Module 4
Random Vectors

4.1. 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 defining a function

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

Example 4.1.1. A fair coin is tossed three times (independently). Then

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

and

$$P(\{\omega\}) = \frac{1}{8}, \quad \forall\, \omega \in \Omega.$$

Suppose that we are simultaneously interested in:

- number of heads in three tosses

and
- 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

$Pr\left((\underset{\sim}{X}, Y) \in \{(0,0), (1,0), (1,1), (2,1), (2,2), (3,2)\}\right) = 1.$

$\boxed{\text{Definition 4.1.1.}}$ Let $(\Omega, \mathcal{F}, P)$ be a given probability space. A function $\underset{\sim}{X} = (X_1, \ldots, X_p): \Omega \to \mathbb{R}^p$ (defined on the sample space Ω) is called a random vector (p-dimensional random vector). A one-dimensional random vector (r.v.) is simply called a random variable (r.v.).

For any function $\underset{\sim}{Y} = (Y_1, \ldots, Y_p): \Omega \to \mathbb{R}^p$ and $A \subseteq \mathbb{R}^p$, define

$$\underset{\sim}{Y}^{-1}(A) = \{\omega \in \Omega : \underset{\sim}{Y}(\omega) \in A\}.$$

For a probability space $(\Omega, \mathcal{F}, P)$ and a p-dimensional random vector $\underset{\sim}{X} = (X_1, \ldots, X_p): \Omega \to \mathbb{R}^p$, define

$$P_{\underset{\sim}{X}}(B) = P(\underset{\sim}{X}^{-1}(B)), \quad B \in \mathcal{B}_p,$$

where, for all practical purposes, we take $\mathcal{B}_p$ to be power set of $\mathbb{R}^p$.

We will simply write

$$P_{\underset{\sim}{X}}(B) = P(\{\omega \in \Omega : \underset{\sim}{X}(\omega) \in B\})$$

$$= Pr(X \in B), \quad B \in \mathcal{B}_p$$

The following scenario has emerged:

$$(\Omega, \mathcal{F}, P) \xrightarrow{\underset{\sim}{X} \text{ (r.v.)}} (\mathbb{R}^p, \mathcal{B}_p, P_{\underset{\sim}{X}})$$

$\boxed{\text{Theorem 4.1.1.}}$ $(\mathbb{R}^p, \mathcal{B}_p, P_{\underset{\sim}{X}})$, defined above, is a probability space, i.e. $P_{\underset{\sim}{X}}(\cdot)$ is a probability function defined on $\mathcal{B}_p$.

$\boxed{\text{Proof.}}$ Similar to the proof of Theorem 2.1.1.

Definition 4.1.2. The probability function $P_{\underline{X}}(\cdot)$ defined above is called the probability function/measure induced by r.v. $\underline{X}$ and $(\mathbb{R}^p, \mathcal{B}_p, \underline{X})$ is called the probability space induced by r.v. $\underline{X}$

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

Example 4.1.2. Consider the probability space defined in Example 4.1.1, where

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

$$P(\{\omega\}) = \frac{1}{8}, \quad \forall \omega \in \Omega$$

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 $(\mathbb{R}^2, \mathcal{B}_2, P_{\underline{X}})$, 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 $B \in \mathcal{B}_2$

$$P_X(B) = \sum_{(i,j) : B \cap S} P_X(\{(i, j)\}),$$

where $S = \{(0, 0), (1, 0), (1, 1), (2, 1), (2, 2), (3, 2)\}$

Definition 4.1.3. (a) The joint distribution function of a p-dimensional random vector $\underline{X} = (X_1, \dots, X_p)$ is defined by

$$F_{\underline{X}}(x_1, \dots, x_p) = \Pr(X_1 \le x_1, \dots, X_p \le x_p), \quad \underline{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$$

(b) The joint d.f. of any subset of r.v.s $X_1, \dots, X_p$ is called a marginal d.f. of $F_{\underline{X}}(\cdot)$ (or $\underline{X} = (X_1, \dots, X_p)$).

Example 4.1.3. $F_{X_1, X_2}(x, y)$, $(x, y) \in \mathbb{R}^2$, $F_{X_2}(x)$, $x \in \mathbb{R}$ 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 notations for writing down all the vertices of a p-dimensional rectangle in a compact form.

For $-\infty \le a_i < b_i < \infty$, $i = 1, 2$, $\underline{a} = (a_1, a_2)$ & $\underline{b} = (b_1, b_2)$ the vertices of the 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 $-\infty \le a_i < b_i < \infty$, $i = 1, 2, 3$, $\underline{a} = (a_1, a_2, a_3)$ & $\underline{b} = (b_1, b_2, b_3)$, the vertices of the 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

$$\{(b_1, b_2, b_3), (a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3), (a_1, a_2, b_3), (a_1, b_2, a_3), (b_1, a_2, a_3), (a_1, a_2, a_3)\}$$

$$= \{(b_1, b_2, b_3)\} \cup \{(a_1, b_2, b_3), (b_1, a_2, b_3), (b_1, b_2, a_3)\} \cup \{(a_1, a_2, b_3), (a_1, b_2, a_3), (b_1, a_2, a_3)\} \cup \{(a_1, a_2, a_3)\}$$

$$= \Delta_{0,3} \cup \Delta_{1,3} \cup \Delta_{2,3} \cup \Delta_{3,3}$$

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

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

where $(\underline{a}, \underline{b}] = (a_1, b_1] \times (a_2, b_2] \times \ldots \times (a_p, b_p]$. $\longrightarrow$ has $\binom{p}{k}$ elements

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

$\boxed{\text{Move to Page 6/4}}$

Theorem 4.1.2. For constants $-\infty \le a_i < b_i < \infty$, $i=1,\dots,p$

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

Proof. (Special Cases)

Case I. $p=1$

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

Case II $p=2$

Here

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

Thus

R.H.S. $= F_{\underline{X}}(b_1, b_2) - F_{\underline{X}}(a_1, b_2) - F_{\underline{X}}(b_1, a_2) + F_{\underline{X}}(a_1, a_2)$

$= \Pr(X_1 \le b_1, X_2 \le b_2) - \Pr(X_1 \le a_1, X_2 \le b_2) - \Pr(X_1 \le b_1, X_2 \le a_2) + \Pr(X_1 \le a_1, X_2 \le a_2)$

$= \Pr(a_1 < X_1 \le b_1, X_2 \le b_2) - \Pr(a_1 < X_1 \le b_1, X_2 \le a_2)$

$= \Pr(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2) =$ R.H.S.

Case III $p=3$

$\Pr(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, a_3 < X_3 \le b_3)$

$= \Pr(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, X_3 \le b_3) - \Pr(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, X_3 \le a_3)$

$= \Pr(a_1 < X_1 \le b_1, X_2 \le b_2, X_3 \le b_3) - \Pr(a_1 < X_1 \le b_1, X_2 \le a_2, X_3 \le b_3)$
$- \{\Pr(a_1 < X_1 \le b_1, X_2 \le b_2, X_3 \le a_3) - \Pr(a_1 < X_1 \le b_1, X_2 \le a_2, X_3 \le a_3)\}$

$= \Pr(X_1 \le b_1, X_2 \le b_2, X_3 \le b_3) - \Pr(X_1 \le a_1, X_2 \le b_2, X_3 \le b_3)$
$- \Pr(X_1 \le b_1, X_2 \le a_2, X_3 \le b_3) + \Pr(X_1 \le a_1, X_2 \le a_2, X_3 \le b_3)$
$- P(X_1 \le b_1, X_2 \le b_2, X_3 \le a_3) + \Pr(X_1 \le a_1, X_2 \le b_2, X_3 \le a_3)$
$+ \Pr(X_1 \le b_1, X_2 \le a_2, X_3 \le a_3) - \Pr(X_1 \le a_1, X_2 \le a_2, X_3 \le a_3)$

$= F_{\underline{X}}(b_1, b_2, b_3) - F_{\underline{X}}(a_1, b_2, b_3) - F_{\underline{X}}(b_1, a_2, b_3) + F_{\underline{X}}(a_1, a_2, b_3)$
$\quad - F_{\underline{X}}(b_1, b_2, a_3) + F_{\underline{X}}(a_1, b_2, a_3) + F_{\underline{X}}(b_1, a_2, a_3) - F_{\underline{X}}(a_1, a_2, a_3)$

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

The following theorem provides a technique to find marginal distributions.

$\boxed{\text{Theorem 4.1.3.}}$ Let $F(x_1, \ldots, x_p)$, $\underline{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$, be the d.f. of p-dimensional r.v. $\underline{X} = (X_1, \ldots, X_p)$. Then the marginal d.f. of $\underline{Y} = (X_1, \ldots, X_{p-1})$ is

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

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

$$\begin{aligned} G(x_1, \ldots, x_{p-1}) &= \Pr(X_1 \le x_1, \ldots, X_{p-1} \le x_{p-1}) \\ &= \Pr(X_1 \le x_1, \ldots, X_{p-1} \le x_{p-1}, X_p < \infty) \\ &= \Pr\Big(\bigcup_{t=1}^{\infty} \underbrace{\{X_1 \le x_1, \ldots, X_{p-1} \le x_{p-1}, X_p \le t\}}_{\uparrow \text{ in } t}\Big) \\ &= \lim_{t \to \infty} \Pr(\{X_1 \le x_1, \ldots, X_{p-1} \le x_{p-1}, X_p \le t\}) \\ &= \lim_{t \to \infty} F(x_1, \ldots, x_{p-1}, t). \end{aligned}$$

$\boxed{\text{Theorem 4.1.4.}}$ Let $\underline{X} = (X_1, \ldots, X_p)$ be a p-dimensional r.v. with d.f. $F(\cdot)$. Then

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

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

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

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

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

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

$\boxed{\text{Proof.}}$ For simplicity we provide the proof for $p=2$.

(a) $$\lim_{x_1 \to \infty, x_2 \to \infty} F(x_1, x_2) = \lim_{x_1 \to \infty, x_2 \to \infty} Pr(\{X_1 \leq x_1, X_2 \leq x_2\})$$

$$= \lim_{n \to \infty} Pr(\underbrace{\{X_1 \leq n, X_2 \leq n\}}_{\uparrow}) \quad \text{(since limit exists)}$$

$$= Pr\left(\bigcup_{n \geq 1}^{\infty} \{X_1 \leq n, X_2 \leq n\}\right)$$

$$= Pr(X_1 < \infty, X_2 < \infty)$$

$$= 1.$$

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

$$\lim_{x_1 \to -\infty} F(x_1, x_2) = \lim_{n \to \infty} Pr(\underbrace{\{X_1 \leq -n, X_2 \leq x_2\}}_{\downarrow})$$

$$= Pr\left(\bigcap_{n \geq 1}^{\infty} \{X_1 \leq -n, X_2 \leq x_2\}\right)$$

$$= Pr(\phi) = 0.$$

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

$$\lim_{n \to \infty} F(x_1 + h_n, x_2) = \lim_{n \to \infty} Pr(\{X_1 \leq x_1 + h_n, X_2 \leq x_2\}) \quad \text{(limit exists)}$$

$$= \lim_{n \to \infty} Pr(\underbrace{\{X_1 \leq x_1 + \tfrac{1}{n}, X_2 \leq x_2\}}_{\downarrow})$$

$$= Pr\left(\bigcap_{n \geq 1}^{\infty} \{X_1 \leq x_1 + \tfrac{1}{n}, X_2 \leq x_2\}\right)$$

$$= Pr(\{X_1 \leq x_1, X_2 \leq x_2\})$$

$$= F(x_1, x_2),$$

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

$$\sum_{k=0}^{2} (-1)^k \sum_{\underline{z} \in \Delta_{k,p}((\underline{a}, \underline{b}])} F(\underline{z}) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

$$= P(a_1 < X_1 \leq b_1, \; a_2 < X_2 \leq b_2) \geq 0.$$

$\boxed{\text{Remark 4.1.1.}}$ (a) For $p=1$, (d) of the above theorem reduces to

$$F(b) - F(a) \geq 0, \quad \forall -\infty < a < b < \infty,$$

i.e., $F(\cdot)$ is monotone on $\mathbb{R}$.

i.e., $F(\cdot)$ is non-decreasing in each argument.

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

4.2. Independent Random Variables

For an arbitrary (countable or uncountable) set Λ, let $\{X_\lambda : \lambda \in \Lambda\}$ be a family of r.v.s.

$\boxed{\text{Definition 4.2.1.}}$ The random variables X_λ, $\lambda \in \Lambda$, are said to be mutually independent if for any finite subcollection $\{X_{\lambda_1}, \dots, X_{\lambda_p}\}$ in $\{X_\lambda : \lambda \in \Lambda\}$

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

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

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

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

Remark 4.2.1. (a) Random variables $\{X_\lambda : \lambda \in \Lambda\}$ are independent iff those in any finite subset of $\{X_\lambda : \lambda \in \Lambda\}$ are independent.

(b) Let $\Lambda_1 \subseteq \Lambda_2$ ...... Then r.v.s $\{X_\lambda : \lambda \in \Lambda_2\}$ are independent $\Rightarrow$ r.v.s $\{X_\lambda : \lambda \in \Lambda_1\}$ are independent. In particular if r.v.s in a collection are independent then they are pairwise independent. The converse may not be true (see Assignment-IV problems)

Theorem 4.2.1. For a positive integer $p (\geq 2)$, the r.v.s $X_1 \ldots X_p$ are independent iff

$$F(x_1 \ldots, x_p) = \prod_{i=1}^{p} F(x_i), \quad \forall\, \underline{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p, \quad \ldots \quad (4.2.1)$$

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

Proof. Obviously if $X_1 \ldots, X_p$ are independent then (4.2.1) holds. Consider a subset of $\{X_1 \ldots, X_p\}$. Conversely suppose that (4.2.1) holds. For simplicity let this subset be $\{X_1 \ldots X_q\}$, for some $2 \leq q \leq p$. Then, for $\underline{x} = (x_1 \ldots, x_q) \in \mathbb{R}^q$ the joint (marginal) df of $(X_1 \ldots X_q)$ is

$$\begin{aligned} G(x_1 \ldots, x_q) &= \lim_{\substack{x_i \to \infty \\ i = q+1 \ldots p}} F(x_1 \ldots x_q, x_{q+1}, \ldots, x_p) \\ &= \lim_{\substack{x_i \to \infty \\ i = q+1 \ldots p}} \prod_{j=1}^{p} F_{X_j}(x_j) \\ &= \prod_{j=1}^{q} F_{X_j}(x_j), \quad \underline{x} = (x_1 \ldots, x_q) \in \mathbb{R}^q \end{aligned}$$

where $F_{X_j}(\cdot)$ is the marginal d.f. of X_j, $j = 1 \ldots q$.

4.3. Discrete Random Vectors

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

Definition 4.3.1. The ~~r.v.~~ random vector $\underline{X} = (X_1 \ldots, X_p)$ is said to be a discrete random vector if there exists a countable set S (finite or infinite) such that

$$Pr(\underline{X}=\underline{x}) > 0, \quad \forall \underline{x} \in S$$

and $$Pr(\underline{X} \in S) = 1.$$

The set S is called the support of r.v. $\underline{X}$ (or of F).

(b) The (joint) probability mass function of a p-dimensional discrete r.v. $\underline{X}$ having support S is defined by

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

$\boxed{\text{Remark 4.3.1.}}$ (a) Let $\underline{X} = (X_1, \dots, X_p)$ be a p-dimensional discrete r.v. with p.m.f. $f(\cdot)$, d.f. F and support S. Then, for any $A \subseteq \mathbb{R}$

$$Pr(\underline{X} \in A) = Pr(\underline{X} \in A \cap S) \qquad (Pr(\underline{X} \in S) = 1)$$

$$= \sum_{\underline{x} \in A \cap S} f(\underline{x}) \qquad (A \cap S \subseteq S \text{ and thus } A \cap S \text{ is a countable set})$$

Moreover

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

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

(i) $f(\underline{x}) > 0, \ \forall \underline{x} \in S$ and $f(\underline{x}) = 0, \ \forall \underline{x} \in S^c$

and (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) $g(\underline{x}) > 0, \ \forall \underline{x} \in T$ and $g(\underline{x}) = 0, \ \forall \underline{x} \in T^c$

and (ii) $\sum_{\underline{x} \in T} g(\underline{x}) = 1.$

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

(c) Marginal distributions of a discrete r.v. are discrete.

Theorem 4.3.1. Let $\underline{X} = (X_1, \dots, X_p)$ be a p-dimensional discrete r.v. with p.m.f. $f(\cdot)$ and support S. Then the marginal distribution of any subset of $\{X_1, \dots, X_p\}$ (say that of $\underline{Y} = (X_1, \dots, X_q)$, $1 \le q < p$) is again discrete with p.m.f.

$$g(x_1, \dots, x_q) = \begin{cases} \displaystyle\sum_{\substack{x_{q+1} \dots x_p \\ \underline{x} = (x_1 \dots x_q, x_{q+1} \dots x_p) \in S}} \cdots \sum f(\underline{x}) & , \text{ if } \underline{x} \in T \\ 0 & , \text{ otherwise} \end{cases}$$

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

Proof. Follows using theorem of total probability.

Conditional distributions of discrete r.v.s

Let $\underline{Y} = (Y_1, \dots, Y_p)$, $\underline{Z} = (Z_1, \dots, Z_q)$ and $\underline{X} = (\underline{Y}, \underline{Z}) = (Y_1, \dots, Y_p, Z_1, \dots, Z_q)$ be random vectors with p.m.f.s f_1, f_2 and f, respectively. Suppose that supports of $\underline{X}$, $\underline{Y}$ and $\underline{Z}$ are S, S_1 and S_2, respectively. For fixed $\underline{z} \in S_2$, define

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

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

$$f(\underline{y} | \underline{z}) = \Pr(\underline{Y} = \underline{y} \mid \underline{Z} = \underline{z})$$

$$= \frac{\Pr(\underline{X} = (\underline{y}, \underline{z}))}{\Pr(\underline{Z} = \underline{z})}$$

$$= \begin{cases} \dfrac{f(\underline{y}, \underline{z})}{f_2(\underline{z})}, & \text{if } \underline{y} \in T_{\underline{z}} \\ 0, & \text{otherwise} \end{cases}.$$

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 $\underline{z} \in S_2$,

$$\Pr(Y_1 \le y_1, \dots, Y_q \le y_q \mid \underline{Z} = \underline{z}) = \frac{\Pr(Y_1 \le y_1, \dots, Y_q \le y_q, \underline{Z} = \underline{z})}{\Pr(\underline{Z} = \underline{z})} = \sum_{\substack{\underline{s} \in T_{\underline{z}} \\ \underline{s} \le \underline{y}}} \frac{f(\underline{s}, \underline{z})}{f_2(\underline{z})}$$

$$= \sum_{\substack{\underline{s} \in T_{\underline{z}} \\ \underline{s} \le \underline{y}}} f(\underline{s} | \underline{z}).$$

Theorem 4.3.2. Let $\underline{X} = (X_1, \dots, X_p)$ be a p-dimensional r.v. with support S and p.m.f. $f(\cdot)$. Let $f_i(\cdot)$ denote the marginal p.m.f. of X_i, $i = 1, \dots, p$. Then $X_1, \dots, X_p$ are independent iff

$$f(x_1, \dots, x_p) = \prod_{i=1}^{p} f_i(x_i), \quad \forall\ \underline{x} = (x_1, \dots, x_p) \in S. \quad \dots (4.3\text{-}1)$$

Proof. (For $p=2$). Suppose that

$$f(x_1, x_2) = f_1(x_1)\, f_2(x_2), \quad \forall\ \underline{x} = (x_1, x_2) \in S$$

Then the d.f. of $\underline{X} = (X_1, X_2)$ is

$$F(x_1, x_2) = \sum_{\substack{(y_1, y_2) \in S \\ y_1 \le x_1,\ y_2 \le x_2}} \sum f(y_1, y_2)$$

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

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

$$S = \{(y_1, y_2) \in \mathbb{R}^2 : f(y_1, y_2) > 0\}$$

$$= \{(y_1, y_2) \in \mathbb{R}^2 : f_1(y_1)\, f_2(y_2) > 0\}$$

$$= \{(y_1, y_2) \in \mathbb{R}^2 : f_1(y_1) > 0, \text{ and } f_2(y_2) > 0\}$$

$$= \{y_1 \in \mathbb{R} : f_1(y) > 0\} \times \{y_2 \in \mathbb{R} : f_2(y) > 0\}$$

$$= S_1 \times S_2.$$

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

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

$$= \Big(\sum_{\substack{y_1 \in S_1 \\ y_1 \le x_1}} f_1(y_1) \Big) \Big(\sum_{\substack{y_2 \in S_2 \\ y_2 \le x_2}} f_2(y_2) \Big)$$

$$= F_1(x_1)\, F_2(x_2),$$

where F_1 and F_2 are marginal dfs of X_1 and X_2, respectively.

$\Rightarrow$ X_1 and X_2 are independent.

Conversely suppose that X_1 and X_2 are independent. Then

$$F(y_1, y_2) = F_1(y_1)\,F_2(y_2) \quad \forall\ (y_1, y_2) \in \mathbb{R}^2$$

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

$$f(x_1, x_2) = \Pr(X_1 = x_1,\ X_2 = x_2)$$

$$= \Pr\Big(\bigcap_{n=1}^{\infty} \underbrace{\{x_1 - \tfrac{1}{n} < X_1 \le x_1,\ x_2 - \tfrac{1}{n} < X_2 \le x_2\}}_{\downarrow}\Big)$$

$$= \lim_{n\to\infty} \Pr\big(x_1 - \tfrac{1}{n} < X_1 \le x_1,\ x_2 - \tfrac{1}{n} < X_2 \le x_2\big)$$

$$= \lim_{n\to\infty} \Big[F(x_1, x_2) - F(x_1 - \tfrac{1}{n}, x_2) - F(x_1, x_2 - \tfrac{1}{n}) + F(x_1 - \tfrac{1}{n}, x_2 - \tfrac{1}{n})\Big]$$

$$= \lim_{n\to\infty} \Big[F_1(x_1)F_2(x_2) - F_1(x_1 - \tfrac{1}{n})F_2(x_2) - F_1(x_1)F_2(x_2 - \tfrac{1}{n}) + F_1(x_1 - \tfrac{1}{n})F_2(x_2 - \tfrac{1}{n})\Big]$$

$$= F_1(x_1)F_2(x_2) - F_1(x_1-)F_2(x_2) - F_1(x_1)F_2(x_2-) + F_1(x_1-)F_2(x_2-)$$

$$= (F_1(x_1) - F_1(x_1-))F_2(x_2) - (F_1(x_1) - F_1(x_1-))F_2(x_2-)$$

$$= (F_1(x_1) - F_1(x_1-))(F_2(x_2) - F_2(x_2-))$$

$$= f_1(x_1)\,f_2(x_2).$$

Remark 4.3.2. (a) If $\underline{X} = (X_1, \dots, X_p)$ is a discrete r.v. with support S and X_i has support S_i, $i = 1, \dots, p$, then $X_1, \dots, X_p$ are independent $\Rightarrow$ $S = S_1 \times S_2 \times \dots \times S_p$.

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

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

for some functions $g_1, \dots, g_p$ and $S = A_1 \times A_2 \times \dots \times A_p$ and countable set $A_i = \{x \in \mathbb{R} : g_i(x) > 0\}$, $i = 1, \dots, p$. In that case the marginal

$\{x \in \mathbb{R} : g_i(x) > 0\}$

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 $\sum_{x \in A_i} c_i\, g_i(x) = 1$, $i = 1, \dots, p$.

(c) If $\underline{X} = (Y, Z)$ is a two-dimensional discrete r.v. then X and Y are independent iff

$$f(y|z) = f_1(y), \quad \forall\, y \in \mathbb{R} \text{ and } z \in \mathbb{R} \text{ such that } f_2(z) > 0.$$

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 4.2.1 to define independence of a collection of random vectors. Then analogous of Theorem 4.2.1, Remark 4.3.1, Theorem 4.3.1, Theorem 4.3.2 & (c) above hold for random vectors. ~~In fact one can also define conditional distributions in the~~

$\boxed{\text{Example 4.3.1.}}$ Let $\underline{X} = (X_1, X_2, X_3)$ have the joint p.m.f.

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

where c is a real constant.

(a) Find the value of constant c;

(b) Find 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 $\Pr(X_1 = X_2 = X_3)$.

$\boxed{\text{Solution}}$ (a) Here the support of r.v. $\underline{X}$ is

$$S_{\underline{X}} = \{1, 2\} \times \{1, 2, 3\} \times \{1, 3\}$$

$$\sum_{\underline{x} \in S_{\underline{X}}} f(\underline{x}) = 1 \Rightarrow c\,[1+3+2+6+3+9+2+6+4+12+6+18] = 1$$

$$\Rightarrow c = \frac{1}{72}.$$

Clearly $f_{\underline{X}}(\underline{x}) \geq 0 \quad \forall\, \underline{x} \in \mathbb{R}^3$

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

$$f_{X_1}(x_1) = \sum_{(x_2,x_3)\in\{1,2,3\}\times\{1,3\}} \frac{x_1x_2x_3}{72} = \frac{x_1}{72}\Big(\sum_{x_2=1}^{3} x_2\Big)\Big(\sum_{x_3=1,3} x_3\Big) = \frac{x_1}{3}$$

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

Similarly

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

(c) Clearly

$$f(x_1,x_2,x_3) = g_1(x_1)\,g_2(x_2)\,g_3(x_3), \quad (x_1,x_2,x_3)\in\mathbb{R}^3$$

and $S_X = A_1\times A_2\times A_3$,

where $A_1 = \{1,2\}$, $A_2 = \{1,2,3\}$, $A_3 = \{1,3\}$

$$g_1(x_1) = \begin{cases} c_1x_1, & x_1 \in A_1 \\ 0, & \text{otherwise} \end{cases} ; \quad g_2(x_2) = \begin{cases} c_2x_2, & x_2 \in A_2 \\ 0, & \text{otherwise} \end{cases}$$

and

$$g_3(x_3) = \begin{cases} c_3x_3, & x_3 \in A_3 \\ 0, & \text{otherwise} \end{cases}$$

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

$$f(x_1,x_2,x_3) = f_{X_1}(x_1)\,f_{X_2}(x_2)\,f_{X_3}(x_3), \quad \forall\ \underline{x} = (x_1,x_2,x_3)\in\mathbb{R}^3$$

(d) Using (c) it follows that X_1 and X_3 are independent.

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

$$\Pr(X_1 = x_1 \mid (X_2,X_3) = (2,1)) = \frac{\Pr(X_1 = x_1, X_2 = 2, X_3 = 1)}{\Pr(X_2 = 2, X_3 = 1)}$$

$$= \frac{2x_1/72}{\frac{2}{72}(1+2)} = \frac{x_1}{3}$$

Thus

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

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

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

$$\Rightarrow f_{X_1|(X_2,X_3)}(x_1|(2,1)) = \begin{cases} \frac{x_1}{3}, & x_1 \in \{1, 2\} \\ 0, & \text{o.w.} \end{cases}$$

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

$$\Pr(X_1 = X_2 = X_3) = \sum_{\substack{\underline{x} \in S_{\underline{X}} \\ x_1 = x_2 = x_3}} \frac{x_1 x_2 x_3}{72}$$

$$= \Pr(X_1 = X_2 = X_3 = 1) = \frac{1}{72}.$$

4.4. Continuous Random Vectors

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

$\boxed{\text{Definition 4.4.1}}$ The r.v. $\underline{X}$ is called a continuous r.v. if there exists a non-negative function $f: \mathbb{R}^p \to \mathbb{R}$ such that for any reasonable set A in $\mathbb{R}^p$

$$\Pr(\underline{X} \in A) = \int\int \cdots \int_A \xcancel{f(\underline{t})\, d\underline{t}}\ f(\underline{t})\, d\underline{t},$$

where ~~$d\underline{t} = dt_1 \ldots dt_p$~~ $d\underline{t} = dt_p \cdots dt_1$, $\underline{t} = (t_1, \ldots, t_p)$. The function f is called the probability density function of $\underline{X}$ and the set

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

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

$\boxed{\text{Remark 4.4.1.}}$ (a) In particular if, for fixed $\underline{x} = (x_1, \ldots, x_p) \in \mathbb{R}^p$, $A = (-\infty, x_1] \times \cdots \times (-\infty, x_p]$, then

$$F(x_1, \ldots, x_p) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(t_1, \ldots, t_p)\, dt_p \cdots dt_1.$$

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

$$* = \lim_{\substack{h_i \to 0 \\ i=1..p}} \frac{1}{h_1 \cdots h_p} Pr(x_i < X_i \le x_i + h_i \quad i=1,\dots,p)$$

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

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

whenever the derivative is defined.

(c) If $f(\cdot)$ is continuous at $\underline{x} \in \mathbb{R}^p$ then $f(\underline{x}) = \lim_{\substack{h_i \to 0 \\ i=1..p}} \frac{1}{h_1 \cdots h_p} \int_{x_1}^{x_1+h_1} \cdots \int_{x_p}^{x_p+h_p} f(\underline{t})\, d\underline{t}$. $\to *$

For small $dx_1, \dots, dx_p$, if f is continuous at $\underline{x}$, then

$$Pr(x_i < X_i \le x_i + dx_i, \ i=1,\dots,p) = \int_{x_1}^{x_1+dx_1} \cdots \int_{x_p}^{x_p+dx_p} f(t_1,\dots,t_p)\, dt_p \cdots dt_1$$

$$\simeq dx_1 \cdots dx_p \, f(x_1,\dots,x_p).$$

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

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

(e) If $\underline{X}$ is a continuous r.v. with p.d.f. $f(\cdot)$, then

$$Pr(\underline{X}=\underline{a}) = \int \cdots \int_{\{\underline{x}=\underline{a}\}} f(\underline{t})\, d\underline{t} = 0$$

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

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

$$\frac{\partial^p}{\partial x_1 \cdots \partial x_p} F(x_1,\dots,x_p)$$

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

$$\int_{\mathbb{R}-C} \frac{\partial^p}{\partial x_1 \cdots \partial x_p} F(x_1,\dots,x_p)\, dx_1 \cdots dx_p = 1$$

then $\underline{X}$ is a continuous r.v. with pdf

$$f(\underline{x}) = \begin{cases} \frac{\partial^p}{\partial x_1 \cdots \partial x_p} F(x_1,\dots,x_p), & \text{if } \underline{x} \in \mathbb{R}-C \\ 0, & \text{if } \underline{x} \in C. \end{cases}$$

(h) Let $\underline{X} = (X_1, \dots, X_p)$ be a continuous r.v. with joint pdf $f_{\underline{X}}(x)$ and d.f. $F_{\underline{X}}(x)$. Then, for $q \in \{1, \dots, p-1\}$ and $\underline{x} = (x_1, \dots, x_q) \in \mathbb{R}^q$

$$F_{X_1, \dots, X_q}(x_1, \dots, x_q) = \lim_{\substack{x_j \to \infty \\ j = q+1, \dots, p}} F_{X_1 \dots X_q, X_{q+1} \dots, X_p}(x_1, \dots, x_q, x_{q+1} \dots x_p)$$

$$= \lim_{\substack{x_j \to \infty \\ j = q+1 \dots p}} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_q} \int_{-\infty}^{x_{q+1}} \cdots \int_{-\infty}^{x_p} f_{X_1 \dots X_q, X_{q+1} \dots X_p}(t_1, \dots, t_q, t_{q+1} \dots t_p)\, dt_p \dots dt_1$$

$$= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_q} \Big[\underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 \dots X_q, X_{q+1}, \dots, X_p}(t_1, \dots, t_q, t_{q+1}, \dots, t_p)\, dt_p \dots dt_{q+1}}_{g(t_1 \dots t_q)} \Big] dt_q \dots dt_1$$

$\Rightarrow (X_1, \dots, X_q)$ is a continuous r.v. with pdf

$$f_{X_1, \dots, X_q}(x_1 \dots x_q) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1 \dots X_q, X_{q+1} \dots X_p}(x_1 \dots x_q, t_{q+1}, \dots, t_p)\, dt_p \dots dt_{q+1}$$

Thus marginal distributions of a continuous r.v. $\underline{X}$ are continuous with pdf 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 r.v. (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 dist of X_1) we want to define conditional d.f. of X_2 given $X_1 = x_1$. If X_1 is a continuous r.v. then $\Pr(X_1 = x_1) = 0 \ \forall x_1 \in \mathbb{R}$ and therefore $\Pr(X_2 \le x \mid X_1 = x_1)$ is not defined for any $x_1 \in \mathbb{R}$; although it is defined for discrete r.v. 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

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

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

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

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

$$F_{X_2|X_1}(x|x_1) = \frac{F_{X_1 X_2}(x_1, x) - F_{X_1}(x_1-, x)}{F_{X_1}(x_1) - F_{X_1}(x_1-)}$$

$$= \frac{Pr(X_1 = x_1,\ X_2 \le x)}{Pr(X_1 = x_1)}$$

$$= Pr(X_2 \le x \mid X_1 = x_1).$$

Also if $\underline{X} = (X_1, X_2)$ is a continuous r.v. with pdf $f(x_1, x_2)$, then

$$F_{X_2|X_1}(x|x_1) = \lim_{h \downarrow 0} \frac{\frac{1}{h} \int_{-\infty}^{x} \int_{x_1-h}^{x_1} f_{X_1 X_2}(y_1, y_2)\, dy_1\, dy_2}{\frac{F_{X_1}(x_1) - F_{X_1}(x_1 - h)}{h}}$$

$$= \frac{\int_{-\infty}^{x} f_{X_1 X_2}(x_1, y_2)\, dy_2}{f_{X_1}(x_1)}$$ (provided $f_{X_1}(x_1) > 0$)

$\Rightarrow$ Conditional distribution of X_2 given $X_1 = x_1$ is continuous with p.d.f.

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

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 $\underline{X}_1 = (X_1, \ldots, X_q) = (x_1, \ldots, x_q) = \underline{x}_1$ as

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

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

Definition 4.4.2. Let $\underline{X} = (X_1, \dots, X_p)$ be a p-dimensional r.v. with joint p.d.f. $f_{\underline{X}}(\cdot)$. Let $q \in \{1, 2, \dots, p-1\}$, $\underline{X}_1 = (X_1, \dots, X_q)$, and $\underline{X}_2 = (X_{q+1}, \dots, X_p)$. Let $\underline{x}_1 = (x_1, \dots, x_q) \in S_{\underline{X}_1}$ (support of distn. of $\underline{X}_1$). Then the conditional p.d.f. of $\underline{X}_2$ given $\underline{X}_1 = \underline{x}_1$ is defined by

$$f_{\underline{X}_2 | \underline{X}_1}(\underline{x}_2 | \underline{x}_1) = \frac{f_{\underline{X}_1, \underline{X}_2}(\underline{x}_1, \underline{x}_2)}{f_{\underline{X}_1}(\underline{x}_1)}$$

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

Theorem 4.4.1. Let $\underline{X} = (X_1, \dots, X_p)$ be a continuous r.v. with joint p.d.f. $f_{\underline{X}}(\cdot)$ and marginal p.d.f.s $f_{X_i}(\cdot)$, $i = 1, \dots, p$. Then $X_1, \dots, X_p$ are independent iff

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

Proof. Exercise

Remark 4.4.2. (a) Let $S_{\underline{X}}$ be the support of distribution of $\underline{X} = (X_1, \dots, X_p)$ and let S_{X_i} be the support of distribution of X_i, $i = 1, \dots, p$. It can be shown that if $X_1, \dots, X_p$ are independent then

$$S_{\underline{X}} = \underbrace{\prod_{i=1}^{p} S_{X_i}}_{\text{Cartesian product}}$$

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

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

Theorem 4.4.2. Let $\underline{X} = (X_1, \dots, X_p)$ be a p-dimensional continuous r.v. Then $X_1, \dots, X_p$ are independent iff

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

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

Example 4.4.1. Let $\underline{X} = (X_1, X_2, X_3)$ have the joint p.d.f.

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

(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

$$\int_{\mathbb{R}^3} f_{\underline{X}}(\underline{x})\, d\underline{x} = \int_0^1 \int_0^{x_1} \int_0^{x_2} \frac{1}{x_1 x_2}\, dx_3\, dx_2\, dx_1 = 1$$

(b) For $(x_2, x_3) \in \mathbb{R}^2$,

$$f_{X_2, X_3}(x_2, x_3) = \int_{-\infty}^{\infty} f_{X_1, X_2, X_3}(x_1, x_2, x_3)\, dx_1$$

For $0 < x_3 < x_2 < 1$,

$$f_{X_2, X_3}(x_2, x_3) = \int_{x_2}^{1} \frac{1}{x_1 x_2}\, dx_1 = \frac{-\ln x_2}{x_2}$$

Thus

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

(c) For $x_1 \in \mathbb{R}$

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2, X_3}(x_1, x_2, x_3)\, dx_2\, dx_3.$$

Clearly, for $0 < x_1 < 1$

$$f_{X_1}(x_1) = \int_0^{x_1} \int_0^{x_2} \frac{1}{x_1 x_2}\, dx_3\, dx_2 = 1$$

Thus

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

(d) Let $0 < x_3 < x_2 < 1$. Then

$$f_{X_1|(X_2,X_3)}(x_1|x_2,x_3) = \frac{f_{X_1,X_2,X_3}(x_1,x_2,x_3)}{f_{X_2,X_3}(x_2,x_3)}$$

$$= -\frac{1}{x_1 \ln x_2}, \quad x_2 < x_1 < 1.$$

Thus, for fixed $0 < x_3 < x_2 < 1$

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

(e) We have

$$S_{\underline{X}} = \{ \underline{x} \in \mathbb{R}^3 : 0 \le x_3 \le x_2 \le x_1 \le 1 \}$$

$$\ne S_{X_1} \times S_{X_2} \times S_{X_3} = [0,1] \times [0,1] \times [0,1]$$

Hence 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 $0 < x_2 < 1$

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

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_3|X_2}(x_1,x_3|x_2)\, dx_1\, dx_3 = 1 \Rightarrow c(x_2) = -\frac{1}{x_2 \ln x_2}$$

Thus, for fixed $0 < x_2 < 1$,

$$f_{X_1,X_3|X_2}(x_1,x_3|x_2) = g_{x_2}(x_1)\, h_{x_2}(x_3), \quad (x_1,x_3) \in \mathbb{R}^2,$$

where, for fixed $x_2 \in (0,1)$,

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

$\Rightarrow$ given $X_2 = x_2$ ($x_2 \in (0,1)$) X_1 and X_3 are independently distributed.

$\boxed{23/4}$

4.5. Expectations and Moments

$\underline{X} = (X_1, \ldots, X_p)$: p-dimensional r.v. with p.m.f./p.d.f. $f(\cdot)$ and support S.

$g: \mathbb{R}^p \to \mathbb{R}$: a given function

Definition 4.5.1. We say that the expected value of $g(\underline{X})$ (denoted by $E(g(\underline{X}))$) is finite and equals

$$E(g(\underline{X})) = \begin{cases} \sum_{\underline{x} \in S} g(\underline{x}) f(\underline{x}), & \text{if } \underline{X} \text{ is a discrete r.v.} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\underline{x}) f(\underline{x}) d\underline{x}, & \text{if } X \text{ is a continuous r.v.} \end{cases}$$

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

Theorem 4.5.1. Let $Y = g(\underline{X})$. Then Y has a finite expectation iff

$$\sum_{y \in S_Y} |y| f_Y(y) dy < \infty \quad \left(\text{or } \int_{-\infty}^{\infty} |y| f_Y(y) dy < \infty \right)$$

and in that case

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

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

Some Special Expectations

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

$$\mu'_{k_1, \ldots, k_p} = E(X_1^{k_1} \cdots X_p^{k_p}),$$

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 $k_1, \dots, k_p$

$$\mu_{k_1, \dots, k_p} = E\left((X_1 - E(X_1))^{k_1} \cdots (X_p - E(X_p))^{k_p}\right)$$

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

(c) The quantity

$$Cov(X_1, X_2) = E\left((X_1 - E(X_1))(X_2 - E(X_2))\right),$$

provided it is finite, is called Covariance between X_1 and X_2.

$\boxed{\text{Remark 4.5.1.}}$ (a) $Cov(X_1, X_2) = Cov(X_2, X_1)$

$$= E(X_1 X_2) - E(X_1) E(X_2)$$

(b) $Cov(X_1, X_1) = Var(X_1)$.

(c) $Cov(X_1, X_2) = Cov(X_2, X_1)$

$\boxed{\text{Theorem 4.5.2.}}$ Let a_i, $i = 1, \dots, p$, and b_j, $j = 1, \dots, r$, be real constants and let X_i, $i = 1, \dots, p$, and Y_j, $j = 1, \dots, r$, be r.v.s. 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) $Cov\left(\sum_{i=1}^{p} a_i X_i, \sum_{j=1}^{r} b_j Y_j\right) = \sum_{i=1}^{p} \sum_{j=1}^{r} a_i b_j Cov(X_i, Y_j)$, provided involved expectations are finite.

(c) $$Var\left(\sum_{i=1}^{p} a_i X_i\right) = \sum_{i=1}^{p} a_i^2 Var(X_i) + \sum_{i=1}^{p} \sum_{\substack{j=1 \\ i \neq j}}^{p} a_i a_j Cov(X_i, X_j)$$

$$= \sum_{i=1}^{p} a_i^2 Var(X_i) + 2 \sum\sum_{1 \le i < j \le p} a_i a_j Cov(X_i, X_j).$$

Proof. (For the continuous case)

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

$$= \sum_{i=1}^{p} a_i \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_i\, f_{\underline{X}}(\underline{x})\, d\underline{x}$$

$$= \sum_{i=1}^{p} a_i E(X_i).$$

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

$$= E\left[\left(\sum_{i=1}^{p} a_i X_i - \sum_{i=1}^{p} a_i E(X_i)\right)\left(\sum_{j=1}^{r} b_j Y_j - \sum_{j=1}^{r} b_j E(Y_j)\right)\right]$$

$$= E\left[\left(\sum_{i=1}^{p} a_i (X_i - E(X_i))\right)\left(\sum_{j=1}^{r} b_j (Y_j - E(Y_j))\right)\right]$$

$$= E\left[\sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j (X_i - E(X_i))(Y_j - E(Y_j))\right]$$

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

$$= \sum_{i=1}^{p}\sum_{j=1}^{r} a_i b_j\, \mathrm{Cov}(X_i, Y_j)$$

(c) $$\mathrm{Var}\left(\sum_{i=1}^{p} a_i X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{p} a_i X_i, \sum_{j=1}^{p} a_j X_j\right)$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j\, \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{p} a_i^2\, \mathrm{Cov}(X_i, X_i) + \sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} a_i a_j\, \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{p} a_i^2\, \mathrm{Var}(X_i) + \sum_{\substack{i=1 \\ i\neq j}}^{p}\sum_{j=1}^{p} a_i a_j\, \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{p} a_i^2\, \mathrm{Var}(X_i) + 2\sum\sum_{1\le i<j\le p} a_i a_j\, \mathrm{Cov}(X_i, X_j).$$

Theorem 4.5.3. Let $X_1, \dots, X_p$ be independent r.v.s and let $\psi_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \dots, p$, 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 involved expectations are finite

(b) for any $A_1, \dots, A_p \in \mathcal{B}_p$

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

(c) $\psi_1(X_1), \dots, \psi_p(X_p)$ are independent r.v.s

Proof. (For $p=2$ in continuous case)

(a) $$E(\psi_1(X_1)\psi_2(X_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi_1(x_1)\psi_2(x_2) f_{X_1, X_2}(x_1, x_2)\, dx_1\, dx_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi_1(x_1)\psi_2(x_2) f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1\, dx_2$$

(independence of X_1 and X_2)

$$= \left(\int_{-\infty}^{\infty} \psi_1(x_1) f_{X_1}(x_1)\, dx_1\right)\left(\int_{-\infty}^{\infty} \psi_2(x_2) f_{X_2}(x_2)\, dx_2\right)$$

$$= E(\psi_1(X_1))\, E(\psi_2(X_2))$$

(b) Take

$$\psi_i(X_i) = \begin{cases} 1, & \text{if } X_i \in A_i \\ 0, & \text{otherwise} \end{cases}, \quad i = 1, 2,$$

in (a). Note that

$$\psi_1(X_1)\psi_2(X_2) = \begin{cases} 1 & \text{if } X_1 \in A_1 \ \& \ X_2 \in A_2 \\ 0 & \text{otherwise} \end{cases}$$

$E(\psi_i(X_i)) = \Pr(X_i \in A_i)$, $i = 1, 2$

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

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

Then by (a)

$$E(g_1(X_1)\, g_2(X_2)) = E(g_1(X_1))\, E(g_2(X_2))$$

Also

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

$$= \begin{cases} 1, & \text{if } Y_1 \le y_1,\ Y_2 \le y_2 \\ 0, & \text{otherwise} \end{cases}$$

$$E(g_1(X_1)\, g_2(X_2)) = \Pr(Y_1 \le y_1, Y_2 \le y_2)$$

$$E(g_i(X_i)) = \Pr(Y_i \le y_i), \quad i=1,2.$$

Consequently

$$\Pr(Y_1 \le y_1, Y_2 \le y_2) = \Pr(Y_1 \le y_1)\Pr(Y_2 \le y_2), \quad \forall (y_1, y_2) \in \mathbb{R}^2$$

$\Rightarrow$ $Y_1 = \psi_1(X_1)$ and $Y_2 = \psi_2(X_2)$ are independent r.v.s.

Corollary 4.5.1. Let $X_1, \dots, X_p$ be independent r.v.s. Then

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

(b) for real constants $a_1, \dots, a_p$

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

Proof (a) For $i \ne j$

$$\begin{aligned} Cov(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) \\ &= E(X_i)E(X_j) - E(X_i)E(X_j) \\ &= 0 \end{aligned}$$

(b) $$Var\left(\sum_{i=1}^{p} a_i X_i\right) = \sum_{i=1}^{p} a_i^2 Var(X_i) + \sum_{i=1}^{p}\sum_{\substack{j=1 \\ i \neq j}}^{p} a_i a_j Cov(X_i, X_j)$$

$$= \sum_{i=1}^{p} a_i^2 Var(X_i). \qquad \text{(using (a))}.$$

$\boxed{\text{Definition 4.5.2.}}$ (a) The correlation between r.v.s X_1 and X_2 is defined by

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

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

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

$\boxed{\text{Remark 4.5.2.}}$

X_1 and X_2 are independent r.v.s $\Rightarrow$ X_1 and X_2 are uncorrelated

Converse may not be true

$\boxed{\text{Example 4.5.1.}}$ (R.v.s are uncorrelated does not imply that they are independent)

Let (X, Y) have the joint p.m.f.

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

Then

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

Clearly

$$f_{X,Y}(x, y) \neq f_X(x) f_Y(y), \quad \forall (x, y) \in \mathbb{R}^2$$

$\Rightarrow$ X and Y are not independent (in fact $Pr(Y = X^2) = 1$)

However,

$$E(XY) = E(Y) = 0 \quad \text{and} \quad E(X) = \frac{1}{2}$$

$$\Rightarrow \quad Cov(X, Y) = 0$$

$$\Rightarrow \quad \rho(X, Y) = 0.$$

Theorem 4.5.4. (Cauchy – Schwarz Inequality)

For random variables X and Y

$$(E(XY))^2 \le E(X^2)\, E(Y^2), \qquad \dots\dots (4.5.1)$$

provided involved expectations are finite. The equality is attained iff $Pr(Y = cX) = 1$ or $P(X = cY) = 1$, 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 equality in (4.5.1).

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

Then

$$E((Y - cX)^2) \ge 0 \quad \forall\, c \in \mathbb{R}$$

$$\Rightarrow c^2 E(X^2) - 2c\, E(XY) + E(Y^2) \ge 0 \quad \forall\, c \in \mathbb{R}$$

$$\Rightarrow \quad \text{Discriminant} \le 0$$

$$\Rightarrow \quad (2E(XY))^2 - 4(E(X^2)\, E(Y^2)) \le 0$$

$$\Rightarrow \quad (E(XY))^2 \le E(X^2)\, E(Y^2).$$

Clearly equality is attained iff

$$E((Y - cX)^2) = 0, \quad \text{for some } c \in \mathbb{R}$$

$$\Rightarrow \quad Pr(Y = cX) = 1 \quad \text{for some } c \in \mathbb{R}.$$

By symmetry $P(X = cY) = 1$ for some $c \in \mathbb{R}$

Corollary 4.5.2. Let X_1 and X_2 be r.v.s with $E(X_i) = \mu_i \in (-\infty, \infty)$ and $Var(X_i) = \sigma_i^2 \in (0, \infty)$, $i = 1, 2$. Then

(a) $|\rho(X_1, X_2)| \le 1$

(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 C-S inequality

$$(E(XY))^2 \le E(X^2)\, E(Y^2)$$

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

$$\left(E\left(\frac{(X_1-\mu_1)(X_2-\mu_2)}{\sigma_1\sigma_2}\right)\right)^2 \le 1$$

$$\Rightarrow \quad \rho^2(X_1, X_2) \le 1$$

$$\Rightarrow \quad |\rho(X_1, X_2)| \le 1$$

and equality is attained iff

$$\Pr(X = cY) = 1, \quad \text{for some real constant } c$$

$$\Rightarrow P\left(\frac{X_1-\mu_1}{\sigma_1} = c\,\frac{X_2-\mu_2}{\sigma_2}\right) = 1 \quad \text{for some real constant } c.$$

4.6. Conditional Expectation, Conditional Variance and Conditional Covariance

Definition 4.6.1. (a) Let $\underline{X}$ be a p-dimensional r.v. and let $\underline{Y}$ be a q-dimensional r.v.. Let $y \in \mathbb{R}^q$ 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 r.v. $\underline{Y}$. Then

(i) the conditional expectation of $\psi(\underline{X})$ given $\underline{Y} = \underline{y}$ (denoted by $E(\psi(\underline{X}) \mid \underline{Y} = \underline{y})$) 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 $Var(\psi(\underline{X}) \mid \underline{Y} = \underline{y})$) 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 $Cov(X_1, X_2 \mid \underline{Y} = \underline{y})$) 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 $\underline{y} \in \{ \underline{t} \in \mathbb{R}^q : f_{\underline{Y}}(\underline{t}) > 0 \}$,

$$\psi_1(\underline{y}) = E(\psi(\underline{X}) \mid \underline{Y} = \underline{y})$$

$$\psi_2(\underline{y}) = Var(\psi(\underline{X}) \mid \underline{Y} = \underline{y})$$

and $$\psi_3(\underline{y}) = Cov(X_1, X_2 \mid \underline{Y} = \underline{y}).$$

We denote

$$E(\psi(\underline{X}) \mid \underline{Y}) \equiv \psi_1(\underline{Y})$$

$$Var(\psi(\underline{X}) \mid \underline{Y}) \equiv \psi_2(\underline{Y})$$

and $$Cov(X_1, X_2 \mid \underline{Y}) \equiv \psi_3(\underline{Y}).$$

Theorem 4.6.1. Under the above notation

(a) $E(\psi(\underline{X})) = E(E(\psi(\underline{X}) \mid \underline{Y}))$;

(b) $Var(\psi(\underline{X})) = Var(E(\psi(\underline{X}) \mid \underline{Y})) + E(Var(\psi(\underline{X}) \mid \underline{Y}))$

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

Proof. For $p = q = 1$ and for continuous case.

(a) $$E(\underbrace{E(\psi(\underline{X}) \mid \underline{Y})}_{\text{fn of } \underline{Y}}) = \int_{-\infty}^{\infty} E(\psi(\underline{X}) \mid \underline{Y} = \underline{y}) f_{\underline{Y}}(\underline{y})\, dy$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \psi(x) f_{X|Y}(x|y)\, dx \right] f_{\underline{Y}}(y)\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x) f_{X|Y}(x|y) f_Y(y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x) f_{X,Y}(x,y)\, dx\, dy$$

$$= E(\psi(X)).$$

(b) Follows from (c)

(c) $$Cov(X_1, X_2) = E((X_1 - E(X_1))(X_2 - E(X_2)))$$

$$= E\left[E[(X_1 - E(X_1))(X_2 - E(X_2)) \mid \underline{Y}] \right] \quad (\text{by (a)})$$

Now

$$E((X_1 - E(X_1))(X_2 - E(X_2)) \mid \underline{Y}) = E[(X_1 - E(X_1 \mid \underline{Y}) + E(X_1 \mid \underline{Y}) - E(X_1)) (X_2 - E(X_2 \mid Y) + E(X_2 \mid Y) - E(X_2)) \mid \underline{Y})$$

$$= E[(X_1 - E(X_1|Y))(X_2 - E(X_2|Y))|Y]$$
$$+ (E(X_1|Y) - E(X_1))(E(X_2|Y) - E(X_2))$$

$$= Cov(X_1, X_2|Y) + (E(X_1|Y) - E(X_1))(E(X_2|Y) - E(X_2))$$

$$\Rightarrow Cov(X_1, X_2) = E[Cov(X_1, X_2|Y)] + E[(E(X_1|Y) - E(X_1))(E(X_2|Y) - E(X_2))]$$
$$= E[Cov(X_1, X_2|Y)] + Cov(E(X_1|Y), E(X_2|Y)).$$

4.7. Joint Moment Generating Function

$\underline{X} = (X_1, \dots, X_p)$: a p-dimensional r.v. with p.d.f./p.m.f. $f_{\underline{X}}(\cdot)$

$A = \{\underline{t} = (t_1, \dots, t_p) \in \mathbb{R}^p : E(e^{\sum_{i=1}^{p} t_i X_i}) < \infty\}$

Definition 4.7.1. (a) The function $M_{\underline{X}} : A \to \mathbb{R}$ defined by

$$M_{\underline{X}}(\underline{t}) = E(e^{\sum_{i=1}^{p} t_i X_i}), \qquad \underline{t} = (t_1, \dots, t_p) \in A$$

is called the joint moment generating function (m.g.f.) of r.v. $\underline{X} = (X_1, \dots, X_p)$.

Notation: For $\underline{a} = (a_1, \dots, a_p) \in \mathbb{R}^p$, $-\underline{a} = (-a_1, \dots, -a_p)$ and $(-\underline{a}, \underline{a}) = (-a_1, a_1) \times \dots \times (-a_p, a_p)$; $\underline{a} = (a_1, \dots, a_p) > 0 \Leftrightarrow a_i > 0, \ i = 1, \dots, p$

Remark 4.7.1. (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 $X_1, \dots, X_p$ are independent then

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

Conversely suppose that $A \supseteq (-\underline{a}, \underline{a})$ for some $\underline{a} > 0$ and

$$M_{\underline{X}}(\underline{t}) = \prod_{i=1}^{p} M_{X_i}(t_i) \quad \forall\ \underline{t} \in A$$

then it can be shown that $X_1, \ldots, X_p$ are independent.

(iii) Let $X_1, \ldots, X_p$ be independent r.v.s and let $Y = \sum_{i=1}^{p} X_i$. Then

$$M_Y(t) = E\left(e^{t \sum_{i=1}^{p} X_i}\right) = E\left(\prod_{i=1}^{p} e^{tX_i}\right)$$

$$= \prod_{i=1}^{p} E\left(e^{tX_i}\right)$$

$$= \prod_{i=1}^{p} M_{X_i}(t), \quad t \in A$$

In particular if $X_1, \ldots, X_p$ are iid (independent and identically distributed) with common m.g.f. $M(t)$, then

$$M_Y(t) = [M(t)]^p, \quad t \in A.$$

Theorem 4.7.1. Suppose that the joint m.g.f. $M_{\underline{X}}(\underline{t})$ is finite on a rectangle $(-\underline{a}, \underline{a}) \subseteq \mathbb{R}^p$, $\underline{a} > \underline{0}$. Then $M_{\underline{X}}(\underline{t})$ possesses partial derivatives of all orders in $(-\underline{a}, \underline{a})$. Furthermore, for non-negative integers $k_1, \ldots, k_p$,

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

Proof. (Outline).

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

$$= \int_{\mathbb{R}^p} e^{\sum_{i=1}^{p} t_i \lambda_i} f_{\underline{X}}(\underline{\lambda})\, d\underline{\lambda}$$

$$\frac{\partial^{k_1 + \cdots + k_p}}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} M_{\underline{X}}(\underline{t}) = \int_{\mathbb{R}^p} \lambda_1^{k_1} \cdots \lambda_p^{k_p}\, e^{\sum_{i=1}^{p} t_i \lambda_i} f_{\underline{X}}(\underline{\lambda})\, d\underline{\lambda}$$

$$\left[\frac{\partial^{k_1 + \cdots + k_p}}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} M_{\underline{X}}(\underline{t})\right]_{\underline{t} = \underline{0}} = \int_{\mathbb{R}^p} \lambda_1^{k_1} \cdots \lambda_p^{k_p} f_{\underline{X}}(\underline{\lambda})\, d\underline{\lambda}$$

$$= E\left(X_1^{k_1} \cdots X_p^{k_p}\right)$$

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

$$E(X_i) = \left[\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t}) \right]_{\underline{t}=\underline{0}} = \left[\frac{\partial}{\partial t_i} \psi_{\underline{X}}(\underline{t}) \right]_{\underline{t}=\underline{0}}$$

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

$$Var(X_i) = \left[\frac{\partial^2}{\partial t_i^2} M_{\underline{X}}(\underline{t}) \right]_{\underline{t}=\underline{0}} - \left[\left(\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t}) \right)_{\underline{t}=\underline{0}} \right]^2$$

$$= \left[\frac{\partial^2}{\partial t_i^2} \psi_{\underline{X}}(\underline{t}) \right]_{\underline{t}=\underline{0}}, \quad i = 1, \ldots, p,$$

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 $\underline{a} > 0$,

$$Cov(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j) = E[(X_i - E(X_i))(X_j - E(X_j))]$$

$$= \left[\frac{\partial^2}{\partial t_i \partial t_j} M_{\underline{X}}(\underline{t}) \right]_{\underline{t}=\underline{0}} - \left[\frac{\partial}{\partial t_i} M_{\underline{X}}(\underline{t}) \right]_{\underline{t}=\underline{0}} \left[\frac{\partial}{\partial t_j} M_{\underline{X}}(\underline{t}) \right]_{\underline{t}=\underline{0}}$$

$$= \left[\frac{\partial^2}{\partial t_i \partial t_j} \psi_{\underline{X}}(\underline{t}) \right]_{\underline{t}=\underline{0}}.$$

Moreover

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

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

$$= M_{X_i, X_j}(t_i, t_j),$$

provided the m.g.f. is finite.

4.8. Equality in Distribution

$\boxed{\text{Definition 4.8.1.}}$ 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 $F_{\underline{X}}(\underline{x}) = F_{\underline{Y}}(\underline{x})$, $\forall\, \underline{x} \in \mathbb{R}^p$.

$\boxed{\text{Theorem 4.8.1.}}$ (a) Let $\underline{X}$ and $\underline{Y}$ be discrete r.v.s with p.m.f.s $f_{\underline{X}}(\cdot)$ and $f_{\underline{Y}}(\cdot)$, respectively. Then

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

(b) Let $\underline{X}$ and $\underline{Y}$ be continuous r.v.s. Then

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

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 r.v.s and let $\psi: \mathbb{R}^p \to \mathbb{R}^q$ be a given function. Then

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

(d) Let $\underline{X}$ and $\underline{Y}$ be p-dimensional r.v.s 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 $\underline{a} > \underline{0}$. Then

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

4.9. Some Generalizations

$\underset{\sim}{X}_i$: a p_i-dimensional random vector, $i=1,\dots,m$

$F_{\underline{X}_i}$: d.f. of $\underline{X}_i$, $i=1,\dots,m$; $f_{\underline{X}_i}$: p.m.f./p.d.f. of $\underline{X}_i$, $i=1,\dots,m$

$\sum_{i=1}^{p} p_i = p$

$\underline{X} = (\underline{X}_1,\dots,\underline{X}_m)$: p-dimensional r.v. with d.f. $F_{\underline{X}}(\cdot)$ and p.m.f./p.d.f. $f_{\underline{X}}(\cdot)$.

Definition 4.9.1. The random vectors $\underline{X}_1,\dots,\underline{X}_m$ are said to be independent if for any subcollection $\{\underline{X}_{i_1},\dots,\underline{X}_{i_q}\}$ of $\{\underline{X}_1,\dots,\underline{X}_m\}$ $(2\le q\le m)$

$$F_{\underline{X}_{i_1},\dots,\underline{X}_{i_q}}(\underline{x}_1,\dots,\underline{x}_q) = \prod_{j=1}^{q} F_{\underline{X}_{i_j}}(\underline{x}_j), \quad \forall\, \underline{x}=(\underline{x}_1,\dots,\underline{x}_q)\in\mathbb{R}^{\sum_{j=1}^{q} p_{i_j}}$$

Remark 4.9.1. $\underline{X}_1,\dots,\underline{X}_m$ are independent $\Rightarrow$ r.v.s in any subset of $\{\underline{X}_1,\dots,\underline{X}_m\}$ are independent

Theorem 4.9.1. (a) The following statements are equivalent

(i) $\underline{X}_1,\dots,\underline{X}_m$ are independent random vectors

(ii) $F_{\underline{X}_1,\dots,\underline{X}_m}(\underline{x}_1,\dots,\underline{x}_m) = \prod_{i=1}^{m} F_{\underline{X}_i}(\underline{x}_i) \quad \forall\, \underline{x}=(\underline{x}_1,\dots,\underline{x}_m)\in\mathbb{R}^p$

(iii) $f_{\underline{X}_1,\dots,\underline{X}_m}(\underline{x}_1,\dots,\underline{x}_m) = \prod_{i=1}^{m} f_{\underline{X}_i}(\underline{x}_i), \quad \forall\, \underline{x}=(\underline{x}_1,\dots,\underline{x}_m)\in\mathbb{R}^p$

(iv) $f_{\underline{X}_1,\dots,\underline{X}_m}(\underline{x}_1,\dots,\underline{x}_m) = \prod_{i=1}^{m} g_i(\underline{x}_i), \quad \forall\, \underline{x}=(\underline{x}_1,\dots,\underline{x}_m)\in\mathbb{R}^p$, for some real-valued functions $g_i:\mathbb{R}^{p_i}\to\mathbb{R}$, $i=1,\dots,m$.

(v) $\Pr(\underline{X}_i\in A_i,\ i=1,\dots,m) = \prod_{i=1}^{m}\Pr(\underline{X}_i\in A_i), \quad \forall\, A_i\in\mathcal{B}_{p_i},\ i=1,\dots,m$

(b) If $\underline{X}_1,\dots,\underline{X}_m$ are independent random vectors then

(i) $E\left(\prod_{i=1}^{m}\psi_i(\underline{X}_i)\right) = \prod_{i=1}^{m} E(\psi_i(\underline{X}_i))$ for any functions ψ_i, $i=1,\dots,m$.

(ii) $\psi_1(\underline{X}_1),\dots,\psi_m(\underline{X}_m)$ are independent random vectors for any functions $\psi_1,\dots,\psi_m$.

Definition 4.9.2. Let Λ be an arbitrary index set. The random vectors $\{\underline{X}_\lambda : \lambda \in \Lambda\}$ are said to be independent if r.v.s in any finite subcollection of $\{\underline{X}_\lambda : \lambda \in \Lambda\}$ are independent.

Theorem 4.9.2 Under the notation of Theorem 4.9.1,

$\underline{X}_1, \ldots, \underline{X}_m$ are independent r.v.s

$\Leftrightarrow$ for some $\underline{a} > \underline{0}$ and $\forall\ \underline{t} = (\underline{t}_1, \ldots, \underline{t}_m) \in (-\underline{a}, \underline{a})$

$$M_{\underline{X}}(\underline{t}_1, \ldots, \underline{t}_m) = \prod_{i=1}^{m} M_{\underline{X}_i}(\underline{t}_i).$$

4.10. Functions of a Random Vector

$\underline{X} = (X_1, \ldots, X_p)$: a p-dimensional r.v. with p.m.f./p.d.f. $f(\cdot)$

$g : \mathbb{R}^p \to \mathbb{R}^q$; where $1 \le q \le p$: 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 4.10.1. (a) Let $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ be a collection of independent and identically distributed (i.i.d.) r.v.s each having the same (joint) d.f. F and the same p.m.f./p.d.f. $f(\cdot)$. We call $\underline{X}_1, \ldots, \underline{X}_n$ a random sample (r.s.) of size n from a distribution having d.f. F (p.m.f./p.d.f. f). In other words a random sample is a collection of i.i.d. r.v.s.

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

Example 4.10.1 Let $X_1, \ldots, X_n$ be a random sample from a distribution having p.d.f.

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

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

- $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ or $S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ — Sample Variance
- $X_{r:n}$ = r-th smallest of $X_1, \ldots, X_n$, $r = 1, \ldots, n$ so that $X_{1:n} \le \cdots \le X_{n:n}$ — r-th order statistic, $r = 1, 2, \ldots, n$
- $X_{[np]:n}$, $0 < p < 1$; $[x]$ = largest integer $\le x$ — p-th sample quantile.
- $X_{[\frac{n}{4}]:n}$ — Sample lower quantile
- $X_{[\frac{3n}{4}]:n}$ — Sample upper quantile
- $M = \begin{cases} X_{\frac{n+1}{2}:n}, & \text{if } n \text{ is odd} \\ \dfrac{X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}}{2}, & \text{if } n \text{ is even} \end{cases}$ — Sample Median
- $S_n = \sqrt{S_n^2}$ or $S_{n-1} = \sqrt{S_{n-1}^2}$ — Sample Standard Deviation
- $r = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y})^2\right)}}$ — Sample Correlation Coefficient

Let $X_1, \ldots, X_n$ 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 $\underline{X} = (X_1, \ldots, X_n)$ is

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

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

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

Theorem 4.10.1. If $X_1, \ldots, X_n$ is a random sample, then

$$(X_1, \ldots, X_n) \overset{d}{=} (X_{\beta_1}, \ldots, X_{\beta_n})$$

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

Example 4.10.2 Let $X_1, \ldots, X_n$ be a random sample from a given distribution.

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

(b) If X_1 is a continuous r.v. then, for any $r \in \{1, \ldots, n\}$,

$$\Pr(X_i = X_{r:n}) = \frac{1}{n}, \quad i = 1, \ldots, n;$$

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

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

Solution (a) X_1 is a continuous r.v. $\Rightarrow$ $\underline{X} = (X_1, \ldots, X_n)$ is a continuous r.v. (Why?)

$\Rightarrow$ $(X_1, \ldots, X_n) \overset{d}{=} (X_{\beta_1}, \ldots, X_{\beta_n})$ for any permutation $(\beta_1, \ldots, \beta_n)$ of $(1, \ldots, n)$

and

$$\Pr(\text{all } X_i\text{s are distinct}) = 1$$

$\Rightarrow$ $(X_1, \ldots, X_n) \overset{d}{=} (X_{\beta_1}, \ldots, X_{\beta_n})$, for any permutation $(\beta_1, \ldots, \beta_n)$ of $(1, \ldots, n)$

and

$\sum_{\beta \in S_n} \Pr(X_{\beta_1} < \cdots < X_{\beta_n}) = 1$, where S_n is the set of all permutations of $(1, \ldots, n)$

$\Rightarrow$ $\Pr(X_{\beta_1} < \cdots < X_{\beta_n}) = P(X_1 < \cdots < X_n) = \frac{1}{\underline{|n}}$

(b) For any $i = 1, \ldots, n$

$$(X_1, X_2, \ldots, X_i, \ldots X_n) \overset{d}{=} (X_i, X_2, \ldots, X_1, \ldots, X_n)$$

$\Rightarrow$ $X_{r:n}$ r-th smallest of $(X_1, X_2, \ldots, X_i \ldots X_n)$

$\overset{d}{=}$ r-th smallest of $(X_i, X_2, \ldots, X_1, \ldots, X_n)$

and

$Pr(X_1 =$ r-th smallest of $(X_1 \ldots, X_i, \ldots X_n))$

$= Pr(X_i =$ r-th smallest of $(X_i, X_2, \ldots, X_1 \ldots X_n))$

$\Rightarrow$ $Pr(X_1 = X_{r:n}) = Pr(X_i = X_{r:n})$, $i = 1 \ldots n$.

Since $Pr(X_{1:n} < X_{2:n} < \ldots < X_{n:n}) = 1$ (by (a)), we have

$$\sum_{i=1}^{n} Pr(X_i = X_{r:n}) = 1$$

$$\Rightarrow \quad Pr(X_i = X_{r:n}) = Pr(X_1 = X_{r:n}) = \frac{1}{n}.$$

(c) $(X_1, X_2, \ldots, X_i, \ldots X_n) \overset{d}{=} (X_i, X_2, \ldots, X_1, \ldots X_n)$

$$\Rightarrow \quad E\left(\frac{X_1}{X_1 + X_2 + \cdots + X_i + \cdots + X_n}\right) = E\left(\frac{X_i}{X_i + X_2 + \cdots + X_1 + \cdots + X_n}\right)$$

$$\Rightarrow \quad E\left(\frac{X_1}{\sum_{j=1}^{n} X_j}\right) = E\left(\frac{X_i}{\sum_{j=1}^{n} X_j}\right)$$

But

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

$$\Rightarrow \quad E\left(\frac{X_i}{\sum_{j=1}^{n} X_j}\right) = E\left(\frac{X_1}{\sum_{j=1}^{n} X_j}\right) = \frac{1}{n}, \quad i = 1, \ldots, n$$

(d) $(X_1, X_2, X_i, \ldots, X_n) \stackrel{d}{=} (X_i, X_2, \ldots, X_1, \ldots, X_n)$

$$\Rightarrow E(X_1 \mid \underbrace{X_1+X_2+\cdots+X_i+\cdots+X_n}_{=t}) = E(X_i \mid X_i+X_2+\cdots+X_1+\cdots+X_n = t)$$

$$\Rightarrow E(X_1 \mid \sum_{j=1}^{n} X_j = t) = E(X_i \mid \sum_{j=1}^{n} X_j = t)$$

But

$$\sum_{i=1}^{n} E(X_i \mid \sum_{j=1}^{n} X_j = t) = E(\sum_{i=1}^{n} X_i \mid \sum_{j=1}^{n} X_j = t) = t$$

Therefore

$$E(X_i \mid \sum_{j=1}^{n} X_j = t) = E(X_1 \mid \sum_{j=1}^{n} X_j = t) = \frac{t}{n}, \quad i=1,\ldots,n.$$

4.10.1. Distribution Function Technique

$\underline{X} = (X_1, \ldots, X_p)$: a p-dimensional r.v. with d.f. F and p.m.f./p.d.f $f(\cdot)$

$\underline{g}: \mathbb{R}^p \to \mathbb{R}^q$: $\underline{g} = (g_1, \ldots, g_q)$

$\underline{Y} = (Y_1, \ldots, Y_q) = (g_1(\underline{X}), \ldots, g_q(\underline{X}))$

We are interested in the distribution of r.v. $\underline{Y}$.

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

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

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

Example 4.10.1.1 Let $X_1, \ldots, X_n$ be a random sample from a distribution having d.f. F, p.m.f./p.d.f. f and support S. Let $Y_1 = \min\{X_1, \ldots, X_n\}$ and $Y_2 = \max\{X_1, \ldots, X_n\}$.

(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 (i.e. without using (a));

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

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

$$F_{\underline{Y}}(y_1, y_2) = \Pr(Y_1 \le y_1,\ Y_2 \le y_2)$$

$$= \Pr(\min\{X_1, \dots, X_n\} \le y_1,\ \max\{X_1, \dots, X_n\} \le y_2)$$

$$= \Pr(\max\{X_1, \dots, X_n\} \le y_2) - \Pr(\min\{X_1, \dots, X_n\} > y_1,\ \max\{X_1, \dots, X_n\} \le y_2)$$

$$= \Pr(X_i \le y_2,\ i = 1, \dots, n) - \Pr(X_i > y_1\ i = 1, \dots, n,\ X_i \le y_2\ i = 1, \dots, n)$$

$$= \prod_{i=1}^{n} \Pr(X_i \le y_2) - \Pr(y_1 < X_i \le y_2\ \ i = 1, \dots, n)$$

$$= \prod_{i=1}^{n} \Pr(X_i \le y_2) - \prod_{i=1}^{n} \Pr(y_1 < X_i \le y_2)$$

$$= \begin{cases} [F(y_2)]^n - [F(y_2) - F(y_1)]^n, & -\infty < y_1 < y_2 < \infty \\ [F(y_2)]^n, & -\infty < y_2 \le y_1 < \infty \end{cases}$$

(b) $$F_{Y_1}(y_1) = \lim_{y_2 \to \infty} F_{\underline{Y}}(y_1, y_2) = 1 - (1 - F(y_1))^n \qquad -\infty < y_1 < \infty$$

$$F_{Y_2}(y_2) = \lim_{y_1 \to \infty} F_{\underline{Y}}(y_1, y_2) = [F(y_2)]^n, \qquad -\infty < y_2 < \infty$$

(c) $$F_{Y_1}(y_1) = \Pr(Y_1 \le y_1)$$

$$= \Pr(\min\{X_1, \dots, X_n\} \le y_1)$$

$$= 1 - \Pr(\min\{X_1, \dots, X_n\} > y_1)$$

$$= 1 - \Pr(X_i > y_1\ \ i = 1, \dots, n)$$

$$= 1 - \prod_{i=1}^{n} P(X_i > y_1)$$

$$= 1 - [1 - F(y_1)]^n, \qquad -\infty < y_1 < \infty$$

$$F_{Y_2}(y_2) = \Pr(Y_2 \le y_2)$$

$$= \Pr(\max\{X_1, \dots, X_n\} \le y_2)$$

$$= \Pr(X_i \le y_2,\ i = 1, \dots, n)$$

$$= \prod_{i=1}^{n} \Pr(X_i \le y_2)$$

$$= [F(y_2)]^n, \qquad -\infty < y_2 < \infty$$

(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 (i.e. it does not contain any open interval))

$$f_{Y_1}(y) = \frac{d}{dy}\left[1 - (1-F(y))^n\right]$$

$$= n(1-F(y))^{n-1} f(y), \quad -\infty < y < \infty$$

$$f_{Y_2}(y) = \frac{d}{dy}\left[(F(y))^n\right]$$

$$= n[F(y)]^{n-1} f(y), \quad -\infty < y < \infty$$

Example 4.10.1.2 Let X_1 and X_2 be i.i.d. r.v.s with common p.d.f.

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

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

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

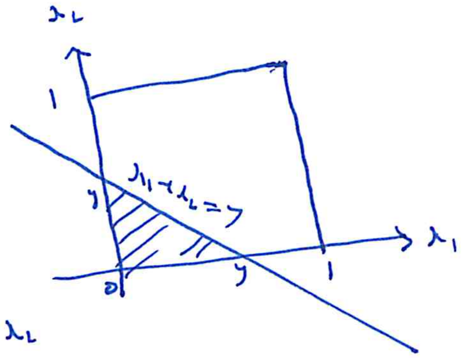
For $y \in \mathbb{R}$

$$F_Y(y) = Pr(Y \le y) = Pr(X_1 + X_2 \le y) = \underset{x_1+x_2 \le y}{\int_0^1 \int_0^1} 4x_1x_2 \, dx_1 \, dx_2$$

Clearly, for $y<0$, $F_Y(y) = 0$ and, for $y \ge 2$, $F_Y(y) = 1$.

Now consider $y \in [0, 1)$.

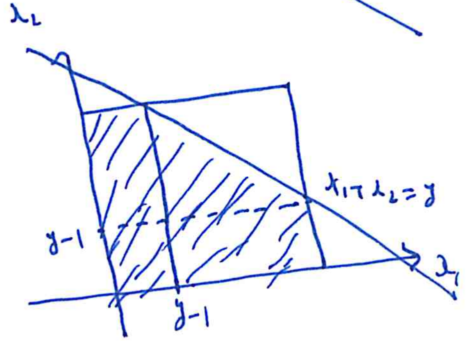
$$F_Y(y) = \int_0^y \int_0^{y-x_1} 4x_1x_2 \, dx_2 \, dy_1 = \frac{y^4}{6}$$



For $y \in [1, 2)$

$$F_Y(y) = \int_0^{y-1} \int_0^1 4x_1x_2 \, dx_2 \, dy_1 + \int_{y-1}^1 \int_0^{y-x_1} 4x_1x_2 \, dx_2 \, dx_1$$

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



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 a continuous r.v. with p.d.f.

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

4.1.10.2. Transformation of Variable Technique

Theorem 4.10.2.1 Let $\underline{X} = (X_1, \ldots, X_p)$ be a p-dimensional discrete r.v. with support S and p.m.f. $f(\cdot)$. Let $g_i : \mathbb{R}^p \to \mathbb{R}$, $i = 1, \ldots, k$ and $Y_i = g_i(\underline{X})$, $i = 1, \ldots, k$, where $1 \le k \le p$ is an integer. Then $\underline{Y} = (Y_1, \ldots, Y_k)$ is a discrete r.v. with support $T = \{(y_1, \ldots, y_k) : y_i = g_i(x_1, \ldots, x_p),\ i = 1, \ldots, k, \text{ for some } \underline{x} = (x_1, \ldots, x_p) \in S\}$,

d.f.

$$G(\underline{y}) = G(y_1, \ldots, y_k) = \sum_{\underline{x} \in A_{\underline{y}}} f(\underline{x}), \quad \underline{y} \in \mathbb{R}^k$$

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

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

and

$$B_{\underline{y}} = \{\underline{x} = (x_1, \ldots, x_p) \in S : g_i(\underline{x}) = y_i,\ i = 1, \ldots, k\}$$

$\boxed{\text{Example 4. 10.2.2}}$ Let $X_1, \dots, X_p$ be independent rvs with X_i having the p.m.f. (Binomial distribution)

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

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

$\boxed{\text{Solution}}$ The joint p.m.f. of $\underline{X} = (X_1, \dots, X_p)$ is

$$f_{\underline{X}}(\underline{x}) = \prod_{i=1}^{p} f_i(x_i)$$

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

where $n = \sum_{i=1}^{p} n_i$

Clearly

$$f_Y(y) = \Pr(X_1 + \dots + X_p = y) = 0, \quad \text{if } y \notin \{0, 1, \dots, n\}.$$

For $y \in \{0, 1, \dots, n\}$.

$$f_Y(y) = \Pr(Y = y)$$

$$= \Pr(X_1 + \dots + X_p = y)$$

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

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

$$= \binom{n}{y} \theta^y (1-\theta)^{n-y}.$$

Thus

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

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

$$f_i(x) = \begin{cases} \dfrac{e^{-\lambda_i} \lambda_i^x}{\lfloor x}, & x \in \{0, 1, 2, \dots\} \\ 0, & \text{otherwise} \end{cases},$$

where $\lambda_i > 0$, $i = 1, \dots, p$ are fixed real constants. Show that the p.m.f. of $Y = \sum_{i=1}^{p} X_i$ is

$$f_Y(y) = \begin{cases} \dfrac{e^{-\lambda} \lambda^y}{\lfloor y}, & y \in \{0, 1, 2, \dots\} \\ 0, & \text{otherwise} \end{cases},$$

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

$\boxed{\text{Theorem 4.10.2.2.}}$ Let $\underline{X} = (X_1, \dots, X_p)$ be a continuous r.v. 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 ($S_i \cap S_j = \phi$ $\forall i \neq j$, and $\bigcup_{i \in \Delta} S_i = S$). Suppose that $h_j : \mathbb{R}^p \to \mathbb{R}$, $j = 1, \dots, p$, are functions such that $\underline{h} = (h_1, \dots, h_p) : S_i^0 \to \mathbb{R}$ is one-to-one with inverse transformation $\underline{h}_i^{-1}(\underline{t}) = (h_{1i}^{-1}(\underline{t}), \dots, h_{pi}^{-1}(\underline{t}))$, $i \in \Delta$; here S_i^0 denotes the interior of S_i, $i \in \Delta$. Further suppose that $h_{ji}^{-1}(\underline{t})$, $j = 1, \dots, p$, $i \in \Delta$, have continuous partial derivatives and the Jacobian determinants

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

Define $\underline{h}(S_j^o) := \{\underline{h}(\underline{x}) = (h_1(\underline{x}), \ldots, h_p(\underline{x})) \in \mathbb{R}^p : \underline{x} \in S_j\}$, $j \in \Lambda$ and $T_l = h_l(x_1, \ldots, x_p)$, $l = 1, \ldots, p$. Then the r.v. $\underline{T} = (T_1, \ldots, T_p)$ is a continuous r.v. with p.d.f.

$$f_{\underline{T}}(\underline{t}) = \sum_{j \in \Lambda} f(h_{1j}^{-1}(\underline{t}), \ldots, h_{pj}^{-1}(\underline{t})) \, |J_j| \, I_{h(S_j^o)}(\underline{t})$$

$\boxed{\text{Corollary 4. .10.2.1.}}$ Under the notation and assumptions of the above theorem suppose that $\underline{h} = (h_1, \ldots, h_p) : S^o \to \mathbb{R}^p$ is one-to-one with inverse transformation $\underline{h}^{-1}(\underline{t}) = (h_1^{-1}(\underline{t}), \ldots, h_p^{-1}(\underline{t}))$ (say); here S^o denotes the interior of S. Further suppose that $h_i^{-1}(\underline{t})$, $i = 1, \ldots, p$ have continuous partial derivatives and the Jacobian determinant

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

Define $\underline{h}(S^o) = \{\underline{h}(\underline{x}) : \underline{x} \in S\}$ and $T_j = h_j(x_1, \ldots, x_p)$, $j = 1, \ldots, p$. Then the r.v. $\underline{T} = (T_1, \ldots, T_p)$ is a continuous r.v. with joint p.d.f.

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

$\boxed{\text{Example 4.10.2.3.}}$ Let X_1 and X_2 be i.i.d. r.v.s with common pdf

$$f(x) = \begin{cases} e^{-x}, & x>0 \\ 0, & \text{o.w.} \end{cases}$$

Find the p.d.f. of $Y = \frac{X_1}{X_1+X_2}$.

$\boxed{\text{Solution}}$ 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 $S = [0, \infty) \times [0, \infty)$, $S^\circ = (0,\infty)\times(0,\infty)$. 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 $\underline{h}: S^\circ \to \mathbb{R}^2$ is 1-1; here $\underline{h} = (h_1, h_2)$. We have

$$h_1(x_1, x_2) = \frac{x_1}{x_1+x_2} = y$$

$$h_2(x_1, x_2) = x_1+x_2 = z$$

$$\Rightarrow \quad x_1 = h_1^{-1}(y, z) = yz \quad \text{and} \quad x_2 = h_2^{-1}(y, z) = z(1-y)$$

$$\underline{x} \in S^\circ \Leftrightarrow x_1>0,\ x_2>0 \Leftrightarrow yz>0,\ z(1-y)>0 \Leftrightarrow 0<y<1,\ z>0$$

Thus

$$\underline{h}(S^\circ) = (0, 1) \times (0, \infty).$$

$$J = \begin{vmatrix} \frac{\partial h_1^{-1}(y,z)}{\partial y} & \frac{\partial h_1^{-1}(y,z)}{\partial z} \\ \frac{\partial h_2^{-1}(y,z)}{\partial y} & \frac{\partial h_2^{-1}(y,z)}{\partial z} \end{vmatrix} = \begin{vmatrix} z & y \\ -z & 1-y \end{vmatrix} = z$$

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

$$f_{Y,Z}(y, z) = f_{\underline{X}}(yz, z(1-y))\, |z|\, I_{(0,1)\times(0,\infty)}(y, z)$$

$$= \begin{cases} z e^{-z}, & 0<y<1,\ z>0 \\ 0, & \text{otherwise} \end{cases}$$

$$= f_Y(y)\, f_Z(z),$$

where

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

and

$$f_Z(z) = \begin{cases} z e^{-z}, & z>0 \\ 0, & \text{otherwise} \end{cases}$$

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

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

Example 4.10.2.4. Let X_1 and X_2 be i.i.d. 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 $S^0 = ((-2,-1) \cup (0,3)) \cup ((-2,-1) \times (0,3)))$

$= S_1^0 \cup S_2^0 \cup S_3^0 \cup S_4^0$, 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_i^0, $\underline{h}(\underline{x}) = (h_1(\lambda_1, \lambda_2), h_2(\lambda_1, \lambda_2)) = (y_1, y_2) = (|\lambda_1| + |\lambda_2|, |\lambda_1|)$ is 1-1. Now proceed.

4.1.10.3 Moment Generating Function Technique

Let $\underline{X} = (X_1, \ldots, X_p)$ be a r.v. 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 need 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 r.v $\underline{Y}$ 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 4.1.10.3.1 Let $\underline{X}$ and $\underline{Y}$ be p-dimensional r.v.s. Suppose that there exists an $h > 0$ such that

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

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

4.11. Order Statistics

Let $X_1 \ldots X_n$ be a random sample (of continuous r.v.s) from a distribution having d.f. F, p.d.f. f and support S [illegible].

Let Y_r = r-th smallest of $X_1, \ldots, X_n$, $r = 1, 2, \ldots, n$. The Y_r is called the r-th order statistic based on random sample $X_1 \ldots X_n$ and $Y_1 \ldots Y_n$ are called order statistics based on random sample $X_1, \ldots, X_n$. Note that if $X_1, \ldots X_n$ are continuous r.v.s then

$$\Pr(Y_1 < Y_2 < \cdots < Y_n) = 1$$

and thus $Y_1 \ldots Y_n$ are uniquely defined with probability one.

$\boxed{\text{Theorem 4.11.1.}}$ Under the above notation

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

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

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

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

$\boxed{\text{Proof.}}$ Since $\underline{X} = (X_1, \ldots, X_n)$ is a continuous r.v. (why?)

$$\Pr(Y_1 < Y_2 < \cdots < Y_n) = 1.$$

Define $S_n = S \times \cdots \times S$, so that support of $\underline{X} = (X_1 \ldots X_n)$ is S_n.
Define

$$S_1^\circ = \{ \underline{x} \in S_n : x_1 < x_2 < \cdots < x_n \}$$

$$S_2^\circ = \{ \underline{x} \in S_n : x_1 < x_2 < \cdots < x_n < x_{n-1} \}$$

$$\vdots$$

$$S_{\lfloor n}^\circ = \{ \underline{x} \in S_n : x_n < x_{n-1} < \cdots < x_1 \}.$$

On each S_i° $\underline{Y} = (Y_1 \ldots Y_n) = (h_{1,i}(\underline{X}), \ldots, h_{n,i}(\underline{X}))$

is 1-1 with inverse transformation $h_i^{-1} = (h_{1,i}^{-1}, \ldots, h_{n,i}^{-1})$,

$i = 1 \ldots \lfloor n$. Note that as a set

$$\{ h_{1,i}^{-1}(\underline{y}) \ldots, h_{n,i}^{-1}(\underline{y}) \} = \{ y_1 \ldots y_n \}, \quad i = 1, \ldots, \lfloor n$$

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

$$\underline{h}(S_i^\circ) = \{ \underline{y} \in S_n : y_1 < y_2 < \cdots < y_n \} = T, \text{ (say)}.$$

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

$$g(\underline{y}) = \sum_{j=1}^{\lfloor n} f_{\underline{X}}(h_{1j}^{-1}(\underline{y}) \ldots, h_{nj}^{-1}(\underline{y})) \, |J_j| \, I_{h(S_j)}(\underline{y})$$

$$= \sum_{j=1}^{\lfloor n} \left(\prod_{i=1}^{n} f(h_{ij}^{-1}(\underline{y})) \right) |J_j| \, I_T(\underline{y})$$

$$= \lfloor n \prod_{i=1}^{n} f(y_i), \qquad -\infty < y_1 < y_2 < \cdots < y_n < \infty$$

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

$$g_r(y) = \int_{-\infty}^{y} \int_{-\infty}^{y_{r-1}} \cdots \int_{-\infty}^{y_2} \int_{y}^{\infty} \int_{y_{r+1}}^{\infty} \cdots \int_{y_{n-1}}^{\infty} \lfloor n \, f(y_1) \cdots f(y_{r-1}) f(y) f(y_{r+1}) \cdots f(y_n) \, dy_n \cdots dy_{r+1} \, dy_1 \cdots dy_{r-1}$$

$$= \frac{\lfloor n}{\lfloor r-1 \, \lfloor n-r} [F(y)]^{r-1} [1-F(y)]^{n-r} f(y), \quad -\infty < y < \infty$$

Similarly, for $1 \le r < s \le n$, the joint pdf of (Y_r, Y_s) is

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

$$-\infty < y < z < \infty$$