Section 2.6. Extension to Several Random Variables

Note. We extend the ideas from this chapter concerning two random variables to more than two (though a finite number) of random variables.

Definition 2.6.1. Consider a random experiment with the sample space $\mathcal{C}$. Let the random variable X_i assign to each element $c \in \mathcal{C}$ one and only one real number $X_i(c) = x_i$ for $i = 1, 2, \ldots, n$. Then $(X_1, X_2, \ldots, X_n)$ is an n-dimensional *random vector*. The *space* (or *range*) of this random vector is the set of ordered n-tuples

$$\mathcal{D} = \{(x_1, x_2, \ldots, x_n) \mid x_1 = X_1(c), x_2 = X_2(c), \ldots, x_n = X_n(c) \text{ and } x \in \mathcal{C}\}.$$

Note/Definition. As in the case of a vector with two entries, we denote the transpose of of row vector $(X_1, X_2, \ldots, X_n)$ as the column vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)' = (X_1, X_2, \ldots, X_n)^T$ and we let $\mathbf{x} = (x_1, x_2, \ldots, x_n)' = (x_1, x_2, \ldots, x_n)^T$. As with two random variables, the joint cumulative density function is

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n).$$

In the discrete case we have

$$F_{\mathbf{X}}(\mathbf{x}) = \sum_{X_1 \le x_1} \sum_{X_2 \le x_2} \cdots \sum_{X_n \le x_n} p(w_1, w_2, \ldots, w_n)$$

and in the continuous case

$$F_{\mathbf{X}}(\mathbf{x}) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(w_1, w_2, \ldots, w_n)\, dw_n \cdots dw_2\, dw_1$$

where p is the probability mass function and f is the probability density functions. In the continuous case we have by Proposition IV.2.1 of my Complex Analysis notes on IV.2. Power Series Representations of Analytic Functions (which requires the integrand to have first partial derivatives) $\frac{\partial^n}{\partial x_1\,\partial x_2\cdots\partial x_n}[F_{\mathbf{X}}(\mathbf{x})] = f(\mathbf{x})$ "except possibly on points that have probability zero" (that is, almost everywhere).

Definition. For a discrete random vector, the *support* is all points in space $\mathcal{D}$ that have positive mass (i.e., positive probability). For a continuous random vector, the *support* is all points in $\mathcal{D}$ that can be embedded in an open set of positive probability. The support of a random vector is denoted $\mathcal{S}$.

Example 2.6.1. Let

$$f(x,y,z) = \begin{cases} e^{-(x+y+z)} & \text{for } 0 < x,y,z < \infty \\ 0 & \text{elsewhere} \end{cases}$$

be the probability density function of the random variables X, Y, and Z. Then the cumulative distribution function is

$$F(x,y,z) = P(Z \le x, Y \le y, Z \le z) = \int_{-\infty}^{x}\int_{-\infty}^{y}\int_{-\infty}^{z} f(u,v,w)\,du\,dv\,dw$$

$$= \int_0^x\int_0^y\int_0^z e^{-(u+v+w)}\,du\,dv\,dw = (1-e^{-x})(1-e^{-y})(1-e^{-z}) \text{ for } 0 \le x,y,z < \infty.$$

Notice $\frac{\partial^3}{\partial x\,\partial y\,\partial z}[F(z,y,z)] = f(z,y,z)$, as expected.

Definition. Let $(X_1, X_2, \ldots, X_n)$ be a random variable and let $Y = u(X_1, X_2, \ldots, X_n)$ for some measurable function u. If

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |u(x_1, x_2, \ldots, x_n)| f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n < \infty$$

for continuous random variables then the *expected value* of Y is

$$E[Y] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |u(x_1, x_2, \ldots, x_n)| f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n.$$

If

$$\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} |u(x_1, x_2, \ldots, x_n)| p(x_1, x_2, \ldots, x_n) < \infty$$

is finite for discrete random variables then the *expected value* of Y is

$$E[Y] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} y(x_1, x_2, \ldots, x_n) f(x_1, x_2, \ldots, x_n).$$

Note. As in the case of two random variables, the expectation operator is linear here as well.

Definition. For continuous random variables $X_1, X_2, \ldots, X_n$ with joint probability density function $f(x_1, x_2, \ldots, x_n)$, the *marginal probability density function* of X_i is

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n.$$

The marginal probability mass function of X_i is similarly defined for discrete random variables.

Note. With several random variables we can also define the marginal probability density function of some subset of the random variables by integrating out the other random variables.

Definition. For random variables $X_1, X_2, \ldots, X_n$ with joint probability density function $f(x_1, x_2, \ldots, x_n)$, with $f_i(x_i) > 0$ we have the *joint conditional probability density function* of $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n$ given $X_i = x_i$ is

$$f_{1,2,\ldots,i-1,i+1,\ldots,n|i}(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \mid x_i) = \frac{f(x_1, x_2, \ldots, x_n)}{f_i(x_i)}.$$

Note 2.6.A. We can similarly define a joint conditional probability density function given the values of several of the variables by dividing the joint probability density function $f(x_1, x_2, \ldots, x_n)$ by the marginal probability density function of the particular given group of variables. For example, if $f(x_1, x_2, \ldots, x_n)$ is the joint probability density function for X_1, X_2, X_3, X_4 then the marginal probability density function of X_1, X_2 is

$$f_{3,4}(x_3, x_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4)\, dx_1\, dx_2$$

and the joint conditional probability density function of X_1, X_2 given $X_3 = x_3$ and $X_4 = x_4$ is

$$f_{1,2|3,4}(x_1, x_2) = \frac{f(x_1, x_2, x_3, x_4)}{f_{3,4}(x_3, x_4)}.$$

Definition. Let $X_1, X_2, \ldots, X_n$ be continuous random variables and let $u(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots x_i)$ be a measurable function. The *conditional expectation* of $u(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ given $X_i = x_i$ is

$$E[u(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \mid x_i]$$
$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \times$$
$$f_{1,2,\ldots,i-1,i+1,\ldots,n}(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_n,$$

provided $f(x_i) > 0$ and the integral converges absolutely (i.e., the integral is absolutely integrable). We can similarly define the conditional expectation of $u(x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)$ given $X_i = x_i$ and $X_j = x_j$ (and we can also define related conditional expectations given values of several X_i's). These ideas can be similarly defined for discrete random variables by replacing the integrals with sums and the probability density functions replaced with probability mass functions.

Definition. Let the continuous random variables $X_1, X_2, \ldots, X_n$ have the joint probability density function $f(x_1, x_2, \ldots, x_n)$ and the marginal probability density functions $f_1(x_1), f_2(x_2), \ldots, x_n(x_n)$. The random variables are *mutually independent* if $f(x_1, x_2, \ldots, x_n) \equiv f_1(x_1)f_2(x_2) \cdots f_n(x_n)$. In the discrete case, the random variables are mutually independent if $p(x_1, x_2, \ldots, x_n) \equiv p_1(x_1)p_2(x_2) \cdots p_n(x_n)$.

Note 2.6.B. In Theorem 2.4.4, we saw that for two independent random variables, the expectation of the product is the product of expectations, $E[u(X_1)v(X_2)] =$

$E[u(x_1)]E[v(X_2)]$. Similarly, if random variables $X_1, X_2, \ldots, X_n$ are mutually independent and if $E[u_i(X_i)]$ exists for $i = 1, 2, \ldots, n$ then

$$E[u_1(X_1)u_2(X_2)\cdots u_n(X_n)] = E[u_1(X_1)]E[u_2(X_2)]\cdots E[u_n(X_n)]$$

or

$$E\left[\prod_{i=1}^{n} u_i(X_i)\right] = \prod_{i=1}^{n} E[u_i(X_i)].$$

Definition. Let $X_1, X_2, \ldots, X_n$ be random variables and suppose that $E[\exp(t_1X_1 + t_2X_2 + \cdots t_nX_n)]$ exists for $-h_i < t_i < h_i$ for some positive h_i where $i = 1, 2, \ldots, n$. Then this expectation is the *moment generating function* of the joint distribution of the random variables and is denoted:

$$M(t_1, t_2, \ldots, t_n) = E[\exp(t_1X_1 + t_2X_2 + \cdots + t_nX_n)] \text{ or } M_{\mathbf{X}}(\mathbf{t}) = E[\exp(\mathbf{t}'\mathbf{X})].$$

Note 2.6.C. As in the case of a moment generating function for the distribution of a single random variable (see Theorem 1.9.2), the moment generating function of the joint distribution of $X_1, X_2, \ldots, X_n$ uniquely determines the distribution of the random vector $(X_1, X_2, \ldots, X_n)$. Also, a moment generating function of the marginal distributions of X_i is $M(0, \ldots, 0, t_i, 0, \ldots, 0)$ for $i = 1, 2, \ldots, n$, the moment generating function of the marginal distribution of X_i and X_j is $M(0, \ldots, 0, t_i, 0, \ldots, 0, t_j, 0, \ldots, 0)$ and so forth, similar to the case of two random variables (see Note 2.1.D). In addition, we can generalize Theorem 2.4.5 (which states that the independence of two random variables X_1 and X_2 are independent if and only if the joint moment generating function $M(t_1, t_2)$ is identically equal to the product of the marginal moment generating functions $M(t_1, 0)M(0, t_2)$) using mathematical induction to show this equivalence between independence and expressing the joint moment generating function as the product of the marginal moment generating functions holds for any finite number of random variables. So if $X_1, X_2, \ldots, X_n$ are independent if and only if

$$M(t_1, t_2, \ldots, t_n) = \prod_{i=1}^{n} M(0, 0, \ldots, 0, t_i, 0, \ldots, 0).$$

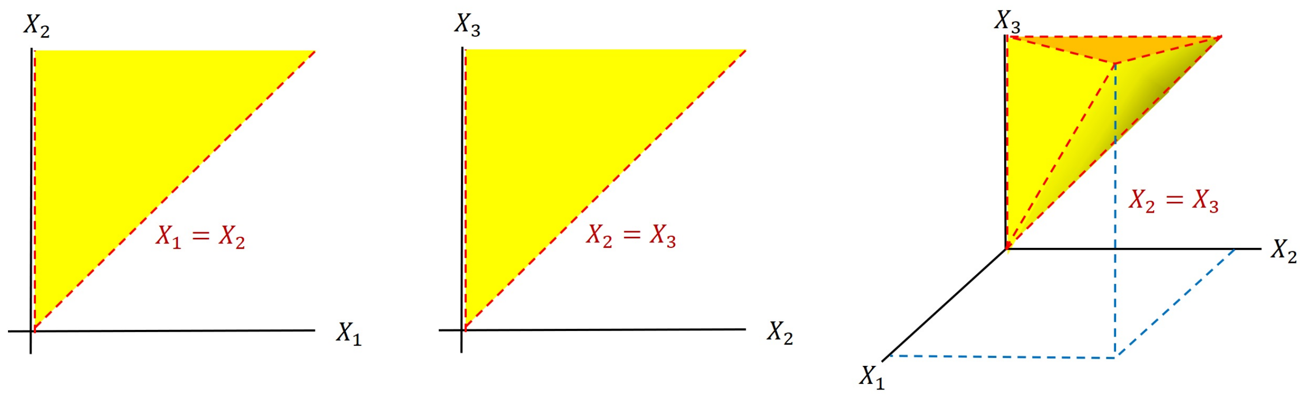
Note. We can find the moment generating function of a linear combination of random variables, as given in the following theorem.

Theorem 2.6.1. Suppose $X_1, X_2, \ldots, X_n$ are n mutually independent random variables. Suppose the Moment generating function for x_i is $M_i(t)$ for $-j_1 < t < h_i$ where $h_i > 0$, for $i = 1, 2, \ldots, n$. Let $T = \sum_{i=1}^{n} k_i X_i$ where $k_1, k_2, \ldots, k_n$ are constants. Then T has the moment generating function given by

$$M_T(i) = \prod_{i=1}^{n} M_i(k_i t) \text{ for } -\min_{1 \leq i \leq n}\{h_i\} \leq t \leq \min_{1 \leq i \leq n}\{h_i\}.$$

Exercise 2.6.2. Let X_1, X_2, X_3 be random variables with joint probability density function $f(x_1, x_2, x_3) = \exp(-(x_1 + x_2 + x_3))$ for $9 < x_1 < \infty$, $0 < x_2 < \infty$, $0 < x_3 < \infty$, and 0 elsewhere. **(a)** Compute $P(X_1 < X_2 < X_3)$ and $P(X_1 = X_2 < X_3)$.

Solution. For $X_1 < X_2 < X_3$, we consider the following sets in the coordinate planes and $\mathbb{R}^3$:



We have

$$P(X_1 < X_2 < X_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_3} \int_{-\infty}^{x_2} f(x_1, x_2, x_3) dx_1\, dx_2\, dx_3$$

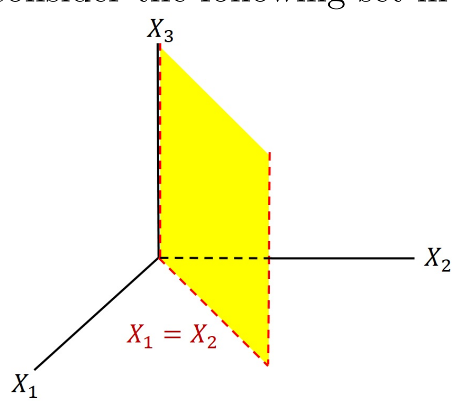
$$= \int_0^{\infty} \int_0^{x_3} \int_0^{x_2} e^{-(x_1+x_2+x_3)}\, dx_1\, dx_2\, dx_3 = \int_0^{\infty} \int_0^{x_3} -e^{-(x_1+x_2+x_3)} \Big|_{x_1=0}^{x_1=x_2} dx_2\, dx_3$$

$$= \int_0^{\infty} \int_0^{x_3} -e^{-(2x_2+x_3)} + e^{-(x_2+x_3)}\, dx_2\, dx_3 = \int_0^{\infty} \frac{1}{2} e^{-(2x_2+x_3)} - e^{-(x_2+x_3)} \Big|_{x_2=0}^{x_2=x_3} dx_3$$

$$= \int_0^{\infty} \frac{1}{2} e^{-3x_3} - e^{-2x_3} - \frac{1}{2} e^{-x_3} + e^{-x_3}\, dx_3 = \int_0^{\infty} \frac{1}{2} e^{-3x_3} - e^{-2x_3} + \frac{1}{2} e^{-x_3}\, dx_3$$

$$= \left(\frac{-1}{6} e^{-3x_3} + \frac{1}{2} e^{-2x_3} - \frac{1}{2} e^{-x_3} \right) \Big|_{x_3=0}^{x_3=\infty} = 0 - \left(\frac{-1}{6} + \frac{1}{2} - \frac{1}{2} \right) = \frac{1}{6}.$$

For $X_1 = X_2 < X_3$, we consider the following set in $\mathbb{R}^3$:



So we similarly have

$$P(X_1 = X_2 < X_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{x_3} \int_{x_2}^{x_2} f(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3$$

$$= \int_0^\infty \int_0^{x_3} \int_{x_2}^{x_2} e^{-(x_1+x_2+x_3)}\, dx_1\, dx_2\, dx_3 = \int_0^\infty \int_0^{x_3} 0\, dx_2\, dx_3 = 0.$$

(b) Determine the joint moment generating function of X_1, X_2, and X_3. Are these random variables independent?

Solution. The moment generating function of the joint distribution of X_1, X_2, X_3 is by definition

$$M(t_1, t_2, t_3) = E[\exp(t_1 X_1 + t_2 X_2 + t_2 X_3)]$$

$$\begin{aligned}
&= \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty e^{t_1x_1+t_2x_2+t_3x_3} e^{-(x_1+x_2+x_3)}\, dx_1\, dx_2\, dx_3 \\
&= \int_0^\infty \int_0^\infty \int_0^\infty e^{(t_1-1)x_1+(t_2-1)x_2+(t_3-1)x_3}\, dx_1\, dx_2\, dx_3 \\
&= \int_0^\infty \int_0^\infty \frac{1}{t_1-1} e^{(t_1-1)x_1+(t_2-1)x_2+(t_3-1)x_3} \Big|_{x_1=0}^{x_1=\infty} dx_2\, dx_3 \text{ if } t_1 \neq 1 \\
&= \int_0^\infty \int_0^\infty \frac{1}{1-t_1} e^{(t_2-1)x_2+(t_3-1)x_3}\, dx_2\, dx_3 \text{ if } t_1 < 1 \\
&= \int_0^\infty \frac{1}{(1-t_1)(t_2-1)} e^{(t_2-1)x_2+(t_3-1)x_3} \Big|_{x_2=0}^{x_2=\infty} dx_3 \text{ if } t_2 \neq 1 \\
&= \int_0^\infty \frac{1}{(t_1-1)(t_2-1)} e^{(t_3-1)x_3}\, dx_3 \text{ if } t_2 < 1 \\
&= \frac{1}{(1-t_1)(1-t_2)(t_3-1)} e^{(t_3-1)x_3} \Big|_{x_3=0}^{x_3=\infty} \text{ if } t_3 \neq 1 \\
&= \frac{1}{(1-t_1)(1-t_2)(1-t_3)} \text{ if } t_3 < 1
\end{aligned}$$

and notice that this holds (and is positive) for $t_1 < 1$, $t_2 < 1$, and $t_3 < 1$ (so that we can take $h_1 = h_2 = h_3 = 1$).

To check independence, we have the marginal probability density functions

$$f_1(x_1) = \int_{-\infty}^\infty \int_{-\infty}^\infty f(x_1, x_2, x_3)\, dx_2\, dx_3 = \int_0^\infty \int_0^\infty e^{-(x_1+x_2+x_3)}\, dx_2\, dx_3$$

$$= \int_0^\infty -e^{-(x_1+x_2+x_3)}\Big|_{x_2=0}^{x_2=\infty} dx_3 = \int_0^\infty 0 + e^{-(x_1+x_3)}\, dx_3$$

$$= -e^{-(x_1+x_2)}|_{x_3=0}^{x_3=\infty} = 0 + e^{-x_1} = e^{-x_1}$$

and similarly $f_2(x+2) = e^{-x_2}$, and $f_3(x_3) = e^{-x_3}$. Since

$$f(x_1, x_2, x_3) = e^{-(x_1+x_2+x_3)} = x^{-x_1}e^{-x_2}e^{-x_3} = f_1(x_1)f_2(x_2)f_3(x_3)$$

then be definition X_1, X_2, X_3 are mutually independent.

Note. A finite collection of random variables $X_1, X_2, \ldots, X_n$ can be "pairwise independent" (i.e., X_i and X_j are independent for any distinct i and j) but not mutually independent. In Remark 2.6.1, the text gives an example which they attribute to Sergei N. Bernstein (March 5, 1880–October 26, 1968). Let X_1, X_2, X_3 have the joint probability mass function

$$p(x_1, x_2, x_3) = \begin{cases} 1/4 & \text{if } (x_1, x_2, x_3) \in \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\} \\ 0 & \text{elsewhere.} \end{cases}$$

For $i \neq j$, the joint probability mass function of X_i and X_j (obtained from p by summing over X_k where $i \neq k \neq j$) is

$$p_{ij}(x_i, x_j) = \begin{cases} 1/4 & \text{if } (x_i, x_j) \in \{(0,0), (1,0), (0,1), (1,1)\} \\ 0 & \text{elsewhere,} \end{cases}$$

whereas the marginal probability mass function of X_i (obtained from p_{ij} be summing over j) is

$$p_i(x_i) = \begin{cases} 1/2 & \text{if } x_i \in \{0,1\} \\ 0 & \text{elsewhere.} \end{cases}$$

Since for $i \neq j$, $p_{ij}(x_i, x_j) \equiv p_i(x_i)p_j(x_j)$ (since $1/4 = (1/2)(1/2)$) then X_i and X_j are independent and so X_1, X_2, X_3 are pairwise independent. However, $p(x_1, x_2, x_3) \not\equiv p_1(x_1)p_2(x_2)p_3(x_3)$ since, for example, $p(1, 1, 1) = 1/4$ but $p_1(1)p_2(1)p_3(1) = 1/8$. So X_1, X_2, X_3 are not mutually independent.

Note. By convention, the text will usually simply stated that random variables are "independent" when they maean that the random variables are mutually independent (so the term takes on a stronger meaning than pairwise independent).

Definition. If random variables $X_1, X_2, \ldots, X_n$ are mutually independent and have each has the same probability density/mass distribution then the random variables are *independent and identically distributed* (abbreviated *idd*).

Note. The proof of the next result (a corollary to Theorem 2.6.1) is to be given in Exercise 2.6.8.

Corollary 2.6.1. Suppose $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with common moment generating function $M(t)$ for $-h < t < h$, where $h > 0$. Let $T = \sum_{i=1}^n X_i$. Then T has the moment generating function given by $M_T(t) = (M(t))^n$ for $-h < t < h$.

Note. The remainder of this section is declared optional by Hogg, McKean, Craig (since it requires matrix algebra!). We want to extend the idea of covariance for two random variables (defined in Section 2.5) to several random variables.

Definition. Let $X_1, X_2, \ldots, X_n$ be random variables and let $\mathbf{X} = (X_1, X_2, \ldots, X_n)' = (X_1, X_2, \ldots, X_n)^T$. We define the *expectation* of $\mathbf{X}$ as

$$E[\mathbf{X}] = (E(X_1), E(X_2), \ldots, E(X_n))' = (E(X_1), E(X_2), \ldots, E(X_n))^T.$$

For a set of random variables $\{X_{ij} \mid 1 \le i \le m, 1 \le j \le m\}$, define the $m \times n$ *random matrix* $\mathbf{W} = [W_{ij}]$ and the *expectation* of random matrix $\mathbf{W}$ as the $m \times n$ matrix $E[\mathbf{W}] = E(X_{ij})]$.

Theorem 2.6.2. Let $\mathbf{V}$ and $\mathbf{W}$ be $m \times n$ matrices of random variables, let $\mathbf{A}$ and $\mathbf{C}$ be $k \times m$ matrices of constants, and let $\mathbf{B}$ be an $n \times \ell$ matrix of constants. Then $E[\mathbf{AV} + \mathbf{CW}] = \mathbf{A}E[\mathbf{V}] + \mathbf{C}E[\mathbf{W}]$ and $E[\mathbf{AWB}] = \mathbf{A}E[\mathbf{E}]\mathbf{B}$; that is, E is a linear operator on matrices of random variables.

Definition. Let $X_1, X_2, \ldots, X_n$ be random variables, let $\mathbf{X} = (X_1, X_2, \ldots, X_n)' = (X_1, X_2, \ldots, X_n)^T$, and suppose $\sigma_i^2 = \text{Var}(_i) < \infty$ for $i = 1, 2, \ldots, n$. The *mean* of $\mathbf{X}$ is $\mu = E[\mathbf{X}]$. The *variance-covariance matrix* is

$$\text{Cov}(\mathbf{X}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = [\sigma_{ij}],$$

where σ_{ii} denotes σ_i^2.

Theorem 2.6.3. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)' = (X_1, X_2, \ldots, X_n)^T$ be an n-dimensional random vector, such that $\sigma_i^2 = \sigma_{ii} = \text{Var}(X_i) < \infty$. Let $\mathbf{A}$ be an $m \times n$ matrix of constants. Then $\text{Cov}(\mathbf{X}) = E[\mathbf{X}\mathbf{X}'] = \boldsymbol{\mu}\boldsymbol{\mu}'$ and $\text{Cov}(\mathbf{A}\mathbf{X}) = \mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}'$.

Definition. An $n \times n$ matrix $\mathbf{A}$ is *positive semi-definite* if for all (column) vectors $\mathbf{a} \in \mathbb{R}^n$ we have the scalar quantity $\mathbf{a}'\text{Cov}(\mathbf{X})\mathbf{a} = \mathbf{a}^T\text{Cov}(\mathbf{X})\mathbf{a} \geq 0$.

Corollary 2.6.A. All variance-covariance matrices are positive semi-definite.

Revised: 10/14/2022