E(X_1^r) &= \frac{1}{2^{n_1/2}\Gamma(n_1/2)} \int_0^\infty x^{\frac{n_1+2r}{2}-1} e^{-x/2} \mathrm{d}x \\
&= \frac{2^{\frac{n_1+2r}{2}}\Gamma(\frac{n_1+2r}{2})}{2^{n_1/2}\Gamma(n_1/2)} = \prod_{i=1}^r (n_1 - 2(i-1)), \ \ r \in \{1,2,\dots\}, \\
E\left(\frac{1}{X_2^2}\right) &= \begin{cases} \dfrac{2^{\frac{n_1-2r}{2}}\Gamma(\frac{n_2-2r}{2})}{2^{n_2/2}\Gamma(n_2/2)}, \ n_2 > 2r, \\ \infty, \ \text{if } n_2 \leq 2r \end{cases} = \begin{cases} \prod_{i=1}^r (n_2 - 2i), \ n_2 > 2r, \\ \infty, \ \text{if } n_2 \leq 2r, \end{cases} \\
\implies E(X^r) &= \begin{cases} \left(\frac{n_2}{n_1}\right)^r \prod_{i=1}^r \left(\frac{n_1+2(i-1)}{n_2-2i}\right), \ n_2 > 2r, \\ \infty, \ \text{if } n_2 \leq 2r. \end{cases}
\end{aligned}
$$

(f) Exercise. □

Corollary 0.110. *Let $X_1, X_2, \dots, X_n$ $(n \geq 2)$ be a random sample from $N(\mu, \sigma^2)$ distribution, where $\mu \in \mathbb{R}$ and $\sigma > 0$. Let*

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

denote the sample mean and sample variance, respectively. Then,

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}.$$

Proof. We know that

$$\begin{aligned}
&\bar{X} \sim N(\mu, \sigma^2/n) \text{ and } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \text{ independent} \\
\implies &\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1) \text{ and } \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1} \text{ independent} \\
\implies &\frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{S^2/\sigma^2}} \sim t_{n-1}, \text{ that is, } \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}.
\end{aligned}$$

This completes the proof. □

Corollary 0.111. *Let $X_1, X_2, \ldots, X_m$ ($m \geq 2$) and $Y_1, Y_2, \ldots, Y_n$ ($n \geq 2$) be independent random samples (that is, $\underline{X} = (X_1, X_2, \ldots, X_m)$ and $\underline{Y} = (Y_1, Y_2, \ldots, Y_m)$ are independent) from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distribution, respectively where $\mu_i \in \mathbb{R}$, $i = 1, 2$, and $\sigma_i > 0$, $i = 1, 2$. Let*

$$\bar{X} = \frac{1}{m}\sum_{i=1}^{m} X_i, \;\; \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \;\; S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2, \;\; S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

Then, (a) $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0,1)$,

(b) $\dfrac{1}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(m-1)S_1^2/\sigma_1^2 + (n-1)S_1^2/\sigma_2^2}{m+n-2}}} \sim t_{m+n-2}$,

(c) $\dfrac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{m-1,n-1}$.

Proof. $\bar{X} \sim N(\mu_1, \sigma_1^2/m)$, $\bar{Y} \sim N(\mu_2, \sigma_2^2/n)$, $\frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}$ and $\frac{(n-1)S_2^2}{\sigma_2^2}$ are independent r.v.s. Thus,

$$\begin{aligned}
&\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \sigma_1^2/m + \sigma_2^2/n\right) \\
&\frac{(m-1)S_1^2}{\sigma_1^2} + \frac{(n-1)S_2^2}{\sigma_2^2} \sim t_{m+n-2} \\
\implies &\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0,1) \;\text{ and }\; \frac{1}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{(m-1)S_1^2/\sigma_1^2 + (n-1)S_1^2/\sigma_2^2}{m+n-2}}} \sim t_{m+n-2}.
\end{aligned}$$

This completes the proof. □

Remark 0.112. *(a) Note that*

$$\begin{aligned}
&X \sim t_m \\
\implies &X \stackrel{d}{=} \frac{N(0,1)}{\sqrt{\chi^2_m/m}} \Big\rangle \textit{independent} \\
\implies &X^2 \stackrel{d}{=} \frac{(N(0,1))^2}{\chi^2_m/m} \Big\rangle \textit{independent}
\end{aligned}$$

$$= \left.\frac{\chi_1^2}{\chi_m^2/m}\right\rangle \textit{independent} \stackrel{d}{=} F_{1,m}.$$

Thus, $X \sim t_m \implies X^2 \sim F_{1,m}$.

(b) Note that

$$\begin{aligned} & X \sim F_{n_1,n_2} \\ \implies & X \stackrel{d}{=} \left.\frac{\chi_{n_1}^2/n_1}{\chi_{n_2}^2/n_2}\right\rangle \textit{independent} \\ \implies & \frac{1}{X} \stackrel{d}{=} \left.\frac{\chi_{n_2}^2/n_2}{\chi_{n_1}^2/n_1}\right\rangle \textit{independent} \stackrel{d}{=} F_{n_2,n_1}. \end{aligned}$$

Thus, $X \sim F_{n_1,n_2} \implies \frac{1}{X} \sim F_{n_2,n_1}$.

(c) $X \sim t_m \implies$ *Kurtosis =* $\nu_1 = \frac{3(m-2)}{m-4},\ m > 4 \implies t_m$ *distribution* ($m > 4$) *is symmetric and leptokurtic (that is, it has sharper peak and longer fatter tails compared to* $N(0,1)$ *distribution). As* $m \to \infty, \nu_1 \to \infty$. *This suggests that for large d.f.* m, t_m *distribution behaves like* $N(0,1)$ *distribution.*

(d) For various values of $m \in \mathbb{N}$ *and* $\alpha \in (0,1)$, *the d.f. of* t_m *is tabulated in various text books.*

(e) For fixed $n_1 \in \mathbb{N}$, $n_2 \in \mathbb{N}$ *and* $\alpha \in (0,1)$ *let* $f_{n_1,n_2,\alpha}$ *be the* $(1-\alpha)$*-th quantile of* $X \sim F_{n_1,n_2}$. *Thus*

$$P(X \le f_{n_1,n_2,\alpha}) = 1 - \alpha \implies P\left(\frac{1}{X} \le \frac{1}{f_{n_1,n_2,\alpha}}\right) = \alpha \implies f_{n_2,n_1,1-\alpha} = \frac{1}{f_{n_1,n_2,\alpha}} \ \left(\textit{as } \frac{1}{X} \sim F_{n_2,n_1}\right).$$

Example 0.113. *Let* $X_1, X_2, \ldots, X_n$ *be a random sample from* $N(\mu, \sigma^2)$ *distribution, where* $\mu \in \mathbb{R}$, $\sigma > 0$ *and* $n \ge 2$. *Let*

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

be the sample mean and sample variance, respectively. Evaluate $E\left(\frac{\bar{X}}{S}\right)$, *for* $n > 2$.

Solution: We have

$$\begin{aligned} & \bar{X} \sim N(\mu, \sigma^2/n) \ \text{ and } Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2 \text{ are independent} \\ \implies E\left(\frac{\bar{X}}{S}\right) &= \frac{\sqrt{n-1}}{\sigma} E(\bar{X} Y^{-1/2}) \\ &= \frac{\sqrt{n-1}}{\sigma} E(\bar{X}) E(Y^{-1/2}) \ \text{ (indepedence)} \\ &= \frac{\sqrt{n-1}}{\sigma}\mu \int_0^\infty \frac{e^{-y/2} y^{\frac{n-2}{2}-1}}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} \mathrm{d}y \\ &= \frac{\sqrt{n-1}}{\sigma}\mu \frac{2^{\frac{n-2}{2}}\Gamma(\frac{n-2}{2})}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} = \frac{\sqrt{(n-1)/2}}{\sigma} \frac{\Gamma(\frac{n-2}{2})}{\Gamma(\frac{n-1}{2})} \frac{\mu}{\sigma}. \end{aligned}$$

Example 0.114. *Let* $Z_1, Z_2, \ldots, Z_n$ *be iid* $N(0,1)$ *r.v.s and let* $a_i, b_i \in \mathbb{R}$, $i = 1, \ldots, n$ *be such that* $\sum_{i=1}^n a_i^2 > 0$, $\sum_{i=1}^n b_i^2 > 0$ *and* $\sum_{i=1}^n a_i b_i = 0$. *Show that*

(a) $Y_1 = \sqrt{\frac{\sum_{i=1}^n b_i^2}{\sum_{i=1}^n a_i^2}} \frac{\sum_{i=1}^n a_i Z_i}{|\sum_{i=1}^n b_i Z_i|} \sim t_1$;

(b) $Y_2 = \frac{\sum_{i=1}^n b_i^2}{\sum_{i=1}^n a_i^2}\left(\frac{\sum_{i=1}^n a_i Z_i}{\sum_{i=1}^n b_i Z_i}\right)^2 \sim F_{1,1}$;

(c) $Y_3 = \sqrt{\frac{\sum_{i=1}^n b_i^2}{\sum_{i=1}^n a_i^2}}\frac{\sum_{i=1}^n a_i Z_i}{\sum_{i=1}^n b_i Z_i} \sim t_1$.

Solution: Linear combination of $Z_1, Z_2, \ldots, Z_n$:

$$\begin{aligned}
&c_1\sum_{i=1}^n a_i Z_i + c_2\sum_{i=1}^n b_i Z_i \quad \text{(univariate normal distribution)}\\
\Longrightarrow\ &\left(\sum_{i=1}^n a_i Z_i, \sum_{i=1}^n b_i Z_i\right) \sim N_2\\
&E\left(\sum_{i=1}^n a_i Z_i\right) = 0,\ \ \text{Var}\left(\sum_{i=1}^n a_i Z_i\right) = \sum_{i=1}^n a_i^2,\\
&E\left(\sum_{i=1}^n b_i Z_i\right) = 0,\ \ \text{Var}\left(\sum_{i=1}^n b_i Z_i\right) = \sum_{i=1}^n b_i^2,\\
&\text{Cov}\left(\sum_{i=1}^n a_i Z_i, \sum_{i=1}^n b_i Z_i\right) = \sum_{i=1}^n a_i b_i = 0,\\
\Longrightarrow\ &\left(\sum_{i=1}^n a_i Z_i, \sum_{i=1}^n b_i Z_i\right) \sim N_2\left(0, 0, \sum_{i=1}^n a_i^2, \sum_{i=1}^n b_i^2, 0\right)\\
\Longrightarrow\ &\sum_{i=1}^n a_i Z_i \sim N\left(0, \sum_{i=1}^n a_i^2\right) \ \text{ and }\ \sum_{i=1}^n b_i Z_i \sim N\left(0, \sum_{i=1}^n b_i^2\right) \ \text{ are independent}\\
\Longrightarrow\ &\frac{\sum_{i=1}^n a_i Z_i}{\sqrt{\sum_{i=1}^n a_i^2}} \sim N(0,1) \ \text{ and }\ \frac{\sum_{i=1}^n b_i Z_i}{\sqrt{\sum_{i=1}^n b_i^2}} \sim N(0,1) \ \text{ are independent.}
\end{aligned}$$

(a)

$$\begin{aligned}
&\frac{\sum_{i=1}^n a_i Z_i}{\sqrt{\sum_{i=1}^n a_i^2}} \sim N(0,1) \ \text{and}\ \frac{\left(\sum_{i=1}^n b_i Z_i\right)^2}{\sum_{i=1}^n b_i^2} \sim \chi_1^2 \ \text{ are independent}\\
\Longrightarrow\ &\frac{\sum_{i=1}^n a_i Z_i/\sqrt{\sum_{i=1}^n a_i^2}}{\sqrt{\frac{\left(\sum_{i=1}^n b_i Z_i\right)^2}{\sum_{i=1}^n b_i^2}}} \sim t_1,\ \text{ that is, }\ Y_1 \sim t_1.
\end{aligned}$$

(b) Since $t_1^2 \stackrel{d}{=} F_{1,1}$, the result follows on using (a).

(c) $F_{Y_3}(y) = P(Y_3 \le y) = P\left(\frac{Z_1}{Z_2} \le y\right)$, $y \in \mathbb{R}$ (Why?). Clearly,

$$\begin{aligned}
F_{Y_3}(y) &= P\left(\frac{Z_1}{|Z_2|} \le y,\ Z_2 > 0\right) + P\left(-\frac{Z_1}{|Z_2|} \le y,\ Z_2 < 0\right)\\
&= P\left(\frac{Z_1}{|Z_2|} \le y,\ Z_2 > 0\right) + P\left(\frac{Z_1}{|Z_2|} \le y,\ Z_2 < 0\right) \quad ((Z_1, Z_2) \stackrel{d}{=} (-Z_1, Z_2))\\
&= P\left(\frac{Z_1}{|Z_2|} \le y\right),\ \forall\, y \in \mathbb{R} \implies Y_3 \stackrel{d}{=} \frac{Z_1}{|Z_2|} \sim t_1,\ (\text{by } (a)) \implies Y_3 \sim t_1.
\end{aligned}$$

0.7. Statistical Inference

We seek information about some numerical characteristic(s) of a collection of elements called population.

For reasons of time or cost it may not be possible to study each individual element of the population (which althouh is the best thing to do)

Goal to draw conclusions (or make inferences) about the unknown characteristics of the population on the basis of information on characteristic(s) of a suitable selected sample:

$\underline{X}$: a random sample (or random vector) describing the characteristics of a population under study.

F: distribution function (d.f) of X.

Parametric Statistical Inference:

Here the r.v.s $\underline{X}$ has a d.f $F \equiv F_\theta$ with a known functional form (except perhaps for the unknown parameter θ, which may be vector).

$\mathscr{H}$: set of all possible values of unknown parameter θ. $\mathscr{H}$ is called parameter space.

Basic Parametric Statistical Inference Problem:

To decide, on the basis of a suitable selected sample, which member or members of the family $\{F_\theta : \theta \in \mathscr{H}\}$ can represent the d.f of X.

Non Parametric Statistical Inference:

Here we know nothing about the d.f F (except perhaps that F is absolutely continous or discrete)

$\underline{Goal}$: To make inference about unknown d.f F.

In this course we only concentrate on parametric inference.

$\underline{Data\ Collection}$:

The statistician can observe n independent observations (say $x_1, \ldots, x_n$) on r.v's X that describes the population under study.

Here each x_i can be regarded as the value assumed by a random variable X_i, $i = 1, 2, \ldots, n$, where $X_1, \ldots, X_n$ are independent r.v's with common d.f F.

So the values of $(x_1, \ldots, x_n)$ are the values assumed by $(X_1, \ldots, X_n)$.

$X_1, \ldots, X_n$: a sample of size n taken from a population with d.f F.

$(x_1, \ldots, x_n)$: realization of the sample $(X_1, \ldots, X_n)$.

Sample Space:

The space of possible values of $(X_1, \ldots, X_n)$ is called the sample space and denoted by $\mathscr{X}$.

Random Sample:

Let X be a r.v with d.f F and let $X_1, \ldots, X_n$ be a collection of independent and identically distributed (i.i.d) r.v's with common d.f F. Then the collection $X_1, \ldots, X_n$ is known as a random sample of size n from d.f F, or the corresponding population.

Statistics:

Let $T : \mathscr{X} \to \mathbb{R}^k$ be a Borel function. Then the r.v $T(X_1, \ldots, X_n)$ is called a (sample) statistic provided it is not a function of any unknown parameters, *i.e.* T only depends on sample $(X_1, \ldots, X_n)$.

Example:

Let $X_1, \ldots, X_n$ be a random sample (r.s) from $N(\mu, \sigma^2)$.

$\mathscr{H} = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\} = \mathbb{R} \times (0, \infty)$ is unknown. Then $T_1(\underline{X}) = \sum(X_i - \bar{X})^2$, $T_2 = \sum X_i$ are statistics but $T_3(\underline{X}) = \sum(X_i - \mu)^2$ is not statistics.

In this statistical inference we study two main topics:

(i) Point Estimation, Interval Estimation

(ii) Testing of Hypothesis.

Point Estimation

$\underline{Estimator}$:

Any function of the random sample which is used to estimate the unknown value of the given parametric function $g(\theta)$ is called an estimator.

If $\underline{X} = (X_1, \ldots, X_n)$ is a random sample from a population with probability distribution F_θ, a function $T(\underline{X})$ used for estimating $g(\theta)$ is known as estimator.

Let $\underline{x} = (x_1, \ldots, x_n)$ be a realization of $\underline{X}$, then $T(\underline{x})$ is called an estimate.

Let $X_1, \ldots, X_n$ be a random sample from a population described by family $\mathcal{F} = \{f_{\underline{\theta}}(\cdot) : \underline{\theta} \in \mathscr{H}\}$ of pdf's/pmf's where for each $\underline{\theta} \in \mathscr{H}$ form of $f_{\underline{\theta}}(\cdot)$ is known but $\underline{\theta} \in \mathscr{H}$ is unknown.

Here knowledge of unknown $\underline{\theta} \in \mathscr{H}$ yields knowledge of unknown $\underline{\theta} \in \mathscr{H}$ yields knowledge of the entire population. Moreover $\underline{\theta}$ itself may represent an important characterstic of the population (such as population mean, variance *etc.*) and there may be direct interest in obtaing a point estimate of $\underline{\theta}$. Sometimes there may be interest in obtaing point estimate of $g(\theta)$, a given function of $\underline{\theta}$.

$\underline{Goal}$: Based on a random sample $X_1, \ldots, X_n$ from the population, find a good point estimate $g(\underline{\theta})$.

Definition 0.115. *A point estimator of $g(\theta)$ is a function $W(\underline{X})$ of the random sample $\underline{X} = (X_1, \ldots, X_n)$.*

$\underline{Note}$:

(i) An estimator $W(\underline{X})$ is a random variable whereas an estimate is an observed value of the estimator based on an observed sample.

Different Method of Finding Estimator:

(i) $\underline{Method\ of\ moment\ estimator\ (MME)}$:

Let $X_1, \ldots, X_n$ be a random sample from a population with distribution $F_{\underline{\theta}}$, $\underline{\theta} = (\theta_1, \ldots, \theta_k) \in \mathscr{H}$.

Consider k noncentral moments

$$\begin{cases} \mu_1' = E(X_1) = g_1(\underline{\theta}) \\ \mu_2' = E(X_1^2) = g_2(\underline{\theta}) \\ \vdots \qquad\qquad \vdots \\ \mu_k' = E(X_1^k) = g_k(\underline{\theta}) \end{cases} . \tag{0.11}$$

Assume the system of equation (0.11) have solution and solving for $\theta_1, \dots, \theta_k$ we get

$$\begin{cases} \theta_1 = h_1(\mu_1', \dots, \mu_k') \\ \vdots \qquad \vdots \\ \theta_k = h_k(\mu_1', \dots, \mu_k') \end{cases}.$$

Define the 1^{st} k noncentral sample moments

$$\alpha_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \;\; \alpha_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \dots, \alpha_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

In method of moments we estimate k^{th} population moment by k^{th} sample moment *i.e.* $\hat{\mu}_j = \alpha_j, \;\; j = 1, 2, \dots, k$.

Thus the method of moment estimators (MME) of $\theta_1, \dots, \theta_k$ are

$$\hat{\theta}_1 = h_1(\alpha_1, \dots, \alpha_k), \dots, \hat{\theta}_k = h_k(\alpha_1, \dots, \alpha_k).$$

Example:

$X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find MME of $\underline{\theta} = (\mu, \sigma^2)$ and $(\mu + \sigma) = \psi(\underline{\theta})$.

$\underline{Solution}$:

$\mu_1' = E(X_1) = \mu, \;\; \mu_2' = E(X_1^2) = \sigma^2 + \mu^2$. So $\mu = \mu_1'$ and $\sigma^2 = \mu_2' - {\mu_1'}^2$.

So $\hat{\mu}_{MME} = \bar{X}, \; \hat{\sigma}^2_{MME} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Thus,

$$\hat{\mu}_{MME} = \bar{X}, \;\;\; \hat{\sigma}^2_{MME} = \frac{1}{n}\sum (X_i - \bar{X})^2.$$

MME of $\psi(\underline{\theta}) = \mu + \sigma$ is $\bar{X} + \sqrt{\frac{1}{n}\sum (X_i - \bar{X})^2}$.

Example:

$X_1, \dots, X_n \overset{iid}{\sim} Bin(m, p), \; \underline{\theta} = (m, p) \in \mathscr{H}, \; \mathscr{H} = \{1, 2, \dots\} \times (0, 1)$.

$$\mu_1' = E(X_1) = mp, \;\; \mu_2' = E(X_1^2) = mp(1-p) + m^2p^2.$$

Now we have

$$mp(1-p) = \mu_2' - {\mu_1'}^2 \implies (1-p) = \frac{\mu_2' - {\mu_1'}^2}{\mu_1'} \implies p = 1 - \frac{\mu_2' - {\mu_1'}^2}{\mu_1'} \;\text{ and }\; m = \frac{\mu_1'}{p}.$$

$$\hat{p}_{MME} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}{\bar{X}} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}{\bar{X}}, \; \hat{m}_{MME} = \frac{\bar{X}}{\hat{p}_{MME}} = \frac{\bar{X}^2}{1 - \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Remark 0.116. *(i) Some times for k-dimensional parameter we may have to consider more than k equations.*

Example:

$X_1, \dots, X_n \overset{iid}{\sim} U[-\theta, \theta], \;\; \theta > 0$. $\mu_1' = E(X_1) = 0, \;\; \mu_2' = E(X_1^2) = \frac{\theta^2}{3} \implies \theta = \sqrt{3\mu_2'}$,

$\hat{\theta}_{MME} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2}$.

Here the 1^{st} moment did not give any solution so we are using 2^{nd} moment.

(ii) MME may not exist.

(ii) MME is not unique.

Method of Maximum Likliehood Estimation:

For a given observed sample point $\underline{x} = (x_1, \ldots, x_n)$, define

$$L_{\underline{x}}(\underline{\theta}) = \prod_{i=1}^{n} f(x_i|\underline{\theta}), \;\; \underline{\theta} \in \mathscr{H}$$

as a function of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$.

$L_{\underline{x}}(\underline{\theta})$: The probability that the observed sample point $\underline{x}$ came from population represented by pdf/pmf $f(\underline{x}|\underline{\theta})$, $\underline{\theta} \in \mathscr{H}$, *i.e.* the likelihood of observing r.s. $(x_1 \ldots x_n)$ from $f(\cdot|\underline{\theta}), \;\theta \in \mathscr{H}$.

Definition 0.117. *(1) For a given sample point $\underline{x}$, the function $L_{\underline{x}}(\underline{\theta})$ as a function of $\theta \in \mathscr{H}$ is called the likelihood function.*

It means sense to find $\hat{\underline{\theta}}$ that maximizes $L_{\underline{x}}(\underline{\theta})$ for a given sample point $\underline{x}$, as the corresponding population (represented by pdf/pmf) is most likely to have yielded the observed sample $\underline{x}$.

Definition 0.118. *For each $\underline{x} \in \mathfrak{X}$, let $\hat{\theta} = \hat{\theta}(\underline{x})$ be such that*

$$L_{\underline{x}}(\hat{\theta}) = \sup_{\underline{\theta} \in \mathscr{H}} L_x(\underline{\theta}).$$

Then a maximum likelihood estimator (MLE) of the parameter $\underline{\theta} = (\theta_1, \ldots, \theta_k)$, based on random sample $\underline{X}$ is $\hat{\theta}(\underline{X})$.

Finding MLE:

$\hat{\theta}$: MLE if $\quad L_{\underline{x}}(\hat{\theta}) = \sup_{\underline{\theta} \in \mathscr{H}} L_{\underline{x}}(\underline{\theta})$.

Define, $\;l_{\underline{x}}(\underline{\theta}) = \log L_{\underline{x}}(\underline{\theta}), \;\underline{\theta} \in \mathscr{H}$. This is called log-likelihood function.

MLE maximizes $L_{\underline{x}}(\underline{\theta})$ or equivalently $l_{\underline{x}}(\underline{\theta})$. If $l_{\underline{x}}(\underline{\theta})$ is differentiable and maximum at the interior of $\mathscr{H}$ then MLE $\hat{\theta}(\underline{x})$ satisfies

$$\frac{\partial}{\partial \theta_i} l_{\underline{x}}(\underline{\theta})\Big|_{\theta = \hat{\theta}(\underline{x})} = 0, \;\; i = 1, 2, \ldots k.$$

Examples:

(1) Let $X \sim Bin(n, p), \;0 \le p \le 1, \;\theta = p \in [0, 1]$. Here n is known.

$$L_x(p) = f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \;\; x = 0, 1, \ldots, n$$

$$\log L_x(p) = l_x(p) = \log \binom{n}{x} + x \log p + (n - x) \log(1 - p)$$

$$\frac{\partial l_x}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = \frac{x - np}{p(1-p)} \begin{cases} < 0, & \text{if } p > \frac{x}{n} \\ > 0, & \text{if } p < \frac{x}{n} \end{cases}.$$

So,

$$l_x(p) \begin{cases} \uparrow, & \text{if } p < \frac{x}{n} \\ \downarrow, & \text{if } p > \frac{x}{n} \end{cases}.$$

So, $\hat{p}_{ML} = \frac{X}{n}$.

(2) Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{P}(\lambda),\ \lambda > 0$ and $\lambda \le \lambda_0$.

$$L_{\underline{x}}(\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^{n}(x_i!)}$$
$$l_{\underline{x}}(\lambda) = \log L_{\underline{x}}(\lambda) = -n\lambda + \sum x_i \log\lambda - \log\Big\{\prod_{i=1}^{n}(x_i!)\Big\}$$
$$\frac{\partial l_x(\lambda)}{\partial\lambda} = -n + \frac{\sum x_i}{\lambda} = \frac{\sum x_i - n\lambda}{\lambda} \begin{cases} > 0, & \text{if } \lambda < \bar{x} \\ < 0, & \text{if } \lambda > \bar{x} \end{cases}.$$

So,

$$l_x(\lambda) \begin{cases} \uparrow, & \text{if } \lambda < \bar{x} \\ \downarrow, & \text{if } \lambda > \bar{x} \end{cases}.$$

Given $\lambda \le \lambda_0$ *i.e.* $\mathscr{H} = (0, \lambda_0)$.

$$\hat{\lambda}_{ML} = \begin{cases} \bar{x}, & \text{if } \bar{x} \le \lambda_0 \\ \lambda_0, & \text{if } \bar{x} > \lambda_0 \end{cases}.$$

This is restricted MLE.

Remark 0.119. *(1) MLE is not unique.*

Example:

$X_1, \ldots, X_n \overset{iid}{\sim} U[\theta - a, \theta + a],\ a \in \mathbb{R},\ a > 0$ where a is a known constant.

The likelihood function

$$L_{\underline{x}}(\theta) = \begin{cases} \left(\dfrac{1}{2a}\right)^n, & \text{if } \theta - a \le x_{(1)} < x_{(2)} < \cdots < x_{(n)} \le \theta + a \\ 0, & \text{otherwise.} \end{cases}$$

So $L_{\underline{x}}(\theta)$ is maximum when $\theta - a \le x_{(1)}$ and $x_{(n)} \le \theta + a$ *i.e.* $x_{(n)} - a \le \theta \le x_{(1)} + a$. So any value of θ between $x_{(n)} - a$ to $x_{(1)} + a$ is MLE of θ. We may choose the midpoint *i.e.* $\dfrac{x_{(1)} - x_{(n)}}{2}$ as the MLE.

Example:

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\mu \in \mathbb{R},\ \sigma > 0$.

$$L_{\underline{x}}(\theta) = \prod_{i=1}^{n}\left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}\right] = \frac{1}{\sigma^n(2\pi)^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},\ \mu \in \mathbb{R},\ \sigma > 0,\ x_i \in \mathbb{R},\ i = 1(1)n.$$

$$l_{\underline{x}}(\theta) = \log L_{\underline{x}}(\underline{\theta}) = -n\log\sigma - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$
$$\frac{\partial l_{\underline{x}}(\theta)}{\partial\mu} = \frac{1}{\sigma^2}\sum(x_i - \mu),\quad \frac{\partial l_{\underline{x}}(\theta)}{\partial\mu} = 0 \implies \mu = \bar{x}$$
$$\frac{\partial l}{\partial\sigma^2} = 0 \implies \sigma^2 = \frac{1}{n}\sum(x_i-\mu)^2.$$

So $\hat{\mu}_{ML} = \bar{X}$ and $\sigma^2_{ML} = \dfrac{1}{n}\sum(X_1 - \bar{X})^2$. Consider the MLE of μ and σ^2 when $\mu > 0$. In this case $\mu \in (0, \infty)$. We have

$$\frac{\partial l_x}{\partial \mu} = \frac{n(\bar{x} - \mu)}{\sigma} \begin{cases} > 0, & \text{if } \mu < \bar{x} \\ < 0, & \text{if } \mu > \bar{x} \end{cases}.$$

So,

$$\hat{\mu}_{RML} = \begin{cases} \bar{X}, & \text{if } \bar{X} > 0 \\ 0, & \text{if } \bar{X} \leq 0 \end{cases} = \max\{\bar{X}, 0\}.$$

$$\hat{\sigma}^2_{RML} = \frac{1}{n}\sum(X_1 - \hat{\mu}_{RML})^2 = \begin{cases} \dfrac{1}{n}\sum(X_1 - \bar{X})^2, & \text{if } \bar{X} > 0 \\ \dfrac{1}{n}\sum X_1^2, & \text{if } \bar{X} \leq 0 \end{cases}.$$

$\underline{Exercise}$:

(i) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Exp(\mu, \sigma)$. Find MLE of μ and σ.

(ii) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Gamma(r, \lambda)$. Find MLE of r and λ.

Example:

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$. Then we have

$$L_{\underline{x}}(\theta) = \begin{cases} \dfrac{1}{\theta^n}, & 0 < x_i < \theta, \ \ i = 1(1)n \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \dfrac{1}{\theta^n}, & 0 < x_{(1)} < x_{(2)} < \cdots < x_{(n)} < \theta \\ 0, & \text{otherwise} \end{cases}$$

$L_{\underline{x}}(\theta)$ is the decreasing function of θ. So $L_{\underline{x}}(\theta)$ attains its maximum when θ is minimum $\implies \hat{\theta}_{ML} = X_{(n)}$.

$\underline{Exercise}$:

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(\theta, \theta + 1)$. Find MLE of θ.

Remark 0.120. *(3) Finding MLE requires maximization of likelihood function which is sometimes difficult and may require numerical optimization techniques.*

Remark 0.121. *(4) MLE is sometimes sensitive to data, a slightly different data may produce a vastly different MLE.*

Invariance Property of MLE:

Let $\hat{\theta} = \theta(\underline{X})$ be a MLE of $\hat{\theta}$ then the of MLE of $\psi(\theta)$ is $\psi(\hat{\theta})$.

$\underline{Exercise}$:

(1) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Bin(n, \theta)$. Then find a MLE of $\theta(1 - \theta)$.

(2) Let $X_1, X_2, \ldots, X_m \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$. The two sample are independent. Then find the distribution of

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(m-1)S_1^2}{\sigma_1^2} + \frac{(n-1)S_2^2}{\sigma_2^2}}} \sqrt{\frac{m + n - 2}{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}.$$

(3) Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} U(\theta_1, \theta_2)$. Find MLE of θ_1 and θ_2.

Efficiency of Estimators:

Let $g(\theta)$ be a parametric function and $\delta(\underline{X})$ be an estimator.

$$\text{Mean absolue Error} : E|\delta(\underline{X}) - g(\theta)|,$$
$$\text{Mean squared Error} : E(\delta(\underline{X}) - g(\theta))^2$$

Definition 0.122. *We say that estimator δ_1 is better (more efficient) than δ_2 if*

$$MSE(\delta_1) \leq MSE(\delta_2) \ \ \forall \theta \in \mathscr{H}.$$

If $E(\delta(\underline{X})) = g(\theta)$ then $MSE(\delta) = E(\delta(\underline{X}) - g(\theta))^2 = \mathrm{Var}(\delta(\underline{X}))$.

Unbaised Estimator:

A estimator $\delta(\underline{X})$ is said to be an unbaised estimator of g(θ) if

$$E((\delta(\underline{X})) = g(\theta) \ \ \forall \theta \in \mathscr{H}.$$

Example: $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$, $\delta_1(\underline{X}) = \bar{X}$.

$E(\delta_1(\underline{X})) = \mu \implies \delta_1(\underline{X})$ is unbaised for μ.