Chapter 2

Multivariate Distributions and Transformations

2.1 Joint, Marginal and Conditional Distributions

Often there are n random variables $Y_1, ..., Y_n$ that are of interest. For example, *age, blood pressure, weight, gender* and *cholesterol level* might be some of the random variables of interest for patients suffering from heart disease.

Notation. Let $\Re^n$ be the n–dimensional Euclidean space. Then the vector $\boldsymbol{y} = (y_1, ..., y_n) \in \Re^n$ if y_i is an arbitrary real number for $i = 1, ..., n$.

Definition 2.1. If $Y_1, ..., Y_n$ are discrete random variables, then the **joint pmf** (probability mass function) of $Y_1, ..., Y_n$ is

$$f(y_1, ..., y_n) = P(Y_1 = y_1, ..., Y_n = y_n) \tag{2.1}$$

for any $(y_1, ..., y_n) \in \Re^n$. A joint pmf f satisfies $f(\boldsymbol{y}) \equiv f(y_1, ..., y_n) \geq 0$ $\forall \boldsymbol{y} \in \Re^n$ and

$$\sum \cdots \sum_{\boldsymbol{y}\,:\, f(\boldsymbol{y}) > 0} f(y_1, ..., y_n) = 1.$$

For any event $A \in \Re^n$,

$$P[(Y_1, ..., Y_n) \in A] = \sum \cdots \sum_{\boldsymbol{y}\,:\, \boldsymbol{y} \in A \text{ and } f(\boldsymbol{y}) > 0} f(y_1, ..., y_n).$$

Definition 2.2. The **joint cdf** (cumulative distribution function) of $Y_1, ..., Y_n$ is $F(y_1, ..., y_n) = P(Y_1 \leq y_1, ..., Y_n \leq y_n)$ for any $(y_1, ..., y_n) \in \Re^n$.

Definition 2.3. If $Y_1, ..., Y_n$ are continuous random variables, then the **joint pdf** (probability density function) of $Y_1, ..., Y_n$ is a function $f(y_1, ..., y_n)$ that satisfies $F(y_1, ..., y_n) = \int_{-\infty}^{y_n} \cdots \int_{-\infty}^{y_1} f(t_1, ..., t_n) dt_1 \cdots dt_n$ where the y_i are any real numbers. A joint pdf f satisfies $f(\boldsymbol{y}) \equiv f(y_1, ..., y_n) \geq 0$ $\forall \boldsymbol{y} \in \Re^n$ and $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, ..., t_n) dt_1 \cdots dt_n = 1$. For any event $A \in \Re^n$, $P[(Y_1, ..., Y_n) \in A] = \underset{A}{\int \cdots \int} f(t_1, ..., t_n) dt_1 \cdots dt_n$.

Definition 2.4. If $Y_1, ..., Y_n$ has a joint pdf or pmf f, then the *sample space* or **support** of $Y_1, ..., Y_n$ is

$$\mathcal{Y} = \{(y_1, ..., y_n) \in \Re^n : f(y_1, ..., y_n) > 0\}.$$

If $\boldsymbol{Y}$ comes from a family of distributions $f(\boldsymbol{y}|\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$, then the support $\mathcal{Y}_{\boldsymbol{\theta}} = \{\boldsymbol{y} : f(\boldsymbol{y}|\boldsymbol{\theta}) > 0\}$ may depend on $\boldsymbol{\theta}$.

Theorem 2.1. Let $Y_1, ..., Y_n$ have joint cdf $F(y_1, ..., y_n)$ and joint pdf $f(y_1, ..., y_n)$. Then

$$f(y_1, ..., y_n) = \frac{\partial^n}{\partial y_1 \cdots \partial y_n} F(y_1, ..., y_n)$$

wherever the partial derivative exists.

Definition 2.5. The **marginal pmf** of any subset $Y_{i1}, ..., Y_{ik}$ of the coordinates $(Y_1, ..., Y_n)$ is found by summing the joint pmf over all possible values of the other coordinates where the values $y_{i1}, ..., y_{ik}$ are held fixed. For example,

$$f_{Y_1,...,Y_k}(y_1, ..., y_k) = \sum_{y_{k+1}} \cdots \sum_{y_n} f(y_1, ..., y_n)$$

where $y_1, ..., y_k$ are held fixed. In particular, if Y_1 and Y_2 are discrete RVs with joint pmf $f(y_1, y_2)$, then the marginal pmf for Y_1 is

$$f_{Y_1}(y_1) = \sum_{y_2} f(y_1, y_2) \tag{2.2}$$

where y_1 is held fixed. The marginal pmf for Y_2 is

$$f_{Y_2}(y_2) = \sum_{y_1} f(y_1, y_2) \tag{2.3}$$

where y_2 is held fixed.

Example 2.1. For $n = 2$, double integrals are used to find marginal pdfs (defined below) and to show that the joint pdf integrates to 1. If the region of integration Ω is bounded on top by the function $y_2 = \phi_T(y_1)$, on the bottom by the function $y_2 = \phi_B(y_1)$ and to the left and right by the lines $y_1 = a$ and $y_2 = b$ then $\int \int_\Omega f(y_1, y_2)dy_1 dy_2 = \int \int_\Omega f(y_1, y_2)dy_2 dy_2 =$

$$\int_a^b \left[\int_{\phi_B(y_1)}^{\phi_T(y_1)} f(y_1, y_2)dy_2 \right] dy_1.$$

Within the inner integral, treat y_2 as the variable, anything else, including y_1, is treated as a constant.

If the region of integration Ω is bounded on the left by the function $y_1 = \psi_L(y_2)$, on the right by the function $y_1 = \psi_R(y_2)$ and to the top and bottom by the lines $y_2 = c$ and $y_2 = d$ then $\int \int_\Omega f(y_1, y_2)dy_1 dy_2 = \int \int_\Omega f(y_1, y_2)dy_2 dy_2 =$

$$\int_c^d \left[\int_{\psi_L(y_2)}^{\psi_R(y_2)} f(y_1, y_2)dy_1 \right] dy_2.$$

Within the inner integral, treat y_1 as the variable, anything else, including y_2, is treated as a constant.

Definition 2.6. The **marginal pdf** of any subset $Y_{i1}, ..., Y_{ik}$ of the coordinates $(Y_1, ..., Y_n)$ is found by integrating the joint pdf over all possible values of the other coordinates where the values $y_{i1}, ..., y_{ik}$ are held fixed. For example, $f(y_1, ..., y_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, ..., t_n) dt_{k+1} \cdots dt_n$ where $y_1, ..., y_k$ are held fixed. In particular, if Y_1 and Y_2 are continuous RVs with joint pdf $f(y_1, y_2)$, then the marginal pdf for Y_1 is

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_2 = \int_{\phi_B(y_1)}^{\phi_T(y_1)} f(y_1, y_2)dy_2 \tag{2.4}$$

where y_1 is held fixed (to get the region of integration, draw a line parallel to the y_2 axis and use the functions $y_2 = \phi_B(y_1)$ and $y_2 = \phi_T(y_1)$ as the lower and upper limits of integration). The marginal pdf for Y_2 is

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2)dy_1 = \int_{\psi_L(y_2)}^{\psi_R(y_2)} f(y_1, y_2)dy_1 \tag{2.5}$$

where y_2 is held fixed (to get the region of integration, draw a line parallel to the y_1 axis and use the functions $y_1 = \psi_L(y_2)$ and $y_1 = \psi_R(y_2)$ as the lower and upper limits of integration).

Definition 2.7. The **conditional pmf** of any subset $Y_{i1}, ..., Y_{ik}$ of the coordinates $(Y_1, ..., Y_n)$ is found by dividing the joint pmf by the marginal pmf of the remaining coordinates assuming that the values of the remaining coordinates are fixed and that the denominator > 0. For example,

$$f(y_1, ..., y_k | y_{k+1}, ..., y_n) = \frac{f(y_1, ..., y_n)}{f(y_{k+1}, ..., y_n)}$$

if $f(y_{k+1}, ..., y_n) > 0$. In particular, the conditional pmf of Y_1 given $Y_2 = y_2$ is a function of y_1 and

$$f_{Y_1|Y_2=y_2}(y_1|y_2) = \frac{f(y_1, y_2)}{f_{Y_2}(y_2)} \tag{2.6}$$

if $f_{Y_2}(y_2) > 0$, and the conditional pmf of Y_2 given $Y_1 = y_1$ is a function of y_2 and

$$f_{Y_2|Y_1=y_1}(y_2|y_1) = \frac{f(y_1, y_2)}{f_{Y_1}(y_1)} \tag{2.7}$$

if $f_{Y_1}(y_1) > 0$.

Definition 2.8. The **conditional pdf** of any subset $Y_{i1}, ..., Y_{ik}$ of the coordinates $(Y_1, ..., Y_n)$ is found by dividing the joint pdf by the marginal pdf of the remaining coordinates assuming that the values of the remaining coordinates are fixed and that the denominator > 0. For example,

$$f(y_1, ..., y_k | y_{k+1}, ..., y_n) = \frac{f(y_1, ..., y_n)}{f(y_{k+1}, ..., y_n)}$$

if $f(y_{k+1}, ..., y_n) > 0$. In particular, the conditional pdf of Y_1 given $Y_2 = y_2$ is a function of y_1 and

$$f_{Y_1|Y_2=y_2}(y_1|y_2) = \frac{f(y_1, y_2)}{f_{Y_2}(y_2)} \tag{2.8}$$

if $f_{Y_2}(y_2) > 0$, and the conditional pdf of Y_2 given $Y_1 = y_1$ is a function of y_2 and

$$f_{Y_2|Y_1=y_1}(y_2|y_1) = \frac{f(y_1, y_2)}{f_{Y_1}(y_1)} \tag{2.9}$$

if $f_{Y_1}(y_1) > 0$.

Example 2.2: Common Problem. If the joint pmf $f(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2)$ is given by a table, then the function $f(y_1, y_2)$ is a joint pmf if $f(y_1, y_2) \geq 0$, $\forall y_1, y_2$ and if

$$\sum_{(y_1,y_2):f(y_1,y_2)>0} f(y_1, y_2) = 1.$$

The marginal pmfs are found from the row sums and column sums using Definition 2.5, and the conditional pmfs are found with the formulas given in Definition 2.7.

Example 2.3: Common Problem. Given the joint pdf $f(y_1, y_2) = kg(y_1, y_2)$ on its support, find k, find the marginal pdfs $f_{Y_1}(y_1)$ and $f_{Y_2}(y_2)$ and find the conditional pdfs $f_{Y_1|Y_2=y_2}(y_1|y_2)$ and $f_{Y_2|Y_1=y_1}(y_2|y_1)$. Also, $P(a_1 < Y_1 < b_1, a_2 < Y_2 < b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(y_1, y_2) dy_1 dy_2$.

Tips: Often using **symmetry** helps.
The support of the marginal pdf does not depend on the 2nd variable.
The *support* of the conditional pdf can depend on the 2nd variable. For example, the support of $f_{Y_1|Y_2=y_2}(y_1|y_2)$ could have the form $0 \leq y_1 \leq y_2$.

The *support* of continuous random variables Y_1 and Y_2 is the region where $f(y_1, y_2) > 0$. The support is generally given by one to three inequalities such as $0 \leq y_1 \leq 1$, $0 \leq y_2 \leq 1$, and $0 \leq y_1 \leq y_2 \leq 1$. For each variable, set the inequalities to equalities to get boundary lines. For example $0 \leq y_1 \leq y_2 \leq 1$ yields 5 lines: $y_1 = 0, y_1 = 1$, $y_2 = 0, y_2 = 1$, and $y_2 = y_1$. Generally y_2 is on the vertical axis and y_1 is on the horizontal axis for pdfs.

To determine the **limits of integration**, examine the **dummy variable used in the inner integral**, say dy_1. Then within the region of integration, draw a line parallel to the same (y_1) axis as the dummy variable. The limits of integration will be functions of the other variable (y_2), never of the dummy variable (dy_1).

2.2 Expectation, Covariance and Independence

For joint pmfs with $n = 2$ random variables Y_1 and Y_2, the marginal pmfs and conditional pmfs can provide important information about the data. For joint pdfs the integrals are usually too difficult for the joint, conditional and marginal pdfs to be of practical use unless the random variables are independent. (Exceptions are the multivariate normal distribution and the elliptically contoured distributions. See Sections 2.9 and 2.10.)

For independent random variables, the joint cdf is the product of the marginal cdfs, the joint pmf is the product of the marginal pmfs, and the joint pdf is the product of the marginal pdfs. Recall that $\forall$ is read " for all."

Definition 2.9. i) The random variables $Y_1, Y_2, ..., Y_n$ are **independent** if $F(y_1, y_2, ..., y_n) = F_{Y_1}(y_1)F_{Y_2}(y_2) \cdots F_{Y_n}(y_n)$ $\forall y_1, y_2, ..., y_n$.
ii) If the random variables have a joint pdf or pmf f then the random variables $Y_1, Y_2, ..., Y_n$ are independent if $f(y_1, y_2, ..., y_n) = f_{Y_1}(y_1)f_{Y_2}(y_2) \cdots f_{Y_n}(y_n)$ $\forall y_1, y_2, ..., y_n$.
If the random variables are not independent, then they are **dependent**.
In particular random variables Y_1 and Y_2 are **independent**, written $Y_1 \perp\!\!\!\perp Y_2$, if either of the following conditions holds.
i) $F(y_1, y_2) = F_{Y_1}(y_1)F_{Y_2}(y_2) \quad \forall y_1, y_2$.
ii) $f(y_1, y_2) = f_{Y_1}(y_1)f_{Y_2}(y_2) \quad \forall y_1, y_2$.
Otherwise, Y_1 and Y_2 are *dependent*.

Definition 2.10. Recall that the support $\mathcal{Y}$ of $(Y_1, Y_2, ..., Y_n)$ is $\mathcal{Y} = \{\boldsymbol{y} : f(\boldsymbol{y}) > 0\}$. The support is a **cross product** or **Cartesian product** if

$$\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_n = \{\boldsymbol{y} : y_i \in \mathcal{Y}_i \text{ for } \text{i} = 1, ..., \text{n}\}$$

where $\mathcal{Y}_i$ is the support of Y_i. If f is a joint pdf then the support is **rectangular** if $\mathcal{Y}_i$ is an interval for each i. If f is a joint pmf then the support is rectangular if the points in $\mathcal{Y}_i$ are equally spaced for each i.

Example 2.4. In applications the support is often rectangular. For $n = 2$ the support is a cross product if

$$\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 = \{(y_1, y_2) : y_1 \in \mathcal{Y}_1 \text{ and } \text{y}_2 \in \mathcal{Y}_2\}$$

where $\mathcal{Y}_i$ is the support of Y_i. The support is rectangular if $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are intervals. For example, if

$$\mathcal{Y} = \{(y_1, y_2) : a < y_1 < \infty \text{ and } \text{c} \le \text{y}_2 \le \text{d}\},$$

then $\mathcal{Y}_1 = (a, \infty)$ and $\mathcal{Y}_2 = [c, d]$. For a joint pmf, the support is rectangular if the grid of points where $f(y_1, y_2) > 0$ is rectangular.

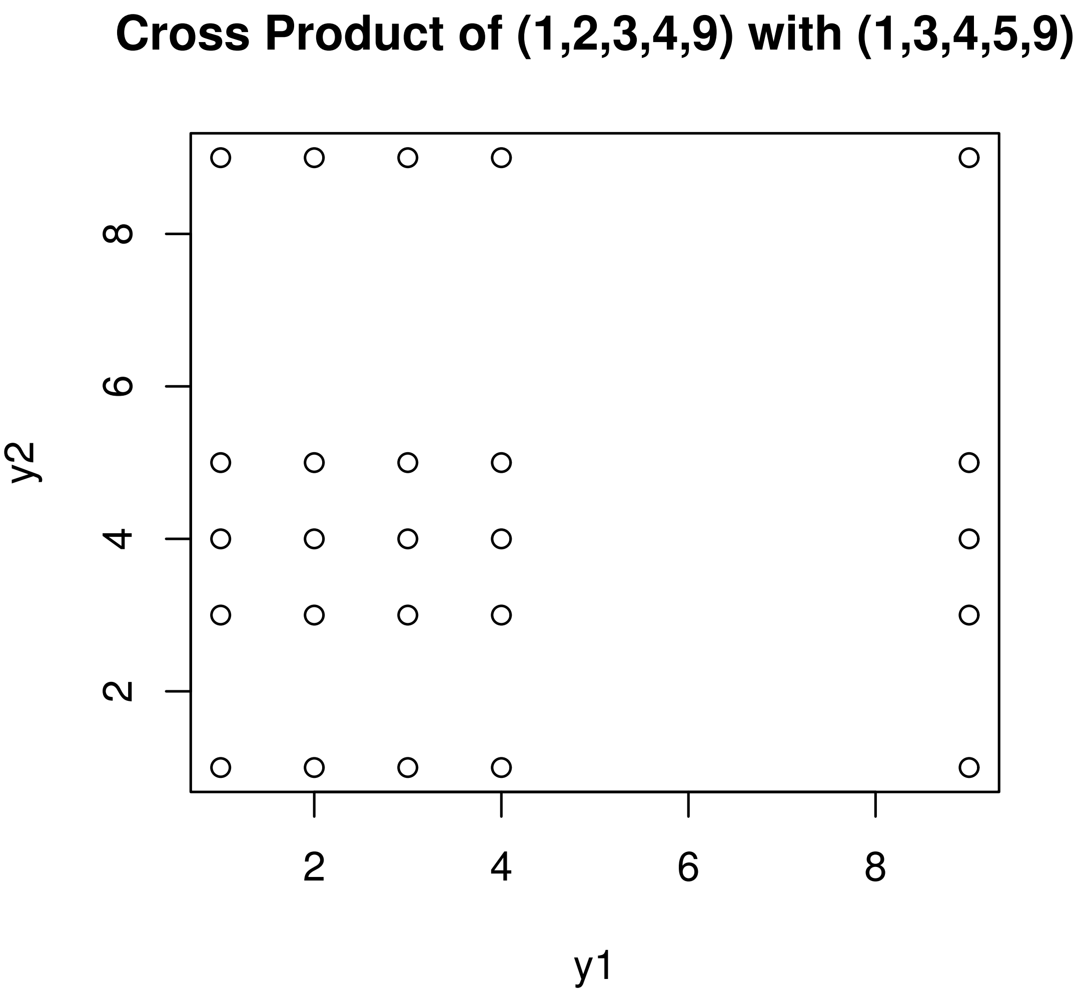

Figure 2.1: Cross Product for a Joint PMF

Figure 2.1 shows the cross product of $\mathcal{Y}_1 \times \mathcal{Y}_2$ where $\mathcal{Y}_1 = \{1, 2, 3, 4, 9\}$ and $\mathcal{Y}_2 = \{1, , 3, 4, 5, 9\}$. Each dot occurs where $f(y_1, y_2) > 0$. Notice that each point in $\mathcal{Y}_1$ occurs with each point in $\mathcal{Y}_2$. This support would not be a cross product if any point was deleted, but would be a cross product if any row of dots or column of dots was deleted.

Theorem 2.2a below is useful because it is often immediate from the formula for the joint pdf or the table for the joint pmf that the support is not a cross product. Hence Y_1 and Y_2 are dependent. For example, if the support of Y_1 and Y_2 is a triangle, then Y_1 and Y_2 are dependent. **A necessary condition for independence is that the support is a cross product.** Theorem 2.2b is useful because factorizing the joint pdf on cross product support is easier than using integration to find the marginal pdfs. Many texts give Theorem 2.2c, but 2.2b is easier to use. Recall that that $\prod_{i=1}^n a_i = a_1 a_2 \cdots a_n$. For example, let $n = 3$ and $a_i = i$ for $i = 1, 2, 3$. Then $\prod_{i=1}^n a_i = a_1 a_2 a_3 = (1)(2)(3) = 6$.

Theorem 2.2. a) Random variables $Y_1, ..., Y_n$ with joint pdf or pmf f are dependent if their support $\mathcal{Y}$ is not a cross product. In particular, Y_1 and Y_2 are dependent if $\mathcal{Y}$ does not have the form $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$.

b) If random variables $Y_1, ..., Y_n$ with joint pdf or pmf f have support $\mathcal{Y}$ that is a cross product, then $Y_1, ..., Y_n$ are independent iff $f(y_1, y_2, ..., y_n) = h_1(y_1) h_2(y_2) \cdots h_n(y_n)$ for all $\boldsymbol{y} \in \mathcal{Y}$ where h_i is a positive function of y_i alone. In particular, if $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$, then $Y_1 \perp\!\!\!\perp Y_2$ iff $f(y_1, y_2) = h_1(y_1) h_2(y_2)$ for all $(y_1, y_2) \in \mathcal{Y}$ where $h_i(y_i) > 0$ for $y_i \in \mathcal{Y}_i$ and $i = 1, 2$.

c) $Y_1, ..., Y_n$ are independent iff $f(y_1, y_2, ..., y_n) = g_1(y_1) g_2(y_2) \cdots g_n(y_n)$ for all $\boldsymbol{y}$ where g_i is a nonnegative function of y_i alone.

d) If discrete Y_1 and Y_2 have cross product support given by a table, find the row and column sums. If $f(y_1, y_2) \neq f_{Y_1}(y_1) f_{Y_2}(y_2)$ for **some entry** (y_1, y_2), then Y_1 and Y_2 are dependent. If $f(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$ for *all table entries, then* Y_1 *and* Y_2 *are independent.*

Proof. a) If the support is not a cross product, then there is a point $\boldsymbol{y}$ such that $f(\boldsymbol{y}) = 0$ but $f_{Y_i}(y_i) > 0$ for $i = 1, ..., n$. Hence $f(\boldsymbol{y}) \neq \prod_{i=1}^n f_{Y_i}(y_i)$ at the point $\boldsymbol{y}$ and $Y_1, ..., Y_n$ are dependent.

b) The proof for a joint pdf is given below. For a joint pmf, replace the integrals by appropriate sums. If $Y_1, ..., Y_n$ are independent, take $h_i(y_i) = f_{Y_i}(y_i) > 0$ for $y_i \in \mathcal{Y}_i$ and $i = 1, ..., n$.

If $f(\boldsymbol{y}) = h_1(y_1) \cdots h_n(y_n)$ for $\boldsymbol{y} \in \mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n$ then $f(\boldsymbol{y}) = 0 = f_{Y_1}(y_1) \cdots f_{Y_n}(y_n)$ if $\boldsymbol{y}$ is not in $\mathcal{Y}$. Hence we need to show that $f(\boldsymbol{y}) = f_{Y_1}(y_1) \cdots f_{Y_n}(y_n) = h_1(y_1) \cdots h_n(y_n)$ if $\boldsymbol{y} \in \mathcal{Y}$. Since f is a joint pdf,

$$1 = \int \cdots \int_{\mathcal{Y}} f(\boldsymbol{y}) \ d\boldsymbol{y} = \prod_{i=1}^{n} \int_{\mathcal{Y}_i} h_i(y_i) \ dy_i = \prod_{i=1}^{n} a_i$$

where $a_i = \int_{\mathcal{Y}_i} h_i(y_i) \ dy_i > 0$. For $y_i \in \mathcal{Y}_i$, the marginal pdfs $f_{Y_i}(y_i) =$

$$\int_{\mathcal{Y}_n} \cdots \int_{\mathcal{Y}_{i+1}} \int_{\mathcal{Y}_{i-1}} \cdots \int_{\mathcal{Y}_1} h_1(y_1) \cdots h_i(y_i) \cdots h(y_n) \ dy_1 \cdots dy_{i-1} dy_{i+1} \cdots dy_n$$

$$= h_i(y_i) \prod_{j=1, j \neq i}^{n} \int_{\mathcal{Y}_j} h_j(y_j) \ dy_j = h_i(y_i) \prod_{j=1, j \neq i}^{n} a_j = h_i(y_i) \frac{1}{a_i}.$$

Thus $a_i f_{Y_i}(y_i) = h_i(y_i)$ for $y_i \in \mathcal{Y}_i$. Since $\prod_{i=1}^n a_i = 1$,

$$f(\boldsymbol{y}) = \prod_{i=1}^{n} h_i(y_i) = \prod_{i=1}^{n} a_i f_{Y_i}(y_i) = (\prod_{i=1}^{n} a_i)(\prod_{i=1}^{n} f_{Y_i}(y_i)) = \prod_{i=1}^{n} f_{Y_i}(y_i)$$

if $\boldsymbol{y} \in \mathcal{Y}$.

c) Take

$$g_i(y_i) = \begin{cases} h_i(y_i), & \text{if } y_i \in \mathcal{Y}_i \\ 0, & \text{otherwise.} \end{cases}$$

Then the result follows from b).

d) Since $f(y_1, y_2) = 0 = f_{Y_1}(y_1) f_{Y_2}(y_2)$ if (y_1, y_2) is not in the support of Y_1 and Y_2, the result follows by the definition of independent random variables. QED

The following theorem shows that finding the marginal and conditional pdfs or pmfs is simple if $Y_1, ..., Y_n$ are independent. Also **subsets of independent random variables are independent**: if $Y_1, ..., Y_n$ are independent and if $\{i_1, ..., i_k\} \subseteq \{1, ..., n\}$ for $k \geq 2$, then $Y_{i_1}, ..., Y_{i_k}$ are independent.

Theorem 2.3. Suppose that $Y_1, ..., Y_n$ are independent random variables with joint pdf or pmf $f(y_1, ..., y_n)$. Then

a) the marginal pdf or pmf of any subset $Y_{i_1}, ..., Y_{i_k}$ is $f(y_{i_1}, ..., y_{i_k}) = \prod_{j=1}^k f_{Y_{i_j}}(y_{i_j})$. Hence $Y_{i_1}, ..., Y_{i_k}$ are independent random variables for $k \geq 2$.

b) The conditional pdf or pmf of $Y_{i_1}, ..., Y_{i_k}$ given any subset of the remaining random variables $Y_{j_1} = y_{j_1}, ..., Y_{j_m} = y_{j_m}$ is equal to the marginal: $f(y_{i_1}, ..., y_{i_k} | y_{j_1}, ..., y_{j_m}) = f(y_{i_1}, ..., y_{i_k}) = \prod_{j=1}^k f_{Y_{i_j}}(y_{i_j})$ if $f(y_{j_1}, ..., y_{j_m}) > 0$.

Proof. The proof for a joint pdf is given below. For a joint pmf, replace the integrals by appropriate sums. a) The marginal

$$f(y_{i_1}, ..., y_{i_k}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{j=1}^{n} f_{Y_{i_j}}(y_{i_j}) \right] \, dy_{i_{k+1}} \cdots dy_{i_n}$$

$$= \left[\prod_{j=1}^{k} f_{Y_{i_j}}(y_{i_j}) \right] \left[\prod_{j=k+1}^{n} \int_{-\infty}^{\infty} f_{Y_{i_j}}(y_{i_j}) \, dy_{i_j} \right]$$

$$= \left[\prod_{j=1}^{k} f_{Y_{i_j}}(y_{i_j}) \right] (1)^{n-k} = \prod_{j=1}^{k} f_{Y_{i_j}}(y_{i_j}).$$

b) follows from a) and the definition of a conditional pdf assuming that $f(y_{j_1}, ..., y_{j_m}) > 0$. QED

Definition 2.11. Suppose that random variables $\boldsymbol{Y} = (Y_1, ..., Y_n)$ have support $\mathcal{Y}$ and joint pdf or pmf f. Then the **expected value** of $h(\boldsymbol{Y}) = h(Y_1, ..., Y_n)$ is

$$E[h(\boldsymbol{Y})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(\boldsymbol{y}) f(\boldsymbol{y}) \, d\boldsymbol{y} = \int \cdots \int_{\mathcal{Y}} h(\boldsymbol{y}) f(\boldsymbol{y}) \, d\boldsymbol{y} \qquad (2.10)$$

if f is a joint pdf and if

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |h(\boldsymbol{y})| f(\boldsymbol{y}) \, d\boldsymbol{y}$$

exists. Otherwise the expectation does not exist. The expected value is

$$E[h(\boldsymbol{Y})] = \sum_{y_1} \cdots \sum_{y_n} h(\boldsymbol{y}) f(\boldsymbol{y}) = \sum_{\boldsymbol{y} \in \Re^n} h(\boldsymbol{y}) f(\boldsymbol{y}) = \sum_{\boldsymbol{y} \in \mathcal{Y}} h(\boldsymbol{y}) f(\boldsymbol{y}) \quad (2.11)$$

if f is a joint pmf and if $\sum_{\boldsymbol{y} \in \Re^n} |h(\boldsymbol{y})| f(\boldsymbol{y})$ exists. Otherwise the expectation does not exist.

The following theorem is useful since multiple integrals with smaller dimension are easier to compute than those with higher dimension.

Theorem 2.4. Suppose that $Y_1, ..., Y_n$ are random variables with joint pdf or pmf $f(y_1, ..., y_n)$. Let $\{i_1, ..., i_k\} \subset \{1, ..., n\}$, and let $f(y_{i_1}, ..., y_{i_k})$ be the marginal pdf or pmf of $Y_{i_1}, ..., Y_{i_k}$ with support $\mathcal{Y}_{Y_{i_1}, ..., Y_{i_k}}$. Assume that $E[h(Y_{i_1}, ..., Y_{i_k})]$ exists. Then

$$E[h(Y_{i_1}, ..., Y_{i_k})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_{i_1}, ..., y_{i_k}) \ f(y_{i_1}, ..., y_{i_k}) \ dy_{i_1} \cdots dy_{i_k} =$$

$$\int \cdots \int_{\mathcal{Y}_{Y_{i_1}, ..., Y_{i_k}}} h(y_{i_1}, ..., y_{i_k}) \ f(y_{i_1}, ..., y_{i_k}) \ dy_{i_1} \cdots dy_{i_k}$$

if f is a pdf, and

$$E[h(Y_{i_1}, ..., Y_{i_k})] = \sum_{y_{i_1}} \cdots \sum_{y_{i_k}} h(y_{i_1}, ..., y_{i_k}) \ f(y_{i_1}, ..., y_{i_k})$$

$$= \sum_{(y_{i_1}, ..., y_{i_k}) \in \mathcal{Y}_{Y_{i_1}, ..., Y_{i_k}}} h(y_{i_1}, ..., y_{i_k}) \ f(y_{i_1}, ..., y_{i_k})$$

if f is a pmf.

Proof. The proof for a joint pdf is given below. For a joint pmf, replace the integrals by appropriate sums. Let $g(Y_1, ..., Y_n) = h(Y_{i_1}, ..., Y_{i_k})$. Then $E[g(\boldsymbol{Y})] =$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_{i_1}, ..., y_{i_k}) f(y_1, ..., y_n) \ dy_1 \cdots dy_n =$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_{i_1}, ..., y_{i_k}) \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1, ..., y_n) \ dy_{i_{k+1}} \cdots dy_{i_n} \right] \ dy_{i_1} \cdots dy_{i_k}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_{i_1}, ..., y_{i_k}) f(y_{i_1}, ..., y_{i_k}) \ dy_{i_1} \cdots dy_{i_k}$$

since the term in the brackets gives the marginal. QED

Example 2.5. Typically $E(Y_i), E(Y_i^2)$ and $E(Y_iY_j)$ for $i \neq j$ are of primary interest. Suppose that (Y_1, Y_2) has joint pdf $f(y_1, y_2)$. Then $E[h(Y_1, Y_2)]$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(y_1, y_2) f(y_1, y_2) dy_2 dy_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(y_1, y_2) f(y_1, y_2) dy_1 dy_2$$

where $-\infty$ to ∞ could be replaced by the limits of integration for dy_i. **In particular**,

$$E(Y_1Y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f(y_1, y_2) dy_2 dy_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 y_2 f(y_1, y_2) dy_1 dy_2.$$

Since finding the marginal pdf is usually easier than doing the double integral, if h is a function of Y_i but not of Y_j, find the marginal for Y_i : $E[h(Y_1)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(y_1) f(y_1, y_2) dy_2 dy_1 = \int_{-\infty}^{\infty} h(y_1) f_{Y_1}(y_1) dy_1$. Similarly, $E[h(Y_2)] = \int_{-\infty}^{\infty} h(y_2) f_{Y_2}(y_2) dy_2$.

In particular, $E(Y_1) = \int_{-\infty}^{\infty} y_1 f_{Y_1}(y_1) dy_1$, and $E(Y_2) = \int_{-\infty}^{\infty} y_2 f_{Y_2}(y_2) dy_2$.

Suppose that (Y_1, Y_2) have a joint pmf $f(y_1, y_2)$. Then the **expectation** $E[h(Y_1, Y_2)] = \sum_{y_2} \sum_{y_1} h(y_1, y_2) f(y_1, y_2) = \sum_{y_1} \sum_{y_2} h(y_1, y_2) f(y_1, y_2)$. **In particular**,

$$E[Y_1Y_2] = \sum_{y_1} \sum_{y_2} y_1 y_2 f(y_1, y_2).$$

Since finding the marginal pmf is usually easier than doing the double summation, if h is a function of Y_i but not of Y_j, find the marginal for pmf for Y_i: $E[h(Y_1)] = \sum_{y_2} \sum_{y_1} h(y_1) f(y_1, y_2) = \sum_{y_1} h(y_1) f_{Y_1}(y_1)$. Similarly, $E[h(Y_2)] = \sum_{y_2} h(y_2) f_{Y_2}(y_2)$. **In particular**, $E(Y_1) = \sum_{y_1} y_1 f_{Y_1}(y_1)$ and $E(Y_2) = \sum_{y_2} y_2 f_{Y_2}(y_2)$.

For pdfs it is sometimes possible to find $E[h(Y_i)]$ but for $k \geq 2$ these expected values tend to be too difficult to compute unless the problem is impractical. Independence makes finding some expected values simple.

Theorem 2.5. Let $Y_1, ..., Y_n$ be independent random variables. If $h_i(Y_i)$ is a function of Y_i alone and if the relevant expected values exist, then

$$E[h_1(Y_1) h_2(Y_2) \cdots h_n(Y_n)] = E[h_1(Y_1)] \cdots E[h_n(Y_n)].$$

In particular, $E[Y_iY_j] = E[Y_i]E[Y_j]$ for $i \neq j$.

Proof. The result will be shown for the case where $\boldsymbol{Y} = (Y_1, ..., Y_n)$ has a joint pdf f. For a joint pmf, replace the integrals by appropriate sums. By independence, the support of $\boldsymbol{Y}$ is a cross product: $\mathcal{Y} = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_n$. Since $f(\boldsymbol{y}) = \prod_{i=1}^n f_{Y_i}(y_i)$, the expectation $E[h_1(Y_1)h_2(Y_2)\cdots h_n(Y_n)] =$

$$\int \cdots \int_{\mathcal{Y}} h_1(y_1)h_2(y_2)\cdots h_n(y_n)\ f(y_1, ..., y_n)\ dy_1 \cdots dy_n$$

$$= \int_{\mathcal{Y}_n} \cdots \int_{\mathcal{Y}_1} \left[\prod_{i=1}^n h_i(y_i) f_{Y_i}(y_i)\right] dy_1 \cdots dy_n$$

$$= \prod_{i=1}^n \left[\int_{\mathcal{Y}_i} h_i(y_i) f_{Y_i}(y_i)\ dy_i\right] = \prod_{i=1}^n E[h_i(Y_i)].\ \text{QED}$$

Corollary 2.6. Let $Y_1, ..., Y_n$ be independent random variables. If $h_j(Y_{i_j})$ is a function of Y_{i_j} alone and if the relevant expected values exist, then

$$E[h_1(Y_{i_1})\cdots h_k(Y_{i_k})] = E[h_1(Y_{i_1})]\cdots E[h_k(Y_{i_k})].$$

Proof. Method 1: Take $X_j = Y_{i_j}$ for $j = 1, ..., k$. Then $X_1, ..., X_k$ are independent and Theorem 2.5 applies.

Method 2: Take $h_j(Y_{i_j}) \equiv 1$ for $j = k+1, ..., n$ and apply Theorem 2.5. QED

Theorem 2.7. Let $Y_1, ..., Y_n$ be independent random variables. If $h_i(Y_i)$ is a function of Y_i alone and $X_i = h_i(Y_i)$, then the random variables $X_1, ..., X_n$ are independent.

Definition 2.12. The **covariance** of Y_1 and Y_2 is

$$\text{Cov}(Y_1, Y_2) = E[(Y_1 - E(Y_1))(Y_2 - E(Y_2))]$$

provided the expectation exists. Otherwise the covariance does not exist.

Theorem 2.8: Short cut formula. If $\text{Cov}(Y_1, Y_2)$ exists then $\text{Cov}(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$.

Theorem 2.9. Let Y_1 and Y_2 be independent random variables.
a) If $\text{Cov}(Y_1, Y_2)$ exists, then $\text{Cov}(Y_1, Y_2) = 0$.

b) **The converse is false**: $\text{Cov}(Y_1, Y_2) = 0$ does not imply $Y_1 \perp\!\!\!\perp Y_2$.

Example 2.6. When $f(y_1, y_2)$ is given by a table, a common problem is to determine whether Y_1 and Y_2 are independent or dependent, find the marginal pmfs $f_{Y_1}(y_1)$ and $f_{Y_2}(y_2)$ and find the conditional pmfs $f_{Y_1|Y_2=y_2}(y_1|y_2)$ and $f_{Y_2|Y_1=y_1}(y_2|y_1)$. Also find $E(Y_1), E(Y_2), V(Y_1), V(Y_2), E(Y_1Y_2)$ and $\text{Cov}(Y_1, Y_2)$.

Example 2.7. Given the joint pdf $f(y_1, y_2) = kg(y_1, y_2)$ on its support, a common problem is to find k, find the marginal pdfs $f_{Y_1}(y_1)$ and $f_{Y_2}(y_2)$ and find the conditional pdfs $f_{Y_1|Y_2=y_2}(y_1|y_2)$ and $f_{Y_2|Y_1=y_1}(y_2|y_1)$. Also determine whether Y_1 and Y_2 are independent or dependent, and find $E(Y_1), E(Y_2), V(Y_1), V(Y_2), E(Y_1Y_2)$ and $\text{Cov}(Y_1, Y_2)$.

Example 2.8. Suppose that the joint probability mass function of Y_1 and Y_2 is $f(y_1, y_2)$ is tabled as shown.

$f(y_1, y_2)$		y_2 = 0	1	2
	0	1/9	2/9	1/9
y_1	1	2/9	2/9	0/9
	2	1/9	0/9	0/9

a) Are Y_1 and Y_2 independent? Explain.
b) Find the marginal pmfs.
c) Find $E(Y_1)$.
d) Find $E(Y_2)$.
e) Find $\text{Cov}(Y_1, Y_2)$.

Solution: a) No, the support is not a cross product. Alternatively, $f(2, 2) = 0 < f_{Y_1}(2)f_{Y_2}(2)$.

b) Find $f_{Y_1}(y_1)$ by finding the row sums. Find $f_{Y_2}(y_2)$ by finding the column sums. In both cases, $f_{Y_i}(0) = f_{Y_i}(1) = 4/9$ and $f_{Y_i}(2) = 1/9$.

c) $E(Y_1) = \sum y_1 f_{Y_1}(y_1) = 0\frac{4}{9} + 1\frac{4}{9} + 2\frac{1}{9} = \frac{6}{9} \approx 0.6667$.

d) $E(Y_2) \approx 0.6667$ is found as in c) with y_2 replacing y_1.

e) $E(Y_1Y_2) = \sum\sum y_1y_2f(y_1, y_2) =$
$0 + 0 + 0$
$+0 + (1)(1)\frac{2}{9} + 0$
$+0 + 0 + 0 = \frac{2}{9}$. Hence $\text{Cov}(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2) = \frac{2}{9} - (\frac{6}{9})(\frac{6}{9}) = -\frac{2}{9} \approx -0.2222$.

Example 2.9. Suppose that the joint pdf of the random variables Y_1

and Y_2 is given by

$$f(y_1, y_2) = 10 y_1 y_2^2, \quad \text{if } 0 < y_1 < y_2 < 1$$

and $f(y_1, y_2) = 0$, otherwise. a) Find the marginal pdf of Y_1. Include the support. b) Is $Y_1 \perp\!\!\!\perp Y_2$?

Solution: a)Notice that for a given value of y_1, the joint pdf is positive for $y_1 < y_2 < 1$. Thus

$$f_{Y_1}(y_1) = \int_{y_1}^{1} 10 y_1 y_2^2 dy_2 = 10 y_1 \frac{y_2^3}{3}\bigg|_{y_1}^{1} = \frac{10 y_1}{3}(1 - y_1^3), 0 < y_1 < 1.$$

b) No, the support is not a cross product.

Example 2.10. Suppose that the joint pdf of the random variables Y_1 and Y_2 is given by

$$f(y_1, y_2) = 4 y_1 (1 - y_2), \text{ if } 0 \le y_1 \le 1, 0 \le y_2 \le 1$$

and $f(y_1, y_2) = 0$, otherwise.
a) Find the marginal pdf of Y_1. Include the support.
b) Find $E(Y_1)$.
c) Find $V(Y_1)$.
d) Are Y_1 and Y_2 independent? Explain.

Solution: a) $f_{Y_1}(y_1) = \int_0^1 4 y_1 (1 - y_2) dy_2 = 4 y_1 (y_2 - \frac{y_2^2}{2})\Big|_0^1 = 4 y_1 (1 - \frac{1}{2}) = 2 y_1, 0 < y_1 < 1$.

b) $E(Y_1) = \int_0^1 y_1 f_{Y_1}(y_1) dy_1 = \int_0^1 y_1 2 y_1 dy_1 = 2 \int_0^1 y_1^2 dy_1 = 2 \frac{y_1^3}{3}\Big|_0^1 = 2/3$.

c) $E(Y_1^2) = \int_0^1 y_1^2 f_{Y_1}(y_1) dy_1 = \int_0^1 y_1^2 2 y_1 dy_1 = 2 \int_0^1 y_1^3 dy_1 = 2 \frac{y_1^4}{4}\Big|_0^1 = 1/2$. So $V(Y_1) = E(Y_1^2) - [E(Y_1)]^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18} \approx 0.0556$.

d) Yes, use Theorem 2.2b with $f(y_1, y_2) = (4 y_1)(1 - y_2) = h_1(y_1) h_2(y_2)$ on cross product support.

2.3 Conditional Expectation and Variance

Notation: $Y|X = x$ is a single conditional distribution while $Y|X$ is a family of distributions. For example, if $Y|X = x \sim N(c + dx, \sigma^2)$, then

$Y|X \sim N(c + dX, \sigma^2)$ is the family of normal distributions with variance σ^2 and mean $\mu_{Y|X=x} = c + dx$.

Definition 2.13. Suppose that $f(y|x)$ is the conditional pmf or pdf of $Y|X = x$ and that $h(Y)$ is a function of Y. Then the *conditional expected value* $E[h(Y)|X = x]$ of $h(Y)$ given $X = x$ is

$$E[h(Y)|X = x] = \sum_y h(y)f(y|x) \tag{2.12}$$

if $f(y|x)$ is a pmf and if the sum exists when $h(y)$ is replaced by $|h(y)|$. In particular,

$$E[Y|X = x] = \sum_y y f(y|x). \tag{2.13}$$

Similarly,

$$E[h(Y)|X = x] = \int_{-\infty}^{\infty} h(y)f(y|x)dy \tag{2.14}$$

if $f(y|x)$ is a pdf and if the integral exists when $h(y)$ is replaced by $|h(y)|$. In particular,

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f(y|x)dy. \tag{2.15}$$

Definition 2.14. Suppose that $f(y|x)$ is the conditional pmf or pdf of $Y|X = x$. Then the *conditional variance*

$$\text{VAR}(Y|X = x) = E(Y^2|X = x) - [E(Y|X = x)]^2$$

whenever $E(Y^2|X = x)$ exists.

Recall that $f(y|x)$ is a function of y with x fixed, but $E(Y|X = x) \equiv m(x)$ is a function of x. In the definition below, both $E(Y|X)$ and $\text{VAR}(Y|X)$ are random variables since $m(X)$ and $v(X)$ are random variables.

Definition 2.15. If $E(Y|X = x) = m(x)$, then $E(Y|X) = m(X)$. Similarly if $\text{VAR}(Y|X = x) = v(x)$, then $\text{VAR}(Y|X) = v(X) = E(Y^2|X) - [E(Y|X)]^2$.

Example 2.11. Suppose that Y = *weight* and X = *height* of college students. Then $E(Y|X = x)$ is a function of x. For example, the weight of 5 feet tall students is less than the weight of 6 feet tall students, on average.

Notation: When computing $E(h(Y))$, the marginal pdf or pmf $f(y)$ is used. When computing $E[h(Y)|X = x]$, the conditional pdf or pmf $f(y|x)$ is used. In a formula such as $E[E(Y|X)]$ the inner expectation uses $f(y|x)$ but the outer expectation uses $f(x)$ since $E(Y|X)$ is a function of X. In the formula below, we could write $E_Y(Y) = E_X[E_{Y|X}(Y|X)]$, but such notation is usually omitted.

Theorem 2.10: Iterated Expectations. Assume the relevant expected values exist. Then

$$E(Y) = E[E(Y|X)].$$

Proof: The result will be shown for the case where (Y, X) has a joint pmf f. For a joint pdf, replace the sums by appropriate integrals. Now

$$E(Y) = \sum_x \sum_y y f(x, y) = \sum_x \sum_y y f_{Y|X}(y|x) f_X(x)$$

$$= \sum_x \left[\sum_y y f_{Y|X}(y|x) \right] f_X(x) = \sum_x E(Y|X = x) f_X(x) = E[E(Y|X)]$$

since the term in brackets is $E(Y|X = x)$. QED

Theorem 2.11: Steiner's Formula or the Conditional Variance Identity. Assume the relevant expectations exist. Then

$$\text{VAR}(Y) = E[\text{VAR}(Y|X)] + \text{VAR}[E(Y|X)].$$

Proof: Following Rice (1988, p. 132), since $\text{VAR}(Y|X) = E(Y^2|X) - [E(Y|X)]^2$ is a random variable,

$$E[\text{VAR}(Y|X)] = E[E(Y^2|X)] - E([E(Y|X)]^2).$$

If W is a random variable then $E(W) = E[E(W|X)]$ by Theorem 2.10 and $\text{VAR}(W) = E(W^2) - [E(W)]^2$ by the short cut formula. Letting $W = E(Y|X)$ gives

$$\text{VAR}(E(Y|X)) = E([E(Y|X)]^2) - (E[E(Y|X)])^2.$$

Since $E(Y^2) = E[E(Y^2|X)]$ and since $E(Y) = E[E(Y|X)]$,

$$\text{VAR}(Y) = E(Y^2) - [E(Y)]^2 = E[E(Y^2|X)] - (E[E(Y|X)])^2.$$

Adding 0 to VAR(Y) gives

$$\text{VAR}(Y) = E[E(Y^2|X)] - E([E(Y|X)]^2) + E([E(Y|X)]^2) - (E[E(Y|X)])^2$$

$= E[\text{VAR}(Y|X)] + \text{VAR}[E(Y|X)]$. QED

A *hierarchical model* models a complicated process by a sequence of models placed in a hierarchy. Interest might be in the marginal expectation $E(Y)$ and marginal variance VAR(Y). One could find the joint pmf from $f(x, y) = f(y|x)f(x)$, then find the marginal distribution $f_Y(y)$ and then find $E(Y)$ and VAR(Y). Alternatively, use Theorems 2.10 and 2.11.

Example 2.12. Suppose $Y|X \sim \text{BIN}(X, \rho)$ and $X \sim$ Poisson (λ). Then $E(Y|X) = X\rho$, $\text{VAR}(Y|X) = X\rho(1-\rho)$ and $E(X) = \text{VAR}(X) = \lambda$. Hence $E(Y) = E[E(Y|X)] = E(X\rho) = \rho E(X) = \rho\lambda$ and $\text{VAR}(Y) = E[\text{VAR}(Y|X)] + \text{VAR}[E(Y|X)] = E[X\rho(1-\rho)] + \text{VAR}(X\rho) = \lambda\rho(1-\rho) + \rho^2\text{VAR}(X) = \lambda\rho(1-\rho) + \rho^2\lambda = \lambda\rho$.

2.4 Location–Scale Families

Many univariate distributions are location, scale or location–scale families. Assume that the random variable Y has a pdf $f_Y(y)$.

Definition 2.16. Let $f_Y(y)$ be the pdf of Y. Then the family of pdfs $f_W(w) = f_Y(w - \mu)$ indexed by the *location parameter* μ, $-\infty < \mu < \infty$, is the *location family* for the random variable $W = \mu + Y$ with *standard pdf* $f_Y(y)$.

Definition 2.17. Let $f_Y(y)$ be the pdf of Y. Then the family of pdfs $f_W(w) = (1/\sigma)f_Y(w/\sigma)$ indexed by the *scale parameter* $\sigma > 0$, is the *scale family* for the random variable $W = \sigma Y$ with *standard pdf* $f_Y(y)$.

Definition 2.18. Let $f_Y(y)$ be the pdf of Y. Then the family of pdfs $f_W(w) = (1/\sigma)f_Y((w - \mu)/\sigma)$ indexed by the *location and scale parameters* μ, $-\infty < \mu < \infty$, and $\sigma > 0$, is the *location–scale family* for the random variable $W = \mu + \sigma Y$ with *standard pdf* $f_Y(y)$.

The most important scale family is the exponential EXP(λ) distribution. Other scale families from Chapter 10 include the chi (p, σ) distribution if p is known, the Gamma G(ν, λ) distribution if ν is known, the lognormal

(μ, σ^2) distribution with scale parameter $\tau = e^\mu$ if σ^2 is known, the one sided stable OSS(σ) distribution, the Pareto PAR(σ, λ) distribution if λ is known, and the Weibull $W(\phi, \lambda)$ distribution with scale parameter $\sigma = \lambda^{1/\phi}$ if ϕ is known.

A location family can be obtained from a location–scale family by fixing the scale parameter while a scale family can be obtained by fixing the location parameter. The most important location–scale families are the Cauchy C(μ, σ), double exponential DE(θ, λ), logistic L(μ, σ), normal N(μ, σ^2) and uniform U(θ_1, θ_2) distributions. Other location–scale families from Chapter 10 include the two parameter exponential EXP(θ, λ), half Cauchy HC(μ, σ), half logistic HL(μ, σ), half normal HN(μ, σ), largest extreme value LEV(θ, σ), Maxwell Boltzmann MB(μ, σ), Rayleigh R(μ, σ) and smallest extreme value SEV(θ, σ) distributions.

2.5 Transformations

Transformations for univariate distributions are important because many "brand name" random variables are transformations of other brand name distributions. These transformations will also be useful for finding the distribution of the complete sufficient statistic for a 1 parameter exponential family. See Chapter 10.

Example 2.13: Common problem. Suppose that Y is a discrete random variable with pmf $f_X(x)$ given by a table. Let the **transformation** $Y = t(X)$ for some function t and find the probability function $f_Y(y)$.
Solution: Step 1) Find $t(x)$ for each value of x.
Step 2) Collect $x : t(x) = y$, and sum the corresponding probabilities:
$f_Y(y) = \sum_{x:t(x)=y} f_X(x)$, and table the resulting pmf $f_Y(y)$ of Y.
For example, if $Y = X^2$ and $f_X(-1) = 1/3, f_X(0) = 1/3$, and $f_X(1) = 1/3$, then $f_Y(0) = 1/3$ and $f_Y(1) = 2/3$.

Definition 2.19. Let $h : D \to \Re$ be a real valued function with domain D. Then h is **increasing** if $h(y_1) < h(y_2)$, *nondecreasing* if $h(y_1) \le h(y_2)$, **decreasing** if $h(y_1) > h(y_2)$ and *nonincreasing* if $h(y_1) \ge h(y_2)$ provided that y_1 and y_2 are any two numbers in D with $y_1 < y_2$. The function h is a monotone function if h is either increasing or decreasing.

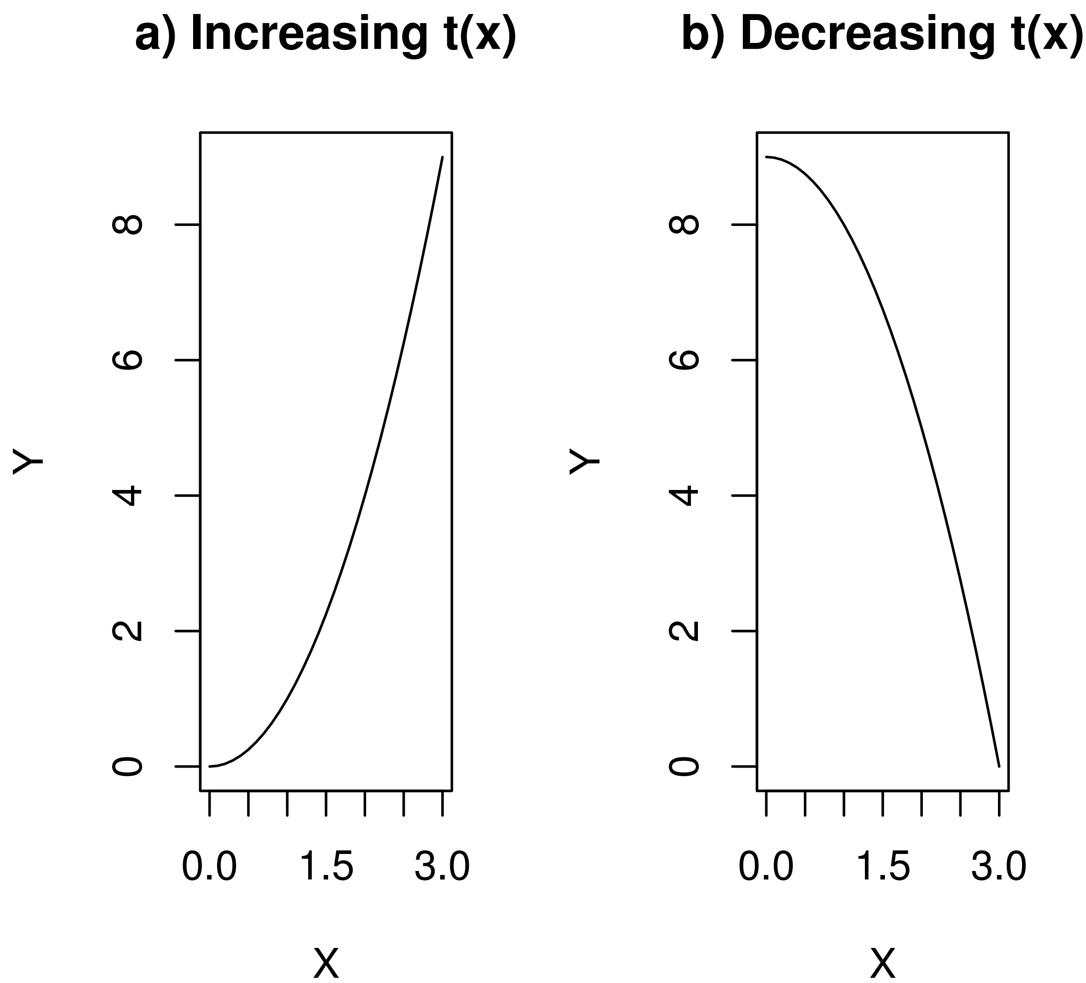

Figure 2.2: Increasing and Decreasing t(x)

Recall that if h is differentiable on an open interval D or continuous on a closed interval D and differentiable on the interior of D, then h is increasing if $h'(y) > 0$ for all y in the interior of D and h is decreasing if $h'(y) < 0$ for all y in the interior of D. Also if h is increasing then $-h$ is decreasing. Similarly, if h is decreasing then $-h$ is increasing.

Suppose that X is a continuous random variable with pdf $f_X(x)$ on support $\mathcal{X}$. Let the transformation $Y = t(X)$ for some monotone function t. Then there are two ways to find the support $\mathcal{Y}$ of $Y = t(x)$ if the support $\mathcal{X}$ of X is an interval with endpoints $a < b$ where $a = -\infty$ and $b = \infty$ are possible. Let $t(a) \equiv \lim_{y \downarrow a} t(y)$ and let $t(b) \equiv \lim_{y \uparrow b} t(y)$. A graph can help. If t is an increasing function, then $\mathcal{Y}$ is an interval with endpoints $t(a) < t(b)$. If t is a decreasing function, then $\mathcal{Y}$ is an interval with endpoints $t(b) < t(a)$. The second method is to find $x = t^{-1}(y)$. Then if $\mathcal{X} = [a, b]$, say, solve $a \leq t^{-1}(y) \leq b$ in terms of y.

If $t(x)$ is increasing then $P(\{Y \leq y\}) = P(\{X \leq t^{-1}(y)\})$ while if $t(x)$ is decreasing $P(\{Y \leq y\}) = P(\{X \geq t^{-1}(y)\})$. To see this, look at Figure 2.2. Suppose the support of Y is $[0, 9]$ and the support of X is $[0, 3]$. Now the height of the curve is $y = t(x)$. Mentally draw a horizontal line from y to $t(x)$ and then drop a vertical line to the x-axis. The value on the x-axis is $t^{-1}(y)$ since $t(t^{-1}(y)) = y$. Hence in Figure 2.2 a) $t^{-1}(4) = 2$ and in Figure 2.2 b) $t^{-1}(8) = 1$. If $w < y$ then $t^{-1}(w) < t^{-1}(y)$ if $t(x)$ is increasing as in Figure 2.2 a), but $t^{-1}(w) > t^{-1}(y)$ if $t(x)$ is decreasing as in Figure 2.2 b). Hence $P(Y \leq y) = P(t^{-1}(Y) \geq t^{-1}(y)) = P(X \geq t^{-1}(y))$.

Theorem 2.12: the CDF Method or Method of Distributions: Suppose that the continuous cdf $F_X(x)$ is known and that $Y = t(X)$. Find the support $\mathcal{Y}$ of Y.
i) If t is an increasing function then, $F_Y(y) = P(Y \leq y) = P(t(X) \leq y) = P(X \leq t^{-1}(y)) = F_X(t^{-1}(y))$.
ii) If t is a decreasing function then, $F_Y(y) = P(Y \leq y) = P(t(X) \leq y) = P(X \geq t^{-1}(y)) = 1 - P(X < t^{-1}(y)) = 1 - P(X \leq t^{-1}(y)) = 1 - F_X(t^{-1}(y))$.
iii) The special case $Y = X^2$ is important. If the support of X is positive, use i). If the support of X is negative, use ii). If the support of X is $(-a, a)$ (where $a = \infty$ is allowed), then $F_Y(y) = P(Y \leq y) =$

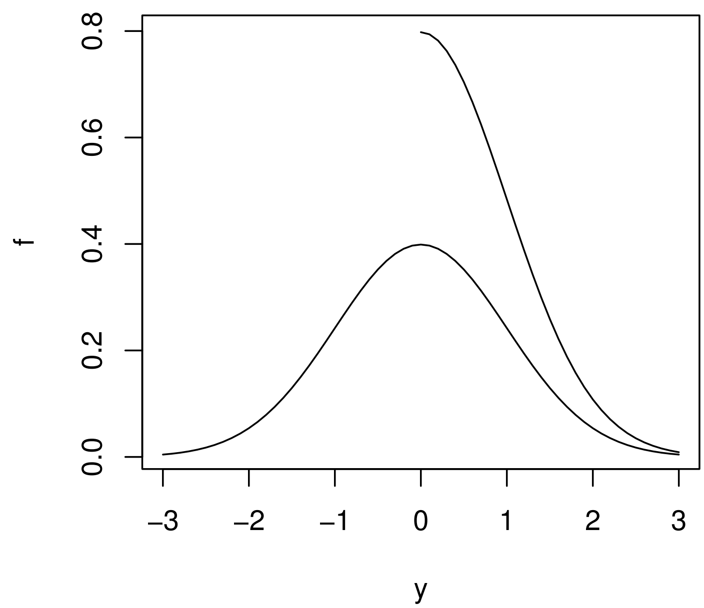

Figure 2.3: Pdfs for N(0,1) and HN(0,1) Distributions

$P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) =$

$$\int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx = F_X(\sqrt{y}) - F_X(-\sqrt{y}),\ 0 \leq y < a^2.$$

After finding the cdf $F_Y(y)$, the pdf of Y is $f_Y(y) = \dfrac{d}{dy}F_Y(y)$ for $y \in \mathcal{Y}$.

Example 2.14. Suppose X has a pdf with support on the real line and that the pdf is symmetric about μ so $f_X(\mu - w) = f_X(\mu + w)$ for all real w. It can be shown that X has a symmetric distribution about μ if $Z = X - \mu$ and $-Z = \mu - X$ have the same distribution. Several named right skewed distributions with support $y \geq \mu$ are obtained by the transformation $Y = \mu + |X - \mu|$. Similarly, let U be a U(0,1) random variable that is independent of Y, then X can be obtained from Y by letting $X = Y$ if $U \leq 0.5$ and $X = 2\mu - Y$ if $U > 0.5$. Pairs of such distributions include the exponential and double exponential, normal and half normal, Cauchy and half Cauchy, and logistic and half logistic distributions. Figure 2.3 shows the N(0,1) and HN(0,1) pdfs.

Notice that for $y \geq \mu$,

$$F_Y(y) = P(Y \leq y) = P(\mu + |X - \mu| \leq y) = P(|X - \mu| \leq y - \mu) =$$

$$P(\mu - y \leq X - \mu \leq y - \mu) = P(2\mu - y \leq X \leq y) = F_X(y) - F_X(2\mu - y).$$

Taking derivatives and using the symmetry of f_X gives $f_Y(y) =$

$$f_X(y) + f_X(2\mu - y) = f_X(\mu + (y - \mu)) + f_X(\mu - (y - \mu)) = 2f_X(\mu + (y - \mu))$$

$= 2f_X(y)$ for $y \geq \mu$. Hence $f_Y(y) = 2f_X(y)I(y \geq \mu)$.

Then X has pdf

$$f_X(x) = \frac{1}{2} f_Y(\mu + |x - \mu|)$$

for all real x, since this pdf is symmetric about μ and $f_X(x) = 0.5 f_Y(x)$ if $x \geq \mu$.

Example 2.15. Often the rules of differentiation such as the multiplication, quotient and chain rules are needed. For example if the support of X is $[-a, a]$ and if $Y = X^2$, then

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})]$$

for $0 \leq y \leq a^2$.

Theorem 2.13: the Transformation Method. Assume that X has pdf $f_X(x)$ and support $\mathcal{X}$. Find the support $\mathcal{Y}$ of $Y = t(X)$. If $t(x)$ is either increasing or decreasing on $\mathcal{X}$ and if $t^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$, then $Y = t(X)$ has pdf

$$f_Y(y) = f_X(t^{-1}(y)) \left| \frac{dt^{-1}(y)}{dy} \right| \tag{2.16}$$

for $y \in \mathcal{Y}$. As always, $f_Y(y) = 0$ for y not in $\mathcal{Y}$.

Proof: Examining Theorem 2.12, if t is increasing then $F_Y(y) = F_X(t^{-1}(y))$ and

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

$$= \frac{d}{dy} F_X(t^{-1}(y)) = f_X(t^{-1}(y)) \frac{d}{dy} t^{-1}(y) = f_X(t^{-1}(y)) \left| \frac{dt^{-1}(y)}{dy} \right|$$

for $y \in \mathcal{Y}$ since the derivative of a differentiable increasing function is positive.

If t is a decreasing function then from Theorem 2.12, $F_Y(y) = 1 - F_X(t^{-1}(x))$. Hence

$$f_Y(y) = \frac{d}{dy}[1 - F_X(t^{-1}(y))] = -f_X(t^{-1}(y)) \frac{d}{dy} t^{-1}(y) = f_X(t^{-1}(y)) \left| \frac{dt^{-1}(y)}{dy} \right|$$

for $y \in \mathcal{Y}$ since the derivative of a differentiable decreasing function is negative.

Tips: To be useful, formula (2.16) should be simplified as much as possible.
a) The pdf of Y will often be that of a gamma random variable. In particular, the pdf of Y is often the pdf of an exponential(λ) random variable.
b) To find the inverse function $x = t^{-1}(y)$, solve the equation $y = t(x)$ for x.
c) The log transformation is often used. Know how to sketch $\log(x)$ and e^x for $x > 0$. Recall that in this text, $\log(x)$ is the natural logarithm of x.
d) If $\mathcal{X}$ is an interval with endpoints a and b, find

$$\mathcal{Y} = (\min\{t(a), t(b)\}, \max\{t(a), t(b)\})$$

as described on p. 49.

Example 2.16. Let X be a random variable with pdf

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

where $x > 0$, μ is real and $\sigma > 0$. Let $Y = \log(X)$ and find the distribution of Y.

Solution: $X = e^Y = t^{-1}(Y)$. So

$$\left| \frac{dt^{-1}(y)}{dy} \right| = |e^y| = e^y,$$

and

$$f_Y(y) = f_X(t^{-1}(y)) \left| \frac{dt^{-1}(y)}{dy} \right| = f_X(e^y)e^y =$$

$$\frac{1}{e^y\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\log(e^y)-\mu)^2}{2\sigma^2}\right) e^y =$$
$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2}\right)$$
for $y \in (-\infty, \infty)$ since $x > 0$ implies that $y = \log(x) \in (-\infty, \infty)$. Notice that X is lognormal (μ, σ^2) and $Y \sim N(\mu, \sigma^2)$.

Example 2.17. If Y has a Topp–Leone distribution, then pdf of Y is

$$f(y) = \nu(2-2y)(2y-y^2)^{\nu-1}$$

for $\nu > 0$ and $0 < y < 1$. Notice that $F(y) = (2y-y^2)^\nu$ for $0 < y < 1$ since $F'(y) = f(y)$. Then the distribution of $W = -\log(2Y - Y^2)$ will be of interest for later chapters.

Let $X = Y - 1$. Then the support of X is $(-1, 0)$ and $F_X(x) =$
$P(X \le x) = P(Y - 1 \le x) = P(Y \le x+1) = F_Y(x+1)$

$$= (2(x+1)-(x+1)^2)^\nu = ((x+1)(2-(x+1)))^\nu = [(x+1)(x-1)]^\nu = (1-x^2)^\nu.$$

So $F_X(x) = (1-x^2)^\nu$ for $-1 < x < 0$. Now the support of W is $w > 0$ and $F_W(w) = P(W \le w) = P(-\log(2Y-Y^2) \le w) = P(\log(2Y-Y^2) \ge -w) =$ $P(2Y-Y^2 \ge e^{-w}) = P(2Y-Y^2-1 \ge e^{-w}-1) = P(-(Y-1)^2 \ge e^{-w}-1) =$ $P((Y-1)^2 \le 1-e^{-w})$. So $F_W(w) = P(X^2 \le 1 - e^{-w}) =$
$P(-\sqrt{a} \le X \le \sqrt{a})$ where $a = 1 - e^{-w} \in (0,1)$. So $F_W(w) =$
$F_X(\sqrt{a}) - F_X(-\sqrt{a}) = 1 - F_X(-\sqrt{a}) = 1 - F_X(-\sqrt{1-e^{-w}})$

$$= 1 - [1 - (-\sqrt{1-e^{-w}})^2]^\nu = 1 - [1-(1-e^{-w})]^\nu = 1 - e^{-w\nu}$$

for $w > 0$. Thus $W = -\log(2Y - Y^2) \sim EXP(1/\nu)$.

Transformations for vectors are often less useful in applications because the transformation formulas tend to be impractical to compute. For the theorem below, typically $n = 2$. If $Y_1 = t_1(X_1, X_2)$ is of interest, choose $Y_2 = t_2(X_1, X_2)$ such that the determinant J is easy to compute. For example, $Y_2 = X_2$ may work. Finding the support $\mathcal{Y}$ can be difficult, but if the joint pdf of X_1, X_2 is $g(x_1, x_2) = h(x_1, x_2)\ I[(x_1, x_2) \in \mathcal{X}]$, then the joint pdf of Y_1, Y_2 is

$$f(y_1, y_2) = h(t_1^{-1}(\boldsymbol{y}), t_2^{-1}(\boldsymbol{y}))\ I[(t_1^{-1}(\boldsymbol{y}), t_2^{-1}(\boldsymbol{y})) \in \mathcal{X}]\ |J|,$$

and using $I[(t_1^{-1}(\boldsymbol{y}), t_2^{-1}(\boldsymbol{y})) \in \mathcal{X}]$ can be useful for finding $\mathcal{Y}$. The fact that $\prod_{j=1}^k I_{A_j}(\boldsymbol{y}) = I_{\cap_{j=1}^k A_j}(\boldsymbol{y})$ can also be useful. See Problem 2.67. Also sketch $\mathcal{X}$ with x_1 on the horizontal axis and x_2 on the vertical axis, and sketch $\mathcal{Y}$ with y_1 on the horizontal axis and y_2 on the vertical axis.

Theorem 2.14: the Multivariate Transformation Method. Let $X_1, ..., X_n$ be random variables with joint pdf $g(x_1, ..., x_n)$ and support $\mathcal{X}$. Let $Y_i = t_i(X_1, ..., X_n)$ for $i = 1, ..., n$. Suppose that $f(y_1, ..., y_n)$ is the joint pdf of $Y_1, ..., Y_n$ and that the multivariate transformation is one to one. Hence the transformation is invertible and can be solved for the equations $x_i = t_i^{-1}(y_1, ..., y_n)$ for $i = 1, ..., n$. Then the Jacobian of this multivariate transformation is

$$J = \det \begin{bmatrix} \frac{\partial t_1^{-1}}{\partial y_1} & \ldots & \frac{\partial t_1^{-1}}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial t_n^{-1}}{\partial y_1} & \ldots & \frac{\partial t_n^{-1}}{\partial y_n} \end{bmatrix}.$$

Let $|J|$ denote the absolute value of the determinant J. Then the pdf of $Y_1, ..., Y_n$ is

$$f(y_1, ..., y_n) = g(t_1^{-1}(\boldsymbol{y}), ..., t_n^{-1}(\boldsymbol{y}))\ |J|. \tag{2.17}$$

Example 2.18. Let X_1 and X_2 have joint pdf

$$g(x_1, x_2) = 2e^{-(x_1+x_2)}$$

for $0 < x_1 < x_2 < \infty$. Let $Y_1 = X_1$ and $Y_2 = X_1 + X_2$. An important step is finding the support $\mathcal{Y}$ of (Y_1, Y_2) from the support of (X_1, X_2)

$$= \mathcal{X} = \{(x_1, x_2) | 0 < x_1 < x_2 < \infty\}.$$

Now $x_1 = y_1 = t_1^{-1}(y_1, y_2)$ and $x_2 = y_2 - y_1 = t_2^{-1}(y_1, y_2)$. Hence $x_1 < x_2$ implies $y_1 < y_2 - y_1$ or $2y_1 < y_2$, and

$$\mathcal{Y} = \{(y_1, y_2) | 0 < 2y_1 < y_2\}.$$

Now

$$\frac{\partial t_1^{-1}}{\partial y_1} = 1, \quad \frac{\partial t_1^{-1}}{\partial y_2} = 0,$$
$$\frac{\partial t_2^{-1}}{\partial y_1} = -1, \quad \frac{\partial t_2^{-1}}{\partial y_2} = 1,$$

and the Jacobian

$$J = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1.$$

Hence $|J| = 1$. Using indicators,

$$g_{X_1,X_2}(x_1, x_2) = 2e^{-(x_1+x_2)} I(0 < x_1 < x_2 < \infty),$$

and

$$f_{Y_1,Y_2}(y_1, y_2) = g_{X_1,X_2}(y_1, y_2 - y_1)|J| = 2e^{-(y_1+y_2-y_1)} I(0 < y_1 < y_2 - y_1)1 =$$

$$2e^{-y_2} I(0 < 2y_1 < y_2).$$

Notice that Y_1 and Y_2 are not independent since the support $\mathcal{Y}$ is not a cross product. The marginals

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} 2e^{-y_2} I(0 < 2y_1 < y_2) dy_2 = \int_{2y_1}^{\infty} 2e^{-y_2} dy_2$$

$$= -2e^{-y_2}\Big|_{y_2=2y_1}^{\infty} = 0 - -2e^{-2y_1} = 2e^{-2y_1}$$

for $0 < y_1 < \infty$, and

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} 2e^{-y_2} I(0 < 2y_1 < y_2) dy_1 = \int_{0}^{y_2/2} 2e^{-y_2} dy_1$$

$$= 2e^{-y_2} y_1 \Big|_{y_1=0}^{y_1=y_2/2} = y_2 e^{-y_2}$$

for $0 < y_2 < \infty$.

Example 2.19. Following Bickel and Doksum (2007, p. 489-490), let X_1 and X_2 be independent gamma (ν_i, λ) RVs for $i = 1, 2$. Then X_1 and X_2 have joint pdf $g(x_1, x_2) = g_1(x_1)g_2(x_2) =$

$$\frac{x_1^{\nu_1-1} e^{-x_1/\lambda}}{\lambda^{\nu_1}\Gamma(\nu_1)} \frac{x_2^{\nu_2-1} e^{-x_2/\lambda}}{\lambda^{\nu_2}\Gamma(\nu_2)} = \frac{1}{\lambda^{\nu_1+\nu_2}\Gamma(\nu_1)\Gamma(\nu_2)} x_1^{\nu_1-1} x_2^{\nu_2-1} \exp[-(x_1 + x_2)/\lambda]$$

for $0 < x_1$ and $0 < x_2$. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1/(X_1 + X_2)$. An important step is finding the support $\mathcal{Y}$ of (Y_1, Y_2) from the support of (X_1, X_2)

$$= \mathcal{X} = \{(x_1, x_2) | 0 < x_1 \text{ and } 0 < \mathrm{x}_2\}.$$

Now $y_2 = x_1/y_1$, so $x_1 = y_1 y_2 = t_1^{-1}(y_1, y_2)$ and $x_2 = y_1 - x_1 = y_1 - y_1 y_2 = t_2^{-1}(y_1, y_2)$. Notice that $0 < y_1$ and $0 < x_1 < x_1 + x_2$. Thus $0 < y_2 < 1$, and

$$\mathcal{Y} = \{(y_1, y_2) | 0 < y_1 \text{ and } 0 < \text{y}_2 < 1\}.$$

Now

$$\frac{\partial t_1^{-1}}{\partial y_1} = y_2, \quad \frac{\partial t_1^{-1}}{\partial y_2} = y_1,$$

$$\frac{\partial t_2^{-1}}{\partial y_1} = 1 - y_2, \quad \frac{\partial t_2^{-1}}{\partial y_2} = -y_1,$$

and the Jacobian

$$J = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1 y_2 - (y_1 - y_1 y_2) = -y_1,$$

and $|J| = y_1$. So the joint pdf

$$f(y_1, y_2) = g(t_1^{-1}(\boldsymbol{y}),\ t_2^{-1}(\boldsymbol{y}))\ |J| = g(y_1 y_2, y_1 - y_1 y_2) y_1 =$$

$$\frac{1}{\lambda^{\nu_1+\nu_2}\Gamma(\nu_1)\Gamma(\nu_2)} y_1^{\nu_1 - 1} y_2^{\nu_1 - 1} y_1^{\nu_2 - 1} (1 - y_2)^{\nu_2 - 1} \exp[-(y_1 y_2 + y_1 - y_1 y_2)/\lambda] y_1 =$$

$$\frac{1}{\lambda^{\nu_1+\nu_2}\Gamma(\nu_1)\Gamma(\nu_2)} y_1^{\nu_1+\nu_2-1} y_2^{\nu_1 - 1} (1 - y_2)^{\nu_2 - 1} e^{-y_1/\lambda} =$$

$$\frac{1}{\lambda^{\nu_1+\nu_2}\Gamma(\nu_1+\nu_2)}\ y_1^{\nu_1+\nu_2-1}\ e^{-y_1/\lambda} \quad \frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)}\ y_2^{\nu_1-1}\ (1 - y_2)^{\nu_2 - 1}.$$

Thus $f(y_1, y_2) = f_1(y_1) f_2(y_2)$ on $\mathcal{Y}$, and $Y_1 \sim \text{gamma}(\nu_1 + \nu_2, \lambda) \perp\!\!\!\perp Y_2 \sim \text{beta}(\nu_1, \nu_2)$ by Theorem 2.2b.

2.6 Sums of Random Variables

An important multivariate transformation of the random variables $\boldsymbol{Y} = (Y_1, ..., Y_n)$ is $T(Y_1, ..., Y_n) = \sum_{i=1}^n Y_i$. Some properties of sums are given below.

Theorem 2.15. Assume that all relevant expectations exist. Let a, $a_1, ..., a_n$ and $b_1, ..., b_m$ be constants. Let $Y_1, ..., Y_n$, and $X_1, ..., X_m$ be random variables. Let $g_1, ..., g_k$ be functions of $Y_1, ..., Y_n$.

i) $E(a) = a$.

ii) $E[aY] = aE[Y]$
iii) $V(aY) = a^2 V(Y)$.
iv) $E[g_1(Y_1, ..., Y_n) + \cdots + g_k(Y_1, ..., Y_n)] = \sum_{i=1}^k E[g_i(Y_1, ..., Y_n)]$.
Let $W_1 = \sum_{i=1}^n a_i Y_i$ and $W_2 = \sum_{i=1}^m b_i X_i$.

v) $E(W_1) = \displaystyle\sum_{i=1}^n a_i E(Y_i)$.

vi) $V(W_1) = \text{Cov}(W_1, W_1) = \displaystyle\sum_{i=1}^n a_i^2 V(Y_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_i a_j \text{Cov}(Y_i, Y_j)$.

vii) $\text{Cov}(W_1, W_2) = \displaystyle\sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(Y_i, X_j)$.

viii) $E(\sum_{i=1}^n Y_i) = \displaystyle\sum_{i=1}^n E(Y_i)$.

ix) If $Y_1, ..., Y_n$ are independent, $V(\sum_{i=1}^n Y_i) = \displaystyle\sum_{i=1}^n V(Y_i)$.

Let $Y_1, ..., Y_n$ be iid RVs with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$, then the **sample mean** $\overline{Y} = \displaystyle\frac{1}{n} \sum_{i=1}^n Y_i$. Then
x) $E(\overline{Y}) = \mu$ and
xi) $V(\overline{Y}) = \sigma^2/n$.

Definition 2.20. $Y_1, ..., Y_n$ are a **random sample** or **iid** if $Y_1, ..., Y_n$ are independent and identically distributed (all of the Y_i have the same distribution).

Example 2.20: Common problem. Let $Y_1, ..., Y_n$ be independent random variables with $E(Y_i) = \mu_i$ and $V(Y_i) = \sigma_i^2$. Let $W = \sum_{i=1}^n Y_i$. Then
a) $E(W) = E(\sum_{i=1}^n Y_i) = \sum_{i=1}^n E(Y_i) = \sum_{i=1}^n \mu_i$, and
b) $V(W) = V(\sum_{i=1}^n Y_i) = \sum_{i=1}^n V(Y_i) = \sum_{i=1}^n \sigma_i^2$.

A **statistic** is a function of the random sample and known constants. A statistic is a random variable and the **sampling distribution** of a statistic is the distribution of the statistic. Important statistics are $\sum_{i=1}^n Y_i$, $\overline{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ and $\sum_{i=1}^n a_i Y_i$ where $a_1, ..., a_n$ are constants. The following theorem shows how to find the mgf and characteristic function of such statistics.

Theorem 2.16. a) The characteristic function uniquely determines the distribution.

b) If the moment generating function exists, then it uniquely determines the distribution.

c) Assume that $Y_1, ..., Y_n$ are independent with characteristic functions $\phi_{Y_i}(t)$. Then the characteristic function of $W = \sum_{i=1}^n Y_i$ is

$$\phi_W(t) = \prod_{i=1}^n \phi_{Y_i}(t). \tag{2.18}$$

d) Assume that $Y_1, ..., Y_n$ are iid with characteristic functions $\phi_Y(t)$. Then the characteristic function of $W = \sum_{i=1}^n Y_i$ is

$$\phi_W(t) = [\phi_Y(t)]^n. \tag{2.19}$$

e) Assume that $Y_1, ..., Y_n$ are independent with mgfs $m_{Y_i}(t)$. Then the mgf of $W = \sum_{i=1}^n Y_i$ is

$$m_W(t) = \prod_{i=1}^n m_{Y_i}(t). \tag{2.20}$$

f) Assume that $Y_1, ..., Y_n$ are iid with mgf $m_Y(t)$. Then the mgf of $W = \sum_{i=1}^n Y_i$ is

$$m_W(t) = [m_Y(t)]^n. \tag{2.21}$$

g) Assume that $Y_1, ..., Y_n$ are independent with characteristic functions $\phi_{Y_i}(t)$. Then the characteristic function of $W = \sum_{j=1}^n (a_j + b_j Y_j)$ is

$$\phi_W(t) = \exp(it \sum_{j=1}^n a_j) \ \prod_{j=1}^n \phi_{Y_j}(b_j t). \tag{2.22}$$

h) Assume that $Y_1, ..., Y_n$ are independent with mgfs $m_{Y_i}(t)$. Then the mgf of $W = \sum_{i=1}^n (a_i + b_i Y_i)$ is

$$m_W(t) = \exp(t \sum_{i=1}^n a_i) \ \prod_{i=1}^n m_{Y_i}(b_i t). \tag{2.23}$$

Proof of g): Recall that $\exp(w) = e^w$ and $\exp(\sum_{j=1}^n d_j) = \prod_{j=1}^n \exp(d_j)$. It can be shown that for the purposes of this proof, that the complex constant

i in the characteristic function (cf) can be treated in the same way as if it were a real constant. Now

$$\phi_W(t) = E(e^{itW}) = E(\exp[it\sum_{j=1}^{n}(a_j + b_jY_j)])$$

$$= \exp(it\sum_{j=1}^{n} a_j)\ E(\exp[\sum_{j=1}^{n} itb_jY_j)])$$

$$= \exp(it\sum_{j=1}^{n} a_j)\ E(\prod_{i=1}^{n} \exp[itb_jY_j)])$$

$$= \exp(it\sum_{j=1}^{n} a_j)\ \prod_{i=1}^{n} E[\exp(itb_jY_j)]$$

since by Theorem 2.5 the expected value of a product of independent random variables is the product of the expected values of the independent random variables. Now in the definition of a cf, the t is a dummy variable as long as t is real. Hence $\phi_Y(t) = E[\exp(itY)]$ and $\phi_Y(s) = E[\exp(isY)]$. Taking $s = tb_j$ gives $E[\exp(itb_jY_j)] = \phi_{Y_j}(tb_j)$. Thus

$$\phi_W(t) = \exp(it\sum_{j=1}^{n} a_j)\ \prod_{i=1}^{n} \phi_{Y_j}(tb_j). \text{ QED}$$

The distribution of $W = \sum_{i=1}^{n} Y_i$ is known as the convolution of $Y_1, ..., Y_n$. Even for $n = 2$ convolution formulas tend to be hard; however, the following two theorems suggest that to find the distribution of $W = \sum_{i=1}^{n} Y_i$, first find the mgf or characteristic function of W using Theorem 2.16. If the mgf or cf is that of a brand name distribution, then W has that distribution. For example, if the mgf of W is a normal (ν, τ^2) mgf, then W has a normal (ν, τ^2) distribution, written $W \sim N(\nu, \tau^2)$. This technique is useful for several brand name distributions. Chapter 10 will show that many of these distributions are exponential families.

Theorem 2.17. a) If $Y_1, ..., Y_n$ are independent binomial BIN(k_i, ρ) random variables, then

$$\sum_{i=1}^{n} Y_i \sim \text{BIN}(\sum_{\text{i}=1}^{\text{n}} \text{k}_\text{i}, \rho).$$

Thus if $Y_1, ..., Y_n$ are iid BIN(k, ρ) random variables, then $\sum_{i=1}^n Y_i \sim$ BIN(nk, ρ).

b) Denote a chi–square χ^2_p random variable by $\chi^2(p)$. If $Y_1, ..., Y_n$ are independent chi–square $\chi^2_{p_i}$, then

$$\sum_{i=1}^n Y_i \sim \chi^2(\sum_{i=1}^n p_i).$$

Thus if $Y_1, ..., Y_n$ are iid χ^2_p, then

$$\sum_{i=1}^n Y_i \sim \chi^2_{np}.$$

c) If $Y_1, ..., Y_n$ are iid exponential EXP(λ), then

$$\sum_{i=1}^n Y_i \sim G(n, \lambda).$$

d) If $Y_1, ..., Y_n$ are independent Gamma $G(\nu_i, \lambda)$ then

$$\sum_{i=1}^n Y_i \sim G(\sum_{i=1}^n \nu_i, \lambda).$$

Thus if $Y_1, ..., Y_n$ are iid $G(\nu, \lambda)$, then

$$\sum_{i=1}^n Y_i \sim G(n\nu, \lambda).$$

e) If $Y_1, ..., Y_n$ are independent normal $N(\mu_i, \sigma_i^2)$, then

$$\sum_{i=1}^n (a_i + b_i Y_i) \sim N(\sum_{i=1}^n (a_i + b_i \mu_i), \sum_{i=1}^n b_i^2 \sigma_i^2).$$

Here a_i and b_i are fixed constants. Thus if $Y_1, ..., Y_n$ are iid $N(\mu, \sigma)$, then $\overline{Y} \sim N(\mu, \sigma^2/n)$.

f) If $Y_1, ..., Y_n$ are independent Poisson POIS(θ_i), then

$$\sum_{i=1}^n Y_i \sim \text{POIS}(\sum_{\mathrm{i=1}}^{\mathrm{n}} \theta_{\mathrm{i}}).$$

Thus if $Y_1, ..., Y_n$ are iid $POIS(\theta)$, then

$$\sum_{i=1}^{n} Y_i \sim \text{POIS}(\text{n}\theta).$$

Theorem 2.18. a) If $Y_1, ..., Y_n$ are independent Cauchy $C(\mu_i, \sigma_i)$, then

$$\sum_{i=1}^{n}(a_i + b_i Y_i) \sim C(\sum_{i=1}^{n}(a_i + b_i \mu_i), \sum_{i=1}^{n} |b_i|\sigma_i).$$

Thus if $Y_1, ..., Y_n$ are iid $C(\mu, \sigma)$, then $\overline{Y} \sim C(\mu, \sigma)$.

b) If $Y_1, ..., Y_n$ are iid geometric geom(p), then

$$\sum_{i=1}^{n} Y_i \sim \text{NB}(\text{n}, \text{p}).$$

c) If $Y_1, ..., Y_n$ are iid inverse Gaussian $IG(\theta, \lambda)$, then

$$\sum_{i=1}^{n} Y_i \sim IG(n\theta, n^2\lambda).$$

Also

$$\overline{Y} \sim IG(\theta, n\lambda).$$

d) If $Y_1, ..., Y_n$ are independent negative binomial NB(r_i, ρ), then

$$\sum_{i=1}^{n} Y_i \sim \text{NB}(\sum_{\text{i}=1}^{\text{n}} \text{r}_\text{i}, \rho).$$

Thus if $Y_1, ..., Y_n$ are iid $NB(r, \rho)$, then

$$\sum_{i=1}^{n} Y_i \sim NB(nr, \rho).$$

Example 2.21: Common problem. Given that $Y_1, ..., Y_n$ are independent random variables from one of the distributions in Theorem 2.17, find the distribution of $W = \sum_{i=1}^{n} Y_i$ or $W = \sum_{i=1}^{n} b_i Y_i$ by finding the mgf or characteristic function of W and recognizing that it comes from a brand name distribution.

Tips: a) in the product, anything that does not depend on the product index i is treated as a constant.
b) $\exp(a) = e^a$ and $\log(y) = \ln(y) = \log_e(y)$ is the **natural logarithm**.
c)

$$\prod_{i=1}^n a^{b\theta_i} = a^{\sum_{i=1}^n b\theta_i} = a^{b\sum_{i=1}^n \theta_i}.$$

In particular, $\prod_{i=1}^n \exp(b\theta_i) = \exp(\sum_{i=1}^n b\theta_i) = \exp(b\sum_{i=1}^n \theta_i)$.

Example 2.22. Suppose $Y_1, ..., Y_n$ are iid $IG(\theta, \lambda)$ where the mgf

$$m_{Y_i}(t) = m(t) = \exp\left[\frac{\lambda}{\theta}\left(1 - \sqrt{1 - \frac{2\theta^2 t}{\lambda}}\right)\right]$$

for $t < \lambda/(2\theta^2)$. Then

$$m_{\sum_{i=1}^n Y_i}(t) = \prod_{i=1}^n m_{Y_i}(t) = [m(t)]^n = \exp\left[\frac{n\lambda}{\theta}\left(1 - \sqrt{1 - \frac{2\theta^2 t}{\lambda}}\right)\right]$$

$$= \exp\left[\frac{n^2\lambda}{n\ \theta}\left(1 - \sqrt{1 - \frac{2(n\theta)^2\ t}{n^2\lambda}}\right)\right]$$

which is the mgf of an $IG(n\theta, n^2\lambda)$ RV. The last equality was obtained by multiplying $\frac{n\lambda}{\theta}$ by $1 = n/n$ and by multiplying $\frac{2\theta^2 t}{\lambda}$ by $1 = n^2/n^2$. Hence $\sum_{i=1}^n Y_i \sim IG(n\theta, n^2\lambda)$.

2.7 Random Vectors

Definition 2.21. $\boldsymbol{Y} = (Y_1, ..., Y_p)$ is a $1 \times p$ **random vector** if Y_i is a random variable for $i = 1, ..., p$. $\boldsymbol{Y}$ is a discrete random vector if each Y_i is discrete, and $\boldsymbol{Y}$ is a continuous random vector if each Y_i is continuous. A random variable Y_1 is the special case of a random vector with $p = 1$.

In the previous sections each $\boldsymbol{Y} = (Y_1, ..., Y_n)$ was a random vector. In this section we will consider n random vectors $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$. Often double subscripts will be used: $\boldsymbol{Y}_i = (Y_{i,1}, ..., Y_{i,p_i})$ for $i = 1, ..., n$.

Notation. The notation for random vectors is rather awkward. In most of the statistical inference literature, $\boldsymbol{Y}$ is a row vector, but in most of the multivariate analysis literature $\boldsymbol{Y}$ is a column vector. In this text, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both vectors, a phrase with $\boldsymbol{Y}$ and $\boldsymbol{X}^T$ means that $\boldsymbol{Y}$ is a column vector and $\boldsymbol{X}^T$ is a row vector where T stands for transpose. Hence in the definition below, first $E(\boldsymbol{Y})$ is a $p \times 1$ row vector, but in the definition of Cov($\boldsymbol{Y}$) below, $E(\boldsymbol{Y})$ and $\boldsymbol{Y} - E(\boldsymbol{Y})$ are $p \times 1$ column vectors and $(\boldsymbol{Y} - E(\boldsymbol{Y}))^T$ is a $1 \times p$ row vector.

Definition 2.22. The *population mean* or **expected value** of a random $1 \times p$ random vector $(Y_1, ..., Y_p)$ is

$$E(\boldsymbol{Y}) = (E(Y_1), ..., E(Y_p))$$

provided that $E(Y_i)$ exists for $i = 1, ..., p$. Otherwise the expected value does not exist. Now let $\boldsymbol{Y}$ be a $p \times 1$ column vector. The $p \times p$ *population covariance matrix*

$$\text{Cov}(\boldsymbol{Y}) = E(\boldsymbol{Y} - E(\boldsymbol{Y}))(\boldsymbol{Y} - E(\boldsymbol{Y}))^T = ((\sigma_{i,j}))$$

where the ij entry of Cov($\boldsymbol{Y}$) is $\text{Cov}(Y_i, Y_j) = \sigma_{i,j}$ provided that each $\sigma_{i,j}$ exists. Otherwise Cov($\boldsymbol{Y}$) does not exist.

The covariance matrix is also called the variance–covariance matrix and variance matrix. Sometimes the notation Var($\boldsymbol{Y}$) is used. Note that Cov($\boldsymbol{Y}$) is a symmetric positive semidefinite matrix. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $p \times 1$ random vectors, $\boldsymbol{a}$ a conformable constant vector and $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable constant matrices, then

$$E(\boldsymbol{a} + \boldsymbol{X}) = \boldsymbol{a} + E(\boldsymbol{X}) \ \text{ and } \ E(\boldsymbol{X} + \boldsymbol{Y}) = E(\boldsymbol{X}) + E(\boldsymbol{Y}) \tag{2.24}$$

and

$$E(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}E(\boldsymbol{X}) \ \text{ and } \ E(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = \boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B}. \tag{2.25}$$

Thus

$$\text{Cov}(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{X}) = \text{Cov}(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}\text{Cov}(\boldsymbol{X})\boldsymbol{A}^T. \tag{2.26}$$

Definition 2.23. Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ be random vectors with joint pdf or pmf $f(\boldsymbol{y}_1, ..., \boldsymbol{y}_n)$. Let $f_{\boldsymbol{Y}_i}(\boldsymbol{y}_i)$ be the marginal pdf or pmf of $\boldsymbol{Y}_i$. Then $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$

are **independent random vectors** if

$$f(\boldsymbol{y}_1, ..., \boldsymbol{y}_n) = f_{\boldsymbol{Y}_1}(\boldsymbol{y}_1) \cdots f_{\boldsymbol{Y}_n}(\boldsymbol{y}_n) = \prod_{i=1}^{n} f_{\boldsymbol{Y}_i}(\boldsymbol{y}_i).$$

The following theorem is a useful generalization of Theorem 2.7.

Theorem 2.19. Let $\boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$ be independent random vectors where $\boldsymbol{Y}_i$ is a $1 \times p_i$ vector for $i = 1, ..., n$. and let $\boldsymbol{h}_i : \Re^{p_i} \to \Re^{p_{j_i}}$ be vector valued functions and suppose that $\boldsymbol{h}_i(\boldsymbol{y}_i)$ is a function of $\boldsymbol{y}_i$ alone for $i = 1, ..., n$. Then the random vectors $\boldsymbol{X}_i = \boldsymbol{h}_i(\boldsymbol{Y}_i)$ are independent. There are three important special cases.
i) If $p_{j_i} = 1$ so that each h_i is a real valued function, then the random variables $X_i = h_i(\boldsymbol{Y}_i)$ are independent.
ii) If $p_i = p_{j_i} = 1$ so that each Y_i and each $X_i = h(Y_i)$ are random variables, then $X_1, ..., X_n$ are independent.
iii) Let $\boldsymbol{Y} = (Y_1, ..., Y_n)$ and $\boldsymbol{X} = (X_1, .., X_m)$ and assume that $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{X}$. If $\boldsymbol{h}(\boldsymbol{Y})$ is a vector valued function of $\boldsymbol{Y}$ alone and if $\boldsymbol{g}(\boldsymbol{X})$ is a vector valued function of $\boldsymbol{X}$ alone, then $\boldsymbol{h}(\boldsymbol{Y})$ and $\boldsymbol{g}(\boldsymbol{X})$ are independent random vectors.

Definition 2.24. The **characteristic function** (cf) of a random vector $\boldsymbol{Y}$ is

$$\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = E(e^{i\boldsymbol{t}^T\boldsymbol{Y}})$$

$\forall \boldsymbol{t} \in \Re^n$ where the complex number $i = \sqrt{-1}$.

Definition 2.25. The **moment generating function** (mgf) of a random vector $\boldsymbol{Y}$ is

$$m_{\boldsymbol{Y}}(\boldsymbol{t}) = E(e^{\boldsymbol{t}^T\boldsymbol{Y}})$$

provided that the expectation exists for all $\boldsymbol{t}$ in some neighborhood of the origin $\boldsymbol{0}$.

Theorem 2.20. If $Y_1, ..., Y_n$ have mgf $m(\boldsymbol{t})$, then moments of all orders exist and

$$E(Y_{i_1}^{k_1} \cdots Y_{i_j}^{k_j}) = \left. \frac{\partial^{k_1+\cdots+k_j}}{\partial t_{i_1}^{k_1} \cdots \partial t_{i_j}^{k_j}} m(\boldsymbol{t}) \right|_{\boldsymbol{t}=\boldsymbol{0}}.$$

In particular,

$$E(Y_i) = \left. \frac{\partial m(\boldsymbol{t})}{\partial t_i} \right|_{\boldsymbol{t}=\boldsymbol{0}}$$

and

$$E(Y_iY_j) = \left.\frac{\partial^2 m(\boldsymbol{t})}{\partial t_i \partial t_j}\right|_{\boldsymbol{t}=\boldsymbol{0}}.$$

Theorem 2.21. If $Y_1, ..., Y_n$ have a cf $\phi_{\boldsymbol{Y}}(\boldsymbol{t})$ and mgf $m_{\boldsymbol{Y}}(\boldsymbol{t})$ then the marginal cf and mgf for $Y_{i_1}, ..., Y_{i_k}$ are found from the joint cf and mgf by replacing t_{i_j} by 0 for $j = k+1, ..., n$. In particular, if $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)$ and $\boldsymbol{t} = (\boldsymbol{t}_1, \boldsymbol{t}_2)$, then

$$\phi_{\boldsymbol{Y}_1}(\boldsymbol{t}_1) = \phi_{\boldsymbol{Y}}(\boldsymbol{t}_1, \boldsymbol{0}) \text{ and } \mathrm{m}_{\boldsymbol{Y}_1}(\boldsymbol{t}_1) = \mathrm{m}_{\boldsymbol{Y}}(\boldsymbol{t}_1, \boldsymbol{0}).$$

Proof. Use the definition of the cf and mgf. For example, if $\boldsymbol{Y}_1 = (Y_1, ..., Y_k)$ and $\boldsymbol{s} = \boldsymbol{t}_1$, then $m(\boldsymbol{t}_1, \boldsymbol{0}) =$

$E[\exp(t_1Y_1 + \cdots + t_kY_k + 0Y_{k+1} + \cdots + 0Y_n)] = E[\exp(t_1Y_1 + \cdots + t_kY_k)] =$

$E[\exp(\boldsymbol{s}^T\boldsymbol{Y}_1)] = m_{\boldsymbol{Y}_1}(\boldsymbol{s})$, which is the mgf of $\boldsymbol{Y}_1$. QED

Theorem 2.22. Partition the $1 \times n$ vectors $\boldsymbol{Y}$ and $\boldsymbol{t}$ as $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)$ and $\boldsymbol{t} = (\boldsymbol{t}_1, \boldsymbol{t}_2)$. Then the random vectors $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are independent iff their joint cf factors into the product of their marginal cfs:

$$\phi_{\boldsymbol{Y}}(\boldsymbol{t}) = \phi_{\boldsymbol{Y}_1}(\boldsymbol{t}_1)\phi_{\boldsymbol{Y}_2}(\boldsymbol{t}_2) \;\; \forall \boldsymbol{t} \in \Re^n.$$

If the joint mgf exists, then the random vectors $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are independent iff their joint mgf factors into the product of their marginal mgfs:

$$m_{\boldsymbol{Y}}(\boldsymbol{t}) = m_{\boldsymbol{Y}_1}(\boldsymbol{t}_1)m_{\boldsymbol{Y}_2}(\boldsymbol{t}_2)$$

$\forall \boldsymbol{t}$ in some neighborhood of $\boldsymbol{0}$.

2.8 The Multinomial Distribution

Definition 2.26. Assume that there are m iid trials with n outcomes. Let Y_i be the number of the m trials that resulted in the ith outcome and let ρ_i be the probability of the ith outcome for $i = 1, ..., n$ where $0 \le \rho_i \le 1$. Thus $\sum_{i=1}^n Y_i = m$ and $\sum_{i=1}^n \rho_i = 1$. Then $\boldsymbol{Y} = (Y_1, ..., Y_n)$ has a multinomial

$M_n(m, \rho_1, ..., \rho_n)$ distribution if the joint pmf of $\boldsymbol{Y}$ is
$f(y_1, ..., y_n) = P(Y_1 = y_1, ..., Y_n = y_n)$

$$= \frac{m!}{y_1! \cdots y_n!} \rho_1^{y_1} \rho_2^{y_2} \cdots \rho_n^{y_n} = m! \prod_{i=1}^{n} \frac{\rho_i^{y_i}}{y_i!}. \tag{2.27}$$

The support of $\boldsymbol{Y}$ is $\mathcal{Y} = \{\boldsymbol{y} : \sum_{i=1}^n y_i = m \text{ and } 0 \le y_i \le m \text{ for i} = 1, ..., \text{n}\}$.

The **multinomial theorem** states that for real x_i and positive integers m and n,

$$(x_1 + \cdots + x_n)^m = \sum_{\boldsymbol{y} \in \mathcal{Y}} \frac{m!}{y_1! \cdots y_n!} x_1^{y_1} x_2^{y_2} \cdots x_n^{y_n}. \tag{2.28}$$

Taking $x_i = \rho_i$ shows that (2.27) is a pmf.

Since Y_n and ρ_n are known if $Y_1, ..., Y_{n-1}$ and $\rho_1, ..., \rho_{n-1}$ are known, it is convenient to act as if $n-1$ of the outcomes $Y_1, ..., Y_{n-1}$ are important and the nth outcome means that none of the $n-1$ important outcomes occurred. With this reasoning, suppose that $\{i_1, ..., i_{k-1}\} \subset \{1, ..., n\}$. Let $W_j = Y_{i_j}$, and let W_k count the number of times that none of $Y_{i_1}, ..., Y_{i_{k-1}}$ occurred. Then $W_k = m - \sum_{j=1}^{k-1} Y_{i_j}$ and $P(W_k) = 1 - \sum_{j=1}^{k-1} \rho_{i_j}$. Here W_k represents the unimportant outcomes and the joint distribution of $W_1, ..., W_{k-1}, W_k$ is multinomial $M_k(m, \rho_{i_1}, ..., \rho_{i_{k-1}}, 1 - \sum_{j=1}^{k-1} \rho_{i_j})$.

Notice that $\sum_{j=1}^k Y_{i_j}$ counts the number of times that the outcome "one of the outcomes $i_1, ..., i_k$ occurred," an outcome with probability $\sum_{j=1}^k \rho_{i_j}$. Hence $\sum_{j=1}^k Y_{i_j} \sim \text{BIN}(\text{m}, \sum_{\text{j}=1}^{\text{k}} \rho_{\text{i}_\text{j}})$.

Now consider conditional distributions. If it is known that $Y_{i_j} = y_{i_j}$ for $j = k+1, ..., n$, then there are $m - \sum_{j=k+1}^n y_{i_j}$ outcomes left to distribute among $Y_{i_1}, ..., Y_{i_k}$. The conditional probabilities of Y_i remains proportional to ρ_i, but the conditional probabilities must sum to one. Hence the conditional distribution is again multinomial. These results prove the following theorem.

Theorem 2.23. Assume that $(Y_1, ..., Y_n)$ has an $M_n(m, \rho_1, ..., \rho_n)$ distribution and that $\{i_1, ..., i_k\} \subset \{1, ..., n\}$ with $k < n$ and $1 \le i_1 < i_2 < \cdots < i_k \le n$.

a) $(Y_{i_1}, ..., Y_{i_{k-1}}, m - \sum_{j=1}^{k-1} Y_{i_j})$ has an $M_k(m, \rho_{i_1}, ..., \rho_{i_{k-1}}, 1 - \sum_{j=1}^{k-1} \rho_{i_j})$ distribution.

b) $\sum_{j=1}^k Y_{i_j} \sim \text{BIN}(\text{m}, \sum_{\text{j}=1}^{\text{k}} \rho_{\text{i}_\text{j}})$. In particular, $Y_i \sim \text{BIN}(\text{m}, \rho_\text{i})$.

c) Suppose that $0 \leq y_{i_j} < m$ for $j = k+1, ..., n$ and that $0 \leq \sum_{j=k+1}^n y_{i_j} < m$. Let $t = m - \sum_{j=k+1}^n y_{i_j}$ and let $\pi_{i_j} = \rho_{i_j} / \sum_{j=1}^k \rho_{i_j}$ for $j = 1, ..., k$. Then the conditional distribution of $Y_{i_1}, ..., Y_{i_k} | Y_{i_{k+1}} = y_{i_{k+1}}, ..., Y_{i_n} = y_{i_n}$ is the $M_k(t, \pi_{i_1}, ..., \pi_{i_k})$ distribution. The support of this conditional distribution is $\{(y_{i_1}, ..., y_{i_k}) : \sum_{j=1}^k y_{i_j} = t, \text{ and } 0 \leq y_{i_j} \leq t \text{ for } \text{j} = 1, ..., \text{k}\}$.

Theorem 2.24. Assume that $(Y_1, ..., Y_n)$ has an $M_n(m, \rho_1, ..., \rho_n)$ distribution. Then the mgf is

$$m(\boldsymbol{t}) = (\rho_1 e^{t_1} + \cdots + \rho_n e^{t_n})^m, \tag{2.29}$$

$E(Y_i) = m\rho_i$, $\text{VAR}(Y_i) = m\rho_i(1 - \rho_i)$ and $\text{Cov}(Y_i, Y_j) = -m\rho_i\rho_j$ for $i \neq j$.

Proof. $E(Y_i)$ and $V(Y_i)$ follow from Theorem 2.23b, and $m(\boldsymbol{t}) =$

$$E[\exp(t_1 Y_1 + \cdots + t_n Y_n)] = \sum_{\mathcal{Y}} \exp(t_1 y_1 + \cdots + t_n y_n) \frac{m!}{y_1! \cdots y_n!} \rho_1^{y_1} \rho_2^{y_2} \cdots \rho_n^{y_n}$$

$$= \sum_{\mathcal{Y}} \frac{m!}{y_1! \cdots y_n!} (\rho_1 e^{t_1})^{y_1} \cdots (\rho_n e^{t_n})^{y_n} = (\rho_1 e^{t_1} + \cdots + \rho_n e^{t_n})^m$$

by the multinomial theorem (2.28). By Theorem 2.20,

$$E(Y_i Y_j) = \frac{\partial^2}{\partial t_i \partial t_j} (\rho_1 e^{t_1} + \cdots + \rho_n e^{t_n})^m \bigg|_{\boldsymbol{t}=\boldsymbol{0}} =$$

$$\frac{\partial}{\partial t_j} m(\rho_1 e^{t_1} + \cdots + \rho_n e^{t_n})^{m-1} \rho_i e^{t_i} \bigg|_{\boldsymbol{t}=\boldsymbol{0}} =$$

$$m(m-1)(\rho_1 e^{t_1} + \cdots + \rho_n e^{t_n})^{m-2} \rho_i e^{t_i} \rho_j e^{t_j} \bigg|_{\boldsymbol{t}=\boldsymbol{0}} = m(m-1)\rho_i \rho_j.$$

Hence $\text{Cov}(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)E(Y_j) = m(m-1)\rho_i\rho_j - m\rho_i m\rho_j = -m\rho_i\rho_j$. QED

2.9 The Multivariate Normal Distribution

Definition 2.27: Rao (1965, p. 437). A $p \times 1$ random vector $\boldsymbol{X}$ has a $p-$dimensional *multivariate normal distribution* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ iff $\boldsymbol{t}^T \boldsymbol{X}$ has a univariate normal distribution for any $p \times 1$ vector $\boldsymbol{t}$.

If $\boldsymbol{\Sigma}$ is positive definite, then $\boldsymbol{X}$ has a joint pdf

$$f(\boldsymbol{z}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-(1/2)(\boldsymbol{z}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})} \tag{2.30}$$

where $|\boldsymbol{\Sigma}|^{1/2}$ is the square root of the determinant of $\boldsymbol{\Sigma}$. Note that if $p = 1$, then the quadratic form in the exponent is $(z-\mu)(\sigma^2)^{-1}(z-\mu)$ and X has the univariate $N(\mu, \sigma^2)$ pdf. If $\boldsymbol{\Sigma}$ is positive semidefinite but not positive definite, then $\boldsymbol{X}$ has a degenerate distribution. For example, the univariate $N(0, 0^2)$ distribution is degenerate (the point mass at 0).

Some important properties of MVN distributions are given in the following three propositions. These propositions can be proved using results from Johnson and Wichern (1988, p. 127-132).

Proposition 2.25. a) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and

$$\text{Cov}(\boldsymbol{X}) = \boldsymbol{\Sigma}.$$

b) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then any linear combination $\boldsymbol{t}^T\boldsymbol{X} = t_1X_1 + \cdots + t_pX_p \sim N_1(\boldsymbol{t}^T\boldsymbol{\mu}, \boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$. Conversely, if $\boldsymbol{t}^T\boldsymbol{X} \sim N_1(\boldsymbol{t}^T\boldsymbol{\mu}, \boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$ for every $p \times 1$ vector $\boldsymbol{t}$, then $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

c) **The joint distribution of independent normal random variables is MVN.** If $X_1, ..., X_p$ are independent univariate normal $N(\mu_i, \sigma_i^2)$ random vectors, then $\boldsymbol{X} = (X_1, ..., X_p)^T$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = (\mu_1, ..., \mu_p)$ and $\boldsymbol{\Sigma} = diag(\sigma_1^2, ..., \sigma_p^2)$ (so the off diagonal entries $\sigma_{i,j} = 0$ while the diagonal entries of $\boldsymbol{\Sigma}$ are $\sigma_{i,i} = \sigma_i^2$.)

d) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if $\boldsymbol{A}$ is a $q \times p$ matrix, then $\boldsymbol{AX} \sim N_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T)$. If $\boldsymbol{a}$ is a $p \times 1$ vector of constants, then $\boldsymbol{a} + \boldsymbol{X} \sim N_p(\boldsymbol{a} + \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

It will be useful to partition $\boldsymbol{X}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$. Let $\boldsymbol{X}_1$ and $\boldsymbol{\mu}_1$ be $q \times 1$ vectors, let $\boldsymbol{X}_2$ and $\boldsymbol{\mu}_2$ be $(p-q) \times 1$ vectors, let $\boldsymbol{\Sigma}_{11}$ be a $q \times q$ matrix, let $\boldsymbol{\Sigma}_{12}$ be a $q \times (p-q)$ matrix, let $\boldsymbol{\Sigma}_{21}$ be a $(p-q) \times q$ matrix, and let $\boldsymbol{\Sigma}_{22}$ be a $(p-q) \times (p-q)$ matrix. Then

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \ \text{and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Proposition 2.26. a) **All subsets of a MVN are MVN:** $(X_{k_1}, ..., X_{k_q})^T \sim N_q(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ where $\tilde{\boldsymbol{\mu}}_i = E(X_{k_i})$ and $\tilde{\boldsymbol{\Sigma}}_{ij} = \text{Cov}(X_{k_i}, X_{k_j})$. In particular, $\boldsymbol{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{X}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$.

b) If $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent, then $\text{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{\Sigma}_{12} = E[(\boldsymbol{X}_1 - E(\boldsymbol{X}_1))(\boldsymbol{X}_2 - E(\boldsymbol{X}_2))^T] = \boldsymbol{0}$, a $q \times (p-q)$ matrix of zeroes.

c) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent iff $\boldsymbol{\Sigma}_{12} = \boldsymbol{0}$.

d) If $\boldsymbol{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ and $\boldsymbol{X}_2 \sim N_{p-q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ are independent, then

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N_p\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right).$$

Proposition 2.27. The conditional distribution of a MVN is MVN. If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the conditional distribution of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is multivariate normal with mean $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and covariance matrix $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. That is,

$$\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Example 2.23. Let $p = 2$ and let $(Y, X)^T$ have a bivariate normal distribution. That is,

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \text{Cov}(Y, X) \\ \text{Cov}(X, Y) & \sigma_X^2 \end{pmatrix} \right).$$

Also recall that the population correlation between X and Y is given by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{VAR}(X)}\sqrt{\text{VAR}(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

if $\sigma_X > 0$ and $\sigma_Y > 0$. Then $Y|X = x \sim N(E(Y|X = x), \text{VAR}(Y|X = x))$ where the conditional mean

$$E(Y|X = x) = \mu_Y + \text{Cov}(Y, X)\frac{1}{\sigma_X^2}(x - \mu_X) = \mu_Y + \rho(X, Y)\sqrt{\frac{\sigma_Y^2}{\sigma_X^2}}(x - \mu_X)$$

and the conditional variance

$$\text{VAR}(Y|X = x) = \sigma_Y^2 - \text{Cov}(X, Y)\frac{1}{\sigma_X^2}\text{Cov}(X, Y)$$

$$= \sigma_Y^2 - \rho(X, Y)\sqrt{\frac{\sigma_Y^2}{\sigma_X^2}}\rho(X, Y)\sqrt{\sigma_X^2}\sqrt{\sigma_Y^2}$$

$$= \sigma_Y^2 - \rho^2(X,Y)\sigma_Y^2 = \sigma_Y^2[1 - \rho^2(X,Y)].$$

Also $aX + bY$ is univariate normal with mean $a\mu_X + b\mu_Y$ and variance

$$a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab \text{ Cov}(X,Y).$$

Remark 2.1. There are several common misconceptions. First, **it is not true that every linear combination $\boldsymbol{t}^T\boldsymbol{X}$ of normal random variables is a normal random variable,** and **it is not true that all uncorrelated normal random variables are independent**. The key condition in Proposition 2.25b and Proposition 2.26c is that the joint distribution of $\boldsymbol{X}$ is MVN. It is possible that $X_1, X_2, ..., X_p$ each has a marginal distribution that is univariate normal, but the joint distribution of $\boldsymbol{X}$ is not MVN. Examine the following example from Rohatgi (1976, p. 229). Suppose that the joint pdf of X and Y is a mixture of two bivariate normal distributions both with $EX = EY = 0$ and VAR(X) = VAR(Y) = 1, but Cov$(X,Y) = \pm\rho$. Hence $f(x,y) =$

$$\frac{1}{2}\ \frac{1}{2\pi\sqrt{1-\rho^2}} \exp(\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)) \ +$$

$$\frac{1}{2}\ \frac{1}{2\pi\sqrt{1-\rho^2}} \exp(\frac{-1}{2(1-\rho^2)}(x^2 + 2\rho xy + y^2)) \equiv \frac{1}{2}f_1(x,y) + \frac{1}{2}f_2(x,y)$$

where x and y are real and $0 < \rho < 1$. Since both marginal distributions of $f_i(x,y)$ are N(0,1) for $i = 1$ and 2 by Proposition 2.26a, the marginal distributions of X and Y are N(0,1). Since $\int\int xyf_i(x,y)dxdy = \rho$ for $i = 1$ and $-\rho$ for $i = 2$, X and Y are uncorrelated, but X and Y are not independent since $f(x,y) \neq f_X(x)f_Y(y)$.

Remark 2.2. In Proposition 2.27, suppose that $\boldsymbol{X} = (Y, X_2, ..., X_p)^T$. Let $X_1 = Y$ and $\boldsymbol{X}_2 = (X_2, ..., X_p)^T$. Then $E[Y|\boldsymbol{X}_2] = \beta_1 + \beta_2 X_2 + \cdots + \beta_p X_p$ and VAR$[Y|\boldsymbol{X}_2]$ is a constant that does not depend on $\boldsymbol{X}_2$. Hence $Y|\boldsymbol{X}_2 = \beta_1 + \beta_2 X_2 + \cdots + \beta_p X_p + e$ follows the multiple linear regression model.

2.10 Elliptically Contoured Distributions

Definition 2.28: Johnson (1987, p. 107-108). A $p \times 1$ random vector has an *elliptically contoured distribution,* also called an *elliptically symmetric distribution,* if $\boldsymbol{X}$ has joint pdf

$$f(\boldsymbol{z}) = k_p|\boldsymbol{\Sigma}|^{-1/2}g[(\boldsymbol{z}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})], \tag{2.31}$$

and we say $\boldsymbol{X}$ has an elliptically contoured $EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ distribution.

If $\boldsymbol{X}$ has an elliptically contoured (EC) distribution, then the characteristic function of $\boldsymbol{X}$ is

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(i\boldsymbol{t}^T\boldsymbol{\mu})\psi(\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}) \tag{2.32}$$

for some function ψ. If the second moments exist, then

$$E(\boldsymbol{X}) = \boldsymbol{\mu} \tag{2.33}$$

and

$$\text{Cov}(\boldsymbol{X}) = c_X\boldsymbol{\Sigma} \tag{2.34}$$

where

$$c_X = -2\psi'(0).$$

Definition 2.29. The *population squared Mahalanobis distance*

$$U \equiv D^2 = D^2(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{X} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \tag{2.35}$$

has density

$$h(u) = \frac{\pi^{p/2}}{\Gamma(p/2)}k_p u^{p/2-1}g(u). \tag{2.36}$$

For $c > 0$, an $EC_p(\boldsymbol{\mu}, c\boldsymbol{I}, g)$ distribution is *spherical about* $\boldsymbol{\mu}$ where $\boldsymbol{I}$ is the $p \times p$ identity matrix. The *multivariate normal distribution* $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has $k_p = (2\pi)^{-p/2}$, $\psi(u) = g(u) = \exp(-u/2)$, and $h(u)$ is the χ^2_p density.

The following lemma is useful for proving properties of EC distributions without using the characteristic function (2.32). See Eaton (1986) and Cook (1998, p. 57, 130).

Lemma 2.28. Let $\boldsymbol{X}$ be a $p \times 1$ random vector with 1st moments; ie, $E(\boldsymbol{X})$ exists. Let $\boldsymbol{B}$ be any constant full rank $p \times r$ matrix where $1 \leq r \leq p$. Then $\boldsymbol{X}$ is elliptically contoured iff for all such conforming matrices $\boldsymbol{B}$,

$$E(\boldsymbol{X}|\boldsymbol{B}^T\ \boldsymbol{X}) = \boldsymbol{\mu} + \boldsymbol{M}_B\boldsymbol{B}^T(\boldsymbol{X} - \boldsymbol{\mu}) = \boldsymbol{a}_B + \boldsymbol{M}_B\boldsymbol{B}^T\boldsymbol{X} \tag{2.37}$$

where the $p \times 1$ constant vector $\boldsymbol{a}_B$ and the $p \times r$ constant matrix $\boldsymbol{M}_B$ both depend on $\boldsymbol{B}$.

A useful fact is that $\boldsymbol{a}_B$ and $\boldsymbol{M}_B$ do not depend on g:

$$\boldsymbol{a}_B = \boldsymbol{\mu} - \boldsymbol{M}_B \boldsymbol{B}^T \boldsymbol{\mu} = (\boldsymbol{I}_p - \boldsymbol{M}_B \boldsymbol{B}^T)\boldsymbol{\mu},$$

and

$$\boldsymbol{M}_B = \boldsymbol{\Sigma} \boldsymbol{B}(\boldsymbol{B}^T \boldsymbol{\Sigma} \boldsymbol{B})^{-1}.$$

Notice that in the formula for $\boldsymbol{M}_B$, $\boldsymbol{\Sigma}$ can be replaced by $c\boldsymbol{\Sigma}$ where $c > 0$ is a constant. In particular, if the EC distribution has second moments, $\text{Cov}(\boldsymbol{X})$ can be used instead of $\boldsymbol{\Sigma}$.

To use Lemma 2.28 to prove interesting properties, partition $\boldsymbol{X}$, $\boldsymbol{\mu}$, and $\boldsymbol{\Sigma}$. Let $\boldsymbol{X}_1$ and $\boldsymbol{\mu}_1$ be $q \times 1$ vectors, let $\boldsymbol{X}_2$ and $\boldsymbol{\mu}_2$ be $(p-q) \times 1$ vectors. Let $\boldsymbol{\Sigma}_{11}$ be a $q \times q$ matrix, let $\boldsymbol{\Sigma}_{12}$ be a $q \times (p-q)$ matrix, let $\boldsymbol{\Sigma}_{21}$ be a $(p-q) \times q$ matrix, and let $\boldsymbol{\Sigma}_{22}$ be a $(p-q) \times (p-q)$ matrix. Then

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Also assume that the $(p+1) \times 1$ vector $(Y, \boldsymbol{X}^T)^T$ is $EC_{p+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ where Y is a random variable, $\boldsymbol{X}$ is a $p \times 1$ vector, and use

$$\begin{pmatrix} Y \\ \boldsymbol{X} \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \mu_Y \\ \boldsymbol{\mu}_X \end{pmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & \boldsymbol{\Sigma}_{XX} \end{pmatrix}.$$

Proposition 2.29. Let $\boldsymbol{X} \sim EC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ and assume that $E(\boldsymbol{X})$ exists.

a) Any subset of $\boldsymbol{X}$ is EC, in particular $\boldsymbol{X}_1$ is EC.

b) (Cook 1998 p. 131, Kelker 1970). If $\text{Cov}(\boldsymbol{X})$ is nonsingular,

$$\text{Cov}(\boldsymbol{X}|\boldsymbol{B}^T\boldsymbol{X}) = d_g(\boldsymbol{B}^T\boldsymbol{X})[\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{\Sigma}\boldsymbol{B})^{-1}\boldsymbol{B}^T\boldsymbol{\Sigma}]$$

where the real valued function $d_g(\boldsymbol{B}^T\boldsymbol{X})$ is constant iff $\boldsymbol{X}$ is MVN.

Proof of a). Let $\boldsymbol{A}$ be an arbitrary full rank $q \times r$ matrix where $1 \leq r \leq q$. Let

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{0} \end{pmatrix}.$$

Then $\boldsymbol{B}^T\boldsymbol{X} = \boldsymbol{A}^T\boldsymbol{X}_1$, and

$$E[\boldsymbol{X}|\boldsymbol{B}^T\boldsymbol{X}] = E[\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} |\boldsymbol{A}^T\boldsymbol{X}_1] =$$

$$\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{M}_{1B} \\ \boldsymbol{M}_{2B} \end{pmatrix} \begin{pmatrix} \boldsymbol{A}^T & \boldsymbol{0}^T \end{pmatrix} \begin{pmatrix} \boldsymbol{X}_1 - \boldsymbol{\mu}_1 \\ \boldsymbol{X}_2 - \boldsymbol{\mu}_2 \end{pmatrix}$$

by Lemma 2.28. Hence $E[\boldsymbol{X}_1|\boldsymbol{A}^T\boldsymbol{X}_1] = \boldsymbol{\mu}_1 + \boldsymbol{M}_{1B}\boldsymbol{A}^T(\boldsymbol{X}_1 - \boldsymbol{\mu}_1)$. Since $\boldsymbol{A}$ was arbitrary, $\boldsymbol{X}_1$ is EC by Lemma 2.28. Notice that $\boldsymbol{M}_B = \boldsymbol{\Sigma}\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{\Sigma}\boldsymbol{B})^{-1} =$

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{0} \end{pmatrix} [\begin{pmatrix} \boldsymbol{A}^T & \boldsymbol{0}^T \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{0} \end{pmatrix}]^{-1}$$

$$= \begin{pmatrix} \boldsymbol{M}_{1B} \\ \boldsymbol{M}_{2B} \end{pmatrix}.$$

Hence

$$\boldsymbol{M}_{1B} = \boldsymbol{\Sigma}_{11}\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{\Sigma}_{11}\boldsymbol{A})^{-1}$$

and $\boldsymbol{X}_1$ is EC with location and dispersion parameters $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_{11}$. QED

Proposition 2.30. Let $(Y, \boldsymbol{X}^T)^T$ be $EC_{p+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ where Y is a random variable.

a) Assume that $E[(Y, \boldsymbol{X}^T)^T]$ exists. Then $E(Y|\boldsymbol{X}) = \alpha + \boldsymbol{\beta}^T\boldsymbol{X}$ where $\alpha = \mu_Y - \boldsymbol{\beta}^T\boldsymbol{\mu}_X$ and

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{XY}.$$

b) Even if the first moment does not exist, the conditional median

$$\text{MED}(Y|\boldsymbol{X}) = \alpha + \boldsymbol{\beta}^T\boldsymbol{X}$$

where α and $\boldsymbol{\beta}$ are given in a).

Proof. a) The trick is to choose $\boldsymbol{B}$ so that Lemma 2.28 applies. Let

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{0}^T \\ \boldsymbol{I}_p \end{pmatrix}.$$

Then $\boldsymbol{B}^T\boldsymbol{\Sigma}\boldsymbol{B} = \boldsymbol{\Sigma}_{XX}$ and

$$\boldsymbol{\Sigma}\boldsymbol{B} = \begin{pmatrix} \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XX} \end{pmatrix}.$$

Now

$$E[\begin{pmatrix} Y \\ \boldsymbol{X} \end{pmatrix} \mid \boldsymbol{X}] = E[\begin{pmatrix} Y \\ \boldsymbol{X} \end{pmatrix} \mid \boldsymbol{B}^T \begin{pmatrix} Y \\ \boldsymbol{X} \end{pmatrix}]$$

$$= \boldsymbol{\mu} + \boldsymbol{\Sigma B}(\boldsymbol{B}^T\boldsymbol{\Sigma B})^{-1}\boldsymbol{B}^T \begin{pmatrix} Y - \mu_Y \\ \boldsymbol{X} - \boldsymbol{\mu}_X \end{pmatrix}$$

by Lemma 2.28. The right hand side of the last equation is equal to

$$\boldsymbol{\mu} + \begin{pmatrix} \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XX} \end{pmatrix} \boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X) = \begin{pmatrix} \mu_Y - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\mu}_X + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{X} \\ \boldsymbol{X} \end{pmatrix}$$

and the result follows since

$$\boldsymbol{\beta}^T = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}.$$

b) See Croux, Dehon, Rousseeuw and Van Aelst (2001) for references.

Example 2.24. This example illustrates another application of Lemma 2.28. Suppose that $\boldsymbol{X}$ comes from a mixture of two multivariate normals with the same mean and proportional covariance matrices. That is, let

$$\boldsymbol{X} \sim (1-\gamma)N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + \gamma N_p(\boldsymbol{\mu}, c\boldsymbol{\Sigma})$$

where $c > 0$ and $0 < \gamma < 1$. Since the multivariate normal distribution is elliptically contoured (and see Proposition 1.14c),

$$E(\boldsymbol{X}|\boldsymbol{B}^T\boldsymbol{X}) = (1-\gamma)[\boldsymbol{\mu} + \boldsymbol{M}_1\boldsymbol{B}^T(\boldsymbol{X} - \boldsymbol{\mu})] + \gamma[\boldsymbol{\mu} + \boldsymbol{M}_2\boldsymbol{B}^T(\boldsymbol{X} - \boldsymbol{\mu})]$$

$$= \boldsymbol{\mu} + [(1-\gamma)\boldsymbol{M}_1 + \gamma\boldsymbol{M}_2]\boldsymbol{B}^T(\boldsymbol{X} - \boldsymbol{\mu}) \equiv \boldsymbol{\mu} + \boldsymbol{M}\boldsymbol{B}^T(\boldsymbol{X} - \boldsymbol{\mu}).$$

Since $\boldsymbol{M}_B$ only depends on $\boldsymbol{B}$ and $\boldsymbol{\Sigma}$, it follows that $\boldsymbol{M}_1 = \boldsymbol{M}_2 = \boldsymbol{M} = \boldsymbol{M}_B$. Hence $\boldsymbol{X}$ has an elliptically contoured distribution by Lemma 2.28.

2.11 Summary

1) Y_1 and Y_2 are dependent if the support $\mathcal{Y} = \{(y_1, y_2)|f(y_1, y_2) > 0\}$ is not a cross product.

2) If the support is a cross product, then Y_1 and Y_2 are independent iff $f(y_1, y_2) = h_1(y_1)h_2(y_2)$ for all $(y_1, y_2) \in \mathcal{Y}$ where $h_i(y_i)$ is a positive function of y_i alone. If no such factorization exists, then Y_1 and Y_2 are dependent.

3) If $Y_1, ..., Y_n$ are independent, then the functions $h_1(Y_1), ..., h_n(Y_n)$ are independent.

4) Given $f(y_1, y_2)$, find $E[h(Y_i)]$ by finding the marginal pdf or pmf $f_{Y_i}(y_i)$ and using the marginal distribution in the expectation.

5) $E[Y] = E[E(Y|X)]$ and $V(Y) = E[V(Y|X)] + V[E(Y|X)]$.

6) Find the pmf of $Y = t(X)$ and the (sample space =) support $\mathcal{Y}$ given the pmf of X by collecting terms $x : y = t(x)$.

7) For increasing or decreasing t, the pdf of $Y = t(X)$ is

$$f_Y(y) = f_X(t^{-1}(y)) \left| \frac{dt^{-1}(y)}{dy} \right|$$

for $y \in \mathcal{Y}$. Also be able to find the support $\mathcal{Y}$.

8) Find the joint pdf of $Y_1 = t_1(X_1, X_2)$ and $Y_2 = t_2(X_1, X_2)$: $f_{Y_1,Y_2}(y_1, y_2)$ $= f_{X_1,X_2}(t_1^{-1}(y_1, y_2), t_2^{-1}(y_1, y_2))|J|$. Finding the support $\mathcal{Y}$ is crucial. Using indicator functions can help. Know that $\prod_{j=1}^k I_{A_j}(\boldsymbol{y}) = I_{\cap_{j=1}^k A_j}(\boldsymbol{y})$. The Jacobian of the bivariate transformation is

$$J = \det \begin{bmatrix} \frac{\partial t_1^{-1}}{\partial y_1} & \frac{\partial t_1^{-1}}{\partial y_2} \\ \\ \frac{\partial t_2^{-1}}{\partial y_1} & \frac{\partial t_2^{-1}}{\partial y_2} \end{bmatrix},$$

and $|J|$ is the absolute value of the determinant J. Recall that

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \text{ad} - \text{bc}.$$

To find $t_i^{-1}(y_1, y_2)$, use $y_i = t_i(x_1, x_2)$ and solve for x_1 and x_2 where $i = 1, 2$.

9) If $Y_1, ..., Y_n$ are independent with mgfs $m_{Y_i}(t)$, then the mgf of $W = \sum_{i=1}^n Y_i$ is

$$m_W(t) = \prod_{i=1}^n m_{Y_i}(t).$$

10) If $Y_1, ..., Y_n$ are iid with mgf $m_Y(t)$, then the mgf of $W = \sum_{i=1}^n Y_i$ is

$$m_W(t) = [m_Y(t)]^n,$$

and the mgf of $\overline{Y}$ is

$$m_{\overline{Y}}(t) = [m_Y(t/n)]^n.$$

11) Know that if $Y_1, ..., Y_n$ are iid with $E(Y) = \mu$ and $\text{V}(Y) = \sigma^2$, then $E(\overline{Y}) = \mu$ and $\text{V}(\overline{Y}) = \sigma^2/n$.

12) Suppose $W = \sum_{i=1}^n Y_i$ or $W = \overline{Y}$ where $Y_1, ..., Y_n$ are independent. For several distributions (especially Y_i iid gamma(ν, λ) and Y_i independent

$N(\mu_i, \sigma_i^2)$), be able to find the distribution of W, the mgf of W, $E(W)$, $\text{Var}(W)$, and $E(W^2) = V(W) + [E(W)]^2$.

13) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{t}^T\boldsymbol{X} = t_1X_1 + ... + t_pX_p \sim N(\boldsymbol{t}^T\boldsymbol{\mu}, \boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$.

14) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if $\boldsymbol{A}$ is a $q \times p$ matrix, then $\boldsymbol{AX} \sim N_q(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T)$. If $\boldsymbol{a}$ is a $p \times 1$ vector of constants, then $\boldsymbol{a} + \boldsymbol{X} \sim N_p(\boldsymbol{a} + \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Suppose $\boldsymbol{X}_1$ is $q \times 1$ and

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

15) $\boldsymbol{X}_1 \sim N_q(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$.

16) If $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the conditional distribution of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is multivariate normal with mean $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and covariance $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$. That is,

$$\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim N_q(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

17)

$$\rho(X_i, X_j) = \frac{\sigma_{i,j}}{\sqrt{\sigma_{ii}\sigma_{jj}}} = \mathrm{Cov}(\mathrm{X_i}, \mathrm{X_j})/\sqrt{\mathrm{V}(\mathrm{X_i})\mathrm{V}(\mathrm{X_j})}.$$

18) Know that (X, Y) can have a joint distribution that is not multivariate normal, yet the marginal distributions of X and Y are both univariate normal. Hence X and Y can be normal, but $aX + bY$ is not normal. (Need the joint distribution of (X, Y) to be MVN for all linear combinations to be univariate normal.)

2.12 Complements

Panjer (1969) provides generalizations of Steiner's formula.

Johnson and Wichern (1988), Mardia, Kent and Bibby (1979) and Press (2005) are good references for multivariate statistical analysis based on the multivariate normal distribution. The elliptically contoured distributions generalize the multivariate normal distribution and are discussed (in increasing order of difficulty) in Johnson (1987), Fang, Kotz, and Ng (1990), Fang and Anderson (1990), and Gupta and Varga (1993). Fang, Kotz, and Ng (1990) sketch the history of elliptically contoured distributions while Gupta and Varga (1993) discuss matrix valued elliptically contoured distributions.

Cambanis, Huang, and Simons (1981), Chmielewski (1981) and Eaton (1986) are also important references. Also see Muirhead (1982, p. 30–42).

Broffitt (1986), Kowalski (1973), Melnick and Tenebien (1982) and Seber and Lee (2003, p. 23) give examples of dependent marginally normal random variables that have 0 correlation. The example in Remark 2.1 appears in Rohatgi (1976, p. 229) and Lancaster (1959).

See Abuhassan (2007) for more information about the distributions in problems 2.52–2.59.

2.13 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

Refer to Chapter 10 for the pdf or pmf of the distributions in the problems below.

Theorem 2.16 is useful for Problems 2.1–2.7.

2.1*. Let $X_1, ..., X_n$ be independent Poisson(λ_i). Let $W = \sum_{i=1}^n X_i$. Find the mgf of W and find the distribution of W.

2.2*. Let $X_1, ..., X_n$ be iid Bernoulli(ρ). Let $W = \sum_{i=1}^n X_i$. Find the mgf of W and find the distribution of W.

2.3*. Let $X_1, ..., X_n$ be iid exponential (λ). Let $W = \sum_{i=1}^n X_i$. Find the mgf of W and find the distribution of W.

2.4*. Let $X_1, ..., X_n$ be independent $N(\mu_i, \sigma_i^2)$. Let $W = \sum_{i=1}^n (a_i + b_i X_i)$ where a_i and b_i are fixed constants. Find the mgf of W and find the distribution of W.

2.5*. Let $X_1, ..., X_n$ be iid negative binomial (1, ρ). Let $W = \sum_{i=1}^n X_i$. Find the mgf of W and find the distribution of W.

2.6*. Let $X_1, ..., X_n$ be independent gamma (ν_i, λ). Let $W = \sum_{i=1}^n X_i$. Find the mgf of W and find the distribution of W.

2.7*. Let $X_1, ..., X_n$ be independent $\chi^2_{p_i}$. Let $W = \sum_{i=1}^n X_i$. Find the mgf of W and find the distribution of W.

2.8. a) Let $f_Y(y)$ be the pdf of Y. If $W = \mu + Y$ where $-\infty < \mu < \infty$, show that the pdf of W is $f_W(w) = f_Y(w - \mu)$.

b) Let $f_Y(y)$ be the pdf of Y. If $W = \sigma Y$ where $\sigma > 0$, show that the pdf of W is $f_W(w) = (1/\sigma) f_Y(w/\sigma)$.

c) Let $f_Y(y)$ be the pdf of Y. If $W = \mu + \sigma Y$ where $-\infty < \mu < \infty$ and $\sigma > 0$, show that the pdf of W is $f_W(w) = (1/\sigma) f_Y((w - \mu)/\sigma)$.

2.9. a) If Y is lognormal $LN(\mu, \sigma^2)$, show that $W = \log(Y)$ is a normal $N(\mu, \sigma^2)$ random variable.

b) If Y is a normal $N(\mu, \sigma^2)$ random variable, show that $W = e^Y$ is a lognormal $LN(\mu, \sigma^2)$ random variable.

2.10. a) If Y is uniform (0,1), Show that $W = -\log(Y)$ is exponential (1).

b) If Y is exponential (1), show that $W = \exp(-Y)$ is uniform (0,1).

2.11. If $Y \sim N(\mu, \sigma^2)$, find the pdf of

$$W = \left(\frac{Y - \mu}{\sigma}\right)^2.$$

2.12. If Y has a half normal distribution, $Y \sim \text{HN}(\mu, \sigma^2)$, show that $W = (Y - \mu)^2 \sim G(1/2, 2\sigma^2)$.

2.13. a) Suppose that Y has a Weibull (ϕ, λ) distribution with pdf

$$f(y) = \frac{\phi}{\lambda} y^{\phi - 1} e^{-\frac{y^\phi}{\lambda}}$$

where $\lambda, y,$ and ϕ are all positive. Show that $W = \log(Y)$ has a smallest extreme value SEV$(\theta = \log(\lambda^{1/\phi}), \sigma = 1/\phi)$ distribution.

b) If Y has a SEV$(\theta = \log(\lambda^{1/\phi}), \sigma = 1/\phi)$ distribution, show that $W = e^Y$ has a Weibull (ϕ, λ) distribution.

2.14. a) Suppose that Y has a Pareto(σ, λ) distribution with pdf

$$f(y) = \frac{\frac{1}{\lambda}\sigma^{1/\lambda}}{y^{1+1/\lambda}}$$

where $y \geq \sigma$, $\sigma > 0$, and $\lambda > 0$. Show that $W = \log(Y) \sim EXP(\theta = \log(\sigma), \lambda)$.

b) If Y as an $EXP(\theta = \log(\sigma), \lambda)$ distribution, show that $W = e^Y$ has a Pareto(σ, λ) distribution.

2.15. a) If Y is chi χ_p, then the pdf of Y is

$$f(y) = \frac{y^{p-1} e^{-y^2/2}}{2^{\frac{p}{2}-1}\Gamma(p/2)}$$

where $y \geq 0$ and p is a positive integer. Show that the pdf of $W = Y^2$ is the χ_p^2 pdf.

b) If Y is a chi–square χ_p^2 random variable, show that $W = \sqrt{Y}$ is a chi χ_p random variable.

2.16. a) If Y is power $POW(\lambda)$, then the pdf of Y is

$$f(y) = \frac{1}{\lambda} y^{\frac{1}{\lambda}-1},$$

where $\lambda > 0$ and $0 < y < 1$. Show that $W = -\log(Y)$ is an exponential (λ) random variable.

b) If Y is an exponential(λ) random variable, show that $W = e^{-Y}$ is a power $POW(\lambda)$ random variable.

2.17. a) If Y is truncated extreme value $TEV(\lambda)$ then the pdf of Y is

$$f(y) = \frac{1}{\lambda} \exp(y - \frac{e^y - 1}{\lambda})$$

where $y > 0$, and $\lambda > 0$. Show that $W = e^Y - 1$ is an exponential (λ) random variable.

b) If Y is an exponential(λ) random variable, show that $W = \log(Y + 1)$ is a truncated extreme value $TEV(\lambda)$ random variable.

c) If Y has an inverse exponential distribution, $Y \sim$ IEXP(θ), show that $W = 1/Y \sim EXP(1/\theta)$.

d) If Y has an inverse Weibull distribution, $Y \sim IW(\phi, \lambda)$, show that $1/Y \sim W(\phi, \lambda)$, the Weibull distribution with parameters $\phi, \lambda)$.

e) If Y has a log-gamma distribution, $Y \sim LG(\nu, \lambda)$, show that $W = e^Y \sim$ gamma (ν, λ).

f) If Y has a two parameter power distribution, $Y \sim power(\tau, \lambda)$, show that $W = -\log(Y) \sim EXP(-\log(\tau), \lambda)$.

2.18. a) If Y is BurrXII(ϕ, λ), show that $W = \log(1 + Y^\phi)$ is an exponential(λ) random variable.

b) If Y is an exponential(λ) random variable, show that $W = (e^Y - 1)^{1/\phi}$ is a BurrXII(ϕ, λ) random variable.

2.19. a) If Y is Pareto $PAR(\sigma, \lambda)$, show that $W = \log(Y/\sigma)$ is an exponential(λ) random variable.

b) If Y is an exponential(λ) random variable, show that $W = \sigma e^Y$ is a Pareto $PAR(\sigma, \lambda)$ random variable.

2.20. a) If Y is Weibull $W(\phi, \lambda)$, show that $W = Y^\phi$ is an exponential (λ) random variable.

b) If Y is an exponential(λ) random variable, show that $W = Y^{1/\phi}$ is a Weibull $W(\phi, \lambda)$ random variable.

2.21. If Y is double exponential (θ, λ), show that $W = |Y - \theta| \sim \text{EXP}(\lambda)$.

2.22. If Y has a generalized gamma distribution, $Y \sim GG(\nu, \lambda, \phi)$, show that $W = Y^\phi \sim G(\nu, \lambda^\phi)$.

2.23. If Y has an inverted gamma distribution, $Y \sim INVG(\nu, \lambda)$, show that $W = 1/Y \sim G(\nu, \lambda)$.

2.24. a) If Y has a largest extreme value distribution $Y \sim LEV(\theta, \sigma)$, show that $W = \exp(-(Y - \theta)/\sigma) \sim \text{EXP}(1)$.

b) If $Y \sim EXP(1)$, show that $W = \theta - \sigma \log(Y) \sim LEV(\theta, \sigma)$.

2.25. a) If Y has a log–Cauchy distribution , $Y \sim LC(\mu, \sigma)$, show that $W = \log(Y)$ has a Cauchy(μ, σ) distribution.

b) If $Y \sim C(\mu, \sigma)$ show that $W = e^Y \sim LC(\mu, \sigma)$.

2.26. a) If Y has a log–logistic distribution, $Y \sim LL(\phi, \tau)$, show that $W = \log(Y)$ has a logistic $L(\mu = -\log(\phi), \sigma = 1/\tau)$ distribution.

b) If $Y \sim L(\mu = -\log(\phi), \sigma = 1/\tau)$, show that $W = e^Y \sim LL(\phi, \tau)$.

2.27. If Y has a Maxwell–Boltzmann distribution, $Y \sim MB(\mu, \sigma)$, show that $W = (Y - \mu)^2 \sim G(3/2, 2\sigma^2)$.

2.28. If Y has a one sided stable distribution, $Y \sim OSS(\sigma)$, show that $W = 1/Y \sim G(1/2, 2/\sigma)$.

2.29. a) If Y has a Rayleigh distribution, $Y \sim R(\mu, \sigma)$, show that $W = (Y - \mu)^2 \sim \text{EXP}(2\sigma^2)$.

b) If $Y \sim \text{EXP}(2\sigma^2)$, show that $W = \sqrt{Y} + \mu \sim R(\mu, \sigma)$.

2.30. If Y has a smallest extreme value distribution, $Y \sim SEV(\theta, \sigma)$, show that $W = -Y$ has an LEV$(-\theta, \sigma)$ distribution.

2.31. Let $Y \sim C(0, 1)$. Show that the Cauchy distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim C(\mu, \sigma)$ where μ is real and $\sigma > 0$.

2.32. Let Y have a chi distribution, $Y \sim \text{chi(p}, 1)$ where p is known. Show that the chi(p, σ) distribution is a scale family for p known by showing that $W = \sigma Y \sim \text{chi(p}, \sigma)$ for $\sigma > 0$.

2.33. Let $Y \sim DE(0, 1)$. Show that the double exponential distribution is a location–scale family by showing that $W = \theta + \lambda Y \sim DE(\theta, \lambda)$ where θ is real and $\lambda > 0$.

2.34. Let $Y \sim \text{EXP}(1)$. Show that the exponential distribution is a scale family by showing that $W = \lambda Y \sim \text{EXP}(\lambda)$ for $\lambda > 0$.

2.35. Let $Y \sim EXP(0, 1)$. Show that the two parameter exponential distribution is a location–scale family by showing that $W = \theta + \lambda Y \sim EXP(\theta, \lambda)$ where θ is real and $\lambda > 0$.

2.36. Let $Y \sim LEV(0, 1)$. Show that the largest extreme value distribution is a location–scale family by showing that $W = \theta + \sigma Y \sim LEV(\theta, \sigma)$ where θ is real and $\sigma > 0$.

2.37. Let $Y \sim G(\nu, 1)$ where ν is known. Show that the gamma (ν, λ) distribution is a scale family for ν known by showing that $W = \lambda Y \sim G(\nu, \lambda)$ for $\lambda > 0$.

2.38. Let $Y \sim HC(0, 1)$. Show that the half Cauchy distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim HC(\mu, \sigma)$ where μ is real and $\sigma > 0$.

2.39. Let $Y \sim HL(0, 1)$. Show that the half logistic distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim HL(\mu, \sigma)$ where μ is real and $\sigma > 0$.

2.40. Let $Y \sim HN(0, 1)$. Show that the half normal distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim HN(\mu, \sigma^2)$ where μ is real and $\sigma > 0$.

2.41. Let $Y \sim L(0, 1)$. Show that the logistic distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim L(\mu, \sigma)$ where μ is real and $\sigma > 0$.

2.42. Let $Y \sim MB(0, 1)$. Show that the Maxwell–Boltzmann distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim MB(\mu, \sigma)$ where μ is real and $\sigma > 0$.

2.43. Let $Y \sim N(0, 1)$. Show that the normal distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim N(\mu, \sigma^2)$ where μ is real and $\sigma > 0$.

2.44. Let $Y \sim \text{OSS}(1)$. Show that the one sided stable distribution is a scale family by showing that $W = \sigma Y \sim \text{OSS}(\sigma)$ for $\sigma > 0$.

2.45. Let $Y \sim PAR(1, \lambda)$ where λ is known. Show that the Pareto (σ, λ) distribution is a scale family for λ known by showing that $W = \sigma Y \sim PAR(\sigma, \lambda)$ for $\sigma > 0$.

2.46. Let $Y \sim R(0, 1)$. Show that the Rayleigh distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim R(\mu, \sigma)$ where μ is real and $\sigma > 0$.

2.47. Let $Y \sim U(0, 1)$. Show that the uniform distribution is a location–scale family by showing that $W = \mu + \sigma Y \sim U(\theta_1, \theta_2)$ where $\mu = \theta_1$ is real and $\sigma = \theta_2 - \theta_1 > 0$.

2.48. Examine the proof of Theorem 2.2b for a joint pdf and prove the result for a joint pmf by replacing the integrals by appropriate sums.

2.49. Examine the proof of Theorem 2.3 for a joint pdf and prove the result for a joint pmf by replacing the integrals by appropriate sums.

2.50. Examine the proof of Theorem 2.4 for a joint pdf and prove the result for a joint pmf by replacing the integrals by appropriate sums.

2.51. Examine the proof of Theorem 2.5 for a joint pdf and prove the result for a joint pmf by replacing the integrals by appropriate sums.

2.52. If $Y \sim hburr(\phi, \lambda)$, then the pdf of Y is

$$f(y) = \frac{2}{\lambda\sqrt{2\pi}} \frac{\phi y^{\phi-1}}{(1+y^{\phi})} \exp\left(\frac{-[\log(1+y^{\phi})]^2}{2\lambda^2}\right) I(y>0)$$

where ϕ and λ are positive.

a) Show that $W = \log(1+Y^{\phi}) \sim HN(0,\lambda)$, the half normal distribution with parameters 0 and λ.

b) If $W \sim HN(0,\lambda)$, then show $Y = [e^W - 1]^{1/\phi} \sim hburr(\phi,\lambda)$.

2.53. If $Y \sim hlev(\theta, \lambda)$, then the pdf of Y is

$$f(y) = \frac{2}{\lambda\sqrt{2\pi}} \exp(\frac{-(y-\theta)}{\lambda}) \exp\left[-\frac{1}{2}[\exp(\frac{-(y-\theta)}{\lambda})]^2\right]$$

where y and θ are real and $\lambda > 0$.

a) Show that $W = \exp(-(Y-\theta)/\lambda) \sim HN(0,1)$, the half normal distribution with parameters 0 and 1.

b) If $W \sim HN(0,1)$, then show $Y = -\lambda\log(W) + \theta \sim hlev(\theta,\lambda)$.

2.54. If $Y \sim hpar(\theta, \lambda)$, then the pdf of Y is

$$f(y) = \frac{2}{\lambda\sqrt{2\pi}} \frac{1}{y} I[y \geq \theta] \exp\left[\frac{-(\log(y)-\log(\theta))^2}{2\lambda^2}\right]$$

where $\theta > 0$ and $\lambda > 0$.

a) Show that $W = \log(Y) \sim HN(\mu = \log(\theta), \sigma = \lambda)$. (See the half normal distribution in Chapter 10.)

b) If $W \sim HN(\mu,\sigma)$, then show $Y = e^W \sim hpar(\theta = e^{\mu}, \lambda = \sigma)$.

2.55. If $Y \sim hpow(\lambda)$, then the pdf of Y is

$$f(y) = \frac{2}{\lambda\sqrt{2\pi}} \frac{1}{y} I_{[0,1]}(y) \ \exp\left[\frac{-(\log(y))^2}{2\lambda^2}\right]$$

where $\lambda > 0$.

a) Show that $W = -\log(Y) \sim HN(0, \sigma = \lambda)$, the half normal distribution with parameters 0 and λ.

b) If $W \sim HN(0,\sigma)$, then show $Y = e^{-W} \sim hpow(\lambda = \sigma)$.

2.56. If $Y \sim hray(\theta, \lambda)$, then the pdf of Y is

$$f(y) = \frac{4}{\lambda\sqrt{2\pi}}(y-\theta)I[y \geq \theta] \exp\left[\frac{-(y-\theta)^4}{2\lambda^2}\right]$$

where $\lambda > 0$ and θ is real.

a) Show that $W = (Y-\theta)^2 \sim HN(0, \sigma = \lambda)$, the half normal distribution with parameters 0 and λ.

b) If $W \sim HN(0, \sigma)$, then show $Y = \sqrt{W} + \theta \sim hray(\theta, \lambda = \sigma)$.

2.57. If $Y \sim hsev(\theta, \lambda)$, then the pdf of Y is

$$f(y) = \frac{2}{\lambda\sqrt{2\pi}} \exp(\frac{y-\theta}{\lambda}) \exp\left(\frac{-1}{2}\left[\exp(\frac{y-\theta}{\lambda})\right]^2\right)$$

where y and θ are real and $\lambda > 0$.

a) Show that $W = \exp[(y-\theta)/\lambda] \sim HN(0, 1)$.

b) If $W \sim HN(0, 1)$, then show $Y = \lambda \log(W) + \theta \sim hsev(\theta, \lambda)$.

2.58. If $Y \sim htev(\lambda)$, then the pdf of Y is

$$f(y) = \frac{2}{\lambda\sqrt{2\pi}} \exp\left(y - \frac{(e^y-1)^2}{2\lambda^2}\right) = \frac{2}{\lambda\sqrt{2\pi}}\ e^y\ \exp\left(\frac{-(e^y-1)^2}{2\lambda^2}\right)$$

where $y > 0$ and $\lambda > 0$.

a) Show that $W = e^Y - 1 \sim HN(0, \sigma = \lambda)$, the half normal distribution with parameters 0 and λ.

b) If $W \sim HN(0, \sigma)$, then show $Y = \log(W + 1) \sim htev(\lambda = \sigma)$.

2.59. If $Y \sim hweib(\phi, \lambda)$, then the pdf of Y is

$$f(y) = \frac{2}{\lambda\sqrt{2\pi}}\ \phi\ y^{\phi-1}\ I[y > 0]\ \exp\left(\frac{-y^{2\phi}}{2\lambda^2}\right)$$

where λ and ϕ are positive.

a) Show that $W = Y^\phi \sim HN(0, \sigma = \lambda)$, the half normal distribution with parameters 0 and λ.

b) If $W \sim HN(0, \sigma)$, then show $Y = W^{1/\phi} \sim hweib(\phi, \lambda = \sigma)$.

Problems from old quizzes and exams.

2.60. If Y is a random variable with pdf

$$f(y) = \lambda y^{\lambda-1} \text{ for } 0 < \text{y} < 1$$

where $\lambda > 0$, show that $W = -\log(Y)$ is an exponential$(1/\lambda)$ random variable.

2.61. If Y is an exponential$(1/\lambda)$ random variable, show that $W = e^{-Y}$ has pdf

$$f_W(w) = \lambda w^{\lambda-1} \text{ for } 0 < \text{w} < 1.$$

2.62. If $Y \sim EXP(\lambda)$, find the pdf of $W = 2\lambda Y$.

2.63*. (Mukhopadhyay 2000, p. 113): Suppose that $X|Y \sim N(\beta_0 + \beta_1 Y, Y^2)$, and that $Y \sim N(3, 10)$. That is, the conditional distribution of X given that $Y = y$ is normal with mean $\beta_0 + \beta_1 y$ and variance y^2 while the (marginal) distribution of Y is normal with mean 3 and variance 10.

a) Find EX.

b) Find Var X.

2.64*. Suppose that

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix} \sim N_4 \left(\begin{pmatrix} 49 \\ 100 \\ 17 \\ 7 \end{pmatrix}, \begin{pmatrix} 3 & 1 & -1 & 0 \\ 1 & 6 & 1 & -1 \\ -1 & 1 & 4 & 0 \\ 0 & -1 & 0 & 2 \end{pmatrix} \right).$$

a) Find the distribution of X_2.

b) Find the distribution of $(X_1, X_3)^T$.

c) Which pairs of random variables X_i and X_j are independent?

d) Find the correlation $\rho(X_1, X_3)$.

2.65*. Recall that if $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the conditional distribution of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is multivariate normal with mean $\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2)$ and covariance matrix $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.

Let $\sigma_{12} = \text{Cov}(Y, X)$ and suppose Y and X follow a bivariate normal distribution

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 49 \\ 100 \end{pmatrix}, \begin{pmatrix} 16 & \sigma_{12} \\ \sigma_{12} & 25 \end{pmatrix} \right).$$

a) If $\sigma_{12} = 0$, find $Y|X$. Explain your reasoning.

b) If $\sigma_{12} = 10$ find $E(Y|X)$.

c) If $\sigma_{12} = 10$, find $\text{Var}(Y|X)$.

2.66. Let $\sigma_{12} = \text{Cov}(Y, X)$ and suppose Y and X follow a bivariate normal distribution

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2\left(\begin{pmatrix} 15 \\ 20 \end{pmatrix}, \begin{pmatrix} 64 & \sigma_{12} \\ \sigma_{12} & 81 \end{pmatrix} \right).$$

a) If $\sigma_{12} = 10$ find $E(Y|X)$.

b) If $\sigma_{12} = 10$, find $\text{Var}(Y|X)$.

c) If $\sigma_{12} = 10$, find $\rho(Y, X)$, the correlation between Y and X.

2.67*. (Mukhopadhyay 2000, p. 197): Suppose that X_1 and X_2 have a joint pdf given by

$$f(x_1, x_2) = 3(x_1 + x_2)I(0 < x_1 < 1)I(0 < x_2 < 1)I(0 < x_1 + x_2 < 1).$$

Consider the transformation $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

a) Find the Jacobian J for the transformation.

b) Find the support $\mathcal{Y}$ of Y_1 and Y_2.

c) Find the joint density $f_{Y_1,Y_2}(y_1, y_2)$.

d) Find the marginal pdf $f_{Y_1}(y_1)$.

e) Find the marginal pdf $f_{Y_2}(y_2)$.

Hint for d) and e): $I_{A_1}(\boldsymbol{y})I_{A_2}(\boldsymbol{y})I_{A_3}(\boldsymbol{y}) = I_{\cap_{j=1}^3 A_j}(\boldsymbol{y}) = I_{\mathcal{Y}}(\boldsymbol{y})$ where $\mathcal{Y}$ is a triangle.

2.68*. (Aug. 2000 Qual): Suppose that the conditional distribution of $Y|\Lambda = \lambda$ is the Poisson(λ) distribution and that the random variable Λ has an exponential(1) distribution.

a) Find E(Y).

b) Find Var(Y).

2.69. Let A and B be positive integers. A hypergeometric random variable $X = W_1 + W_2 + \cdots + W_n$ where the random variables W_i are identically distributed random variables with $P(W_i = 1) = A/(A + B)$ and $P(W_i = 0) = B/(A + B)$. You may use the fact that $E(W_1) = A/(A + B)$ and that $E(X) = nA/(A + B)$.

a) Find $\text{Var}(W_1)$.

b) If $i \neq j$, then $\text{Cov}(W_i, W_j) = \dfrac{-AB}{(A+B)^2(A+B-1)}$. Find Var(X) using the formula

$$Var(\sum_{i=1}^{n} W_i) = \sum_{i=1}^{n} Var(W_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \text{Cov}(W_i, W_j).$$

(Hint: the sum $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}$ has $(n-1)n/2$ terms.)

2.70. Let $X = W_1 + W_2 + \cdots + W_n$ where the joint distribution of the random variables W_i is an n-dimensional multivariate normal distribution with $E(W_i) = 1$ and $\text{Var}(W_i) = 100$ for $i = 1, ..., n$.

a) Find $E(X)$.

b) Suppose that if $i \neq j$, then $\text{Cov}(W_i, W_j) = 10$. Find Var(X) using the formula

$$Var(\sum_{i=1}^{n} W_i) = \sum_{i=1}^{n} Var(W_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \text{Cov}(W_i, W_j).$$

(Hint: the sum $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}$ has $(n-1)n/2$ terms.)

2.71. Find the moment generating function for Y_1 if the joint probability mass function $f(y_1, y_2)$ of Y_1 and Y_2 is tabled as shown.

$f(y_1, y_2)$		y_2 = 0	1	2
y_1	0	0.38	0.14	0.24
	1	0.17	0.02	0.05

2.72. Suppose that the joint pdf of X and Y is $f(x,y) =$

$$\frac{1}{2}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp(\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2))$$

$$+\frac{1}{2}\frac{1}{2\pi\sqrt{1-\rho^2}}\exp(\frac{-1}{2(1-\rho^2)}(x^2 + 2\rho xy + y^2))$$

where x and y are real and $0 < \rho < 1$. It can be shown that the marginal pdfs are

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-1}{2}x^2)$$

for x real and

$$f_Y(y) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-1}{2}y^2)$$

for y real. Are X and Y independent? Explain briefly.

2.73*. Suppose that the conditional distribution of $Y|P = \rho$ is the binomial(k, ρ) distribution and that the random variable P has a beta$(\delta = 4, \nu = 6)$ distribution.

a) Find E(Y).

b) Find Var(Y).

2.74*. Suppose that the joint probability mass function $f(y_1, y_2)$ of Y_1 and Y_2 is given in the following table.

$f(y_1, y_2)$		y_2 = 0	1	2
y_1	0	0.38	0.14	0.24
	1	0.17	0.02	0.05

a) Find the marginal probability function $f_{Y_2}(y_2)$ for Y_2.

b) Find the conditional probability function $f(y_1|y_2)$ of Y_1 given $Y_2 = 2$.

2.75*. Find the pmf of $Y = X^2 + 4$ where the pmf of X is given below.

```
-----------------------------------------------
      X       | -2       -1      0     1     2
 probability | 0.1      0.2     0.4   0.2   0.1
-----------------------------------------------
```

2.76. Suppose that X_1 and X_2 are independent with $X_1 \sim N(0,1)$ and $X_2 \sim N(0,4)$ so $\text{Var}(X_2) = 4$. Consider the transformation $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

a) Find the Jacobian J for the transformation.

b) Find the joint pdf $f(y_1, y_2)$ of Y_1 and Y_2.

c) Are Y_1 and Y_2 independent? Explain briefly.

Hint: can you factor the joint pdf so that $f(y_1, y_2) = g(y_1)h(y_2)$ for every real y_1 and y_2?

2.77. (Aug. 2000 Qual): The number of defects per yard, Y of a certain fabric is known to have a Poisson distribution with parameter λ. However, λ is a random variable with pdf

$$f(\lambda) = e^{-\lambda} I(\lambda > 0).$$

a) Find E(Y).
b) Find Var(Y).