CHAPTER 3

The Normal Distribution on a Vector Space

The univariate normal distribution occupies a central position in the statistical theory of analyzing random samples consisting of one-dimensional observations. This situation is even more pronounced in multivariate analysis due to the paucity of analytically tractable multivariate distributions—one notable exception being the multivariate normal distribution. Ordinarily, the nonsingular multivariate normal distribution is defined on R^n by specifying the density function of the distribution with respect to Lebesgue measure. For our purposes, this procedure poses some problems. First, it is desirable to have a definition that does not require the covariance to be nonsingular. In addition, we have not, as yet, constructed what will be called Lebesgue measure on a finite dimensional inner product space. The definition of the multivariate normal distribution we have chosen circumvents the above technical difficulties by specifying the distribution of each linear function of the random vector. Of course, this necessitates a proof that such normal distributions exist.

After defining the normal distribution in a finite dimensional vector space and establishing some basic properties of the normal distribution, we derive the distribution of a quadratic form in a normal random vector. Conditions for the independence of two quadratic forms are then presented followed by a discussion of conditional distributions for normal random vectors. The chapter ends with a derivation of Lebesgue measure on a finite dimensional vector space and of the density function of a nonsingular normal distribution on a vector space.

3.1. THE NORMAL DISTRIBUTION

Recall that a random variable $Z_0 \in R$ has a normal distribution with mean zero and variance one if the density function of Z_0 is

$$p(z) = (2\pi)^{-1/2}\exp\left[-\tfrac{1}{2}z^2\right], \qquad z \in R$$

with respect to Lebesgue measure. We write $\mathcal{L}(Z_0) = N(0, 1)$ when Z_0 has density p. More generally, a random variable $Z \in R$ has a normal distribution with mean $\mu \in R$ and variance $\sigma^2 \geqslant 0$ if $\mathcal{L}(Z) = \mathcal{L}(\sigma Z_0 + \mu)$ where $\mathcal{L}(Z_0) = N(0, 1)$. In this case, we write $\mathcal{L}(Z) = N(\mu, \sigma^2)$. When $\sigma^2 = 0$, the distribution $N(\mu, \sigma^2)$ is to be interpreted as the distribution degenerate at μ. If $\mathcal{L}(Z) = N(\mu, \sigma^2)$, then the characteristic function of Z is easily shown to be

$$\phi(t) = \exp\left[i\mu t - \tfrac{1}{2}\sigma^2 t^2\right], \qquad t \in R.$$

The phrase "Z has a normal distribution" means that for some μ and some $\sigma \geqslant 0$, $\mathcal{L}(Z) = N(\mu, \sigma^2)$. If $Z_1, \dots, Z_k$ are independent with $\mathcal{L}(Z_j) = N(\mu_j, \sigma_j^2)$, then $\mathcal{L}(\Sigma\alpha_j Z_j) = N(\Sigma\alpha_j\mu_j, \Sigma\alpha_j^2\sigma_j^2)$. To see this, consider the characteristic function

$$\mathcal{E}\exp\left[it\Sigma\alpha_j Z_j\right] = \mathcal{E}\prod_{j=1}^{k}\exp\left[it\alpha_j Z_j\right] = \prod_{j=1}^{k}\mathcal{E}\exp\left[it\alpha_j Z_j\right]$$

$$= \prod_{j=1}^{k}\exp\left[it\alpha_j\mu_j - \tfrac{1}{2}t^2\alpha_j^2\sigma_j^2\right]$$

$$= \exp\left[it(\Sigma\alpha_j\mu_j) - \tfrac{1}{2}t^2(\Sigma\alpha_j^2\sigma_j^2)\right].$$

Thus the characteristic function of $\Sigma\alpha_j Z_j$ is that of a normal distribution with mean $\Sigma\alpha_j\mu_j$ and variance $\Sigma\alpha_j^2\sigma_j^2$. In summary, linear combinations of independent normal random variables are normal.

We are now in a position to define the normal distribution on a finite dimensional inner product space $(V, (\cdot, \cdot))$.

Definition 3.1. A random vector $X \in V$ has a normal distribution if, for each $x \in V$, the random variable (x, X) has a normal distribution on R.

To show that a normal distribution exists on $(V, (\cdot, \cdot))$, let $\{x_1, \dots, x_n\}$ be an orthonormal basis for $(V, (\cdot, \cdot))$. Also, let $Z_1, \dots, Z_n$ be independent

$N(0, 1)$ random variables. Then $X \equiv \Sigma Z_i x_i$ is a random vector and $(x, X) = \Sigma(x, x_i)Z_i$, which is a linear combination of independent normals. Thus (x, X) has a normal distribution for each $x \in V$. Since $\mathcal{E}(x, X) = \Sigma(x_i, x)\mathcal{E}Z_i = 0$, the mean vector of X is $0 \in V$. Also,

$$\operatorname{var}(x, X) = \operatorname{var}(\Sigma(x, x_i)Z_i) = \Sigma(x, x_i)^2\operatorname{var}(Z_i) = \Sigma(x, x_i)^2 = (x, x).$$

Therefore, $\operatorname{Cov}(X) = I \in \mathcal{L}(V, V)$. The particular normal distribution we have constructed on $(V, (\cdot, \cdot))$ has mean zero and covariance equal to the identity linear transformation.

Now, we want to describe all the normal distributions on $(V, (\cdot, \cdot))$. The first result in this direction shows that linear transformations of normal random vectors are again normal random vectors.

Proposition 3.1. Suppose X has a normal distribution on $(V, (\cdot, \cdot))$ and let $A \in \mathcal{L}(V, W)$, $w_0 \in W$. Then $AX + w_0$ has a normal distribution on $(W, [\cdot, \cdot])$.

Proof. It must be shown that, for each $w \in W$, $[w, AX + w_0]$ has a normal distribution on R. But $[w, AX + w_0] = [w, AX] + [w, w_0] = (A'w, X) + [w, w_0]$. By assumption, $(A'w, X)$ is normal. Since $[w, w_0]$ is a constant, $(A'w, X) + [w, w_0]$ is normal. □

If X has a normal distribution on $(V, (\cdot, \cdot))$ with mean zero and covariance I, consider $A \in \mathcal{L}(V, V)$ and $\mu \in V$. Then $AX + \mu$ has a normal distribution on $(V, (\cdot, \cdot))$ and we know $\mathcal{E}(AX + \mu) = A(\mathcal{E}X) + \mu = \mu$ and $\operatorname{Cov}(AX + \mu) = A\operatorname{Cov}(X)A' = AA'$. However, every positive semidefinite linear transformation Σ can be expressed as AA' (take A to be the positive semidefinite square root of Σ). Thus given $\mu \in V$ and a positive semidefinite Σ, there is a random vector that has a normal distribution in V with mean vector μ and covariance Σ. If X has such a distribution, we write $\mathcal{L}(X) = N(\mu, \Sigma)$. To show that all the normal distributions on V have been described, suppose $X \in V$ has a normal distribution. Since (x, X) is normal on R, $\operatorname{var}(x, X)$ exists for each $x \in V$. Thus $\mu = \mathcal{E}X$ and $\Sigma = \operatorname{Cov}(X)$ both exist and $\mathcal{L}(X) = N(\mu, \Sigma)$. Also, $\mathcal{L}((x, X)) = N((x, \mu), (x, \Sigma x))$ for $x \in V$. Hence the characteristic function of (x, X) is

$$\phi(t) = \mathcal{E}\exp[it(x, X)] = \exp\left[it(x, \mu) - \tfrac{1}{2}t^2(x, \Sigma x)\right].$$

Setting $t = 1$, we obtain the characteristic function of X:

$$\xi(x) = \mathcal{E}\exp[i(x, X)] = \exp\left[i(x, \mu) - \tfrac{1}{2}(x, \Sigma x)\right].$$

Summarizing this discussion yields the following.

Proposition 3.2. Given $\mu \in V$ and a positive semidefinite $\Sigma \in \mathcal{L}(V, V)$, there exists a random vector $X \in V$ with distribution $N(\mu, \Sigma)$ and characteristic function

$$\xi(x) = \exp\left[i(x, \mu) - \tfrac{1}{2}(x, \Sigma x)\right].$$

Conversely, if X has a normal distribution on V, then with $\mu = \mathcal{E}X$ and $\Sigma = \text{Cov}(X)$, $\mathcal{L}(X) = N(\mu, \Sigma)$ and the characteristic function of X is given by ξ.

Consider random vectors X_i with values in $(V_i, (\cdot, \cdot)_i)$ for $i = 1, 2$. Then $\{X_1, X_2\}$ is a random vector in the direct sum $V_1 \oplus V_2$. The inner product on $V_1 \oplus V_2$ is $[\cdot, \cdot]$ where

$$[\{v_1, v_2\}, \{v_3, v_4\}] \equiv (v_1, v_3)_1 + (v_2, v_4)_2,$$

$v_1, v_3 \in V_1$ and $v_2, v_4 \in V_2$. If $\text{Cov}(X_i) = \Sigma_{ii}$, $i = 1, 2$, exists, then $\mathcal{E}\{X_1, X_2\} = \{\mu_1, \mu_2\}$ where $\mu_i = \mathcal{E}X_i$, $i = 1, 2$. Also,

$$\Sigma \equiv \text{Cov}\{X_1, X_2\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \in \mathcal{L}(V_1 \oplus V_2, V_1 \oplus V_2)$$

as defined in Chapter 2 and $\Sigma_{21} \equiv \Sigma'_{12}$.

Proposition 3.3. If $\{X_1, X_2\}$ has a normal distribution on $V_1 \oplus V_2$, then X_1 and X_2 are independent iff $\Sigma_{12} = 0$.

Proof. If X_1 and X_2 are independent, then clearly $\Sigma_{12} = 0$. Conversely, if $\Sigma_{12} = 0$, the characteristic function of $\{X_1, X_2\}$ is

$$\begin{aligned}
\mathcal{E}\exp\{i[\{v_1, v_2\}, \{X_1, X_2\}]\} &= \exp\{i[\{v_1, v_2\}, \{\mu_1, \mu_2\}] \\
&\quad - \tfrac{1}{2}[\{v_1, v_2\}, \Sigma\{v_1, v_2\}]\} \\
&= \exp\{i(v_1, \mu_1)_1 + i(v_2, \mu_2)_2 \\
&\quad - \tfrac{1}{2}(v_1, \Sigma_{11}v_1)_1 - \tfrac{1}{2}(v_2, \Sigma_{22}v_2)_2\} \\
&= \exp\{i(v_1, \mu_1)_1 - \tfrac{1}{2}(v_1, \Sigma_{11}v_1)_1\} \\
&\quad \times \exp\{i(v_2, \mu_2)_2 - \tfrac{1}{2}(v_2, \Sigma_{22}v_2)_2\}
\end{aligned}$$

since $\Sigma_{12} = \Sigma'_{21} = 0$. However, for $v_1 \in V_1$, $(v_1, X_1)_1 = [\{v_1, 0\}, \{X_1, X_2\}]$, which has a normal distribution for all $v_1 \in V_1$. Thus $\mathfrak{L}(X_1) = N(\mu_1, \Sigma_1)$ on V_1 and similarly $\mathfrak{L}(X_2) = N(\mu_2, \Sigma_2)$ on V_2. The characteristic function of $\{X_1, X_2\}$ is just the product of the characteristic functions of X_1 and X_2. Thus independence follows and the proof is complete. □

The result of Proposition 3.3 is often paraphrased as "for normal random vectors, X_1 and X_2 are independent iff they are uncorrelated." A useful consequence of Proposition 3.3 is shown in Proposition 3.4.

Proposition 3.4. Suppose $\mathfrak{L}(X) = N(\mu, \Sigma)$ on $(V, (\cdot, \cdot))$, and consider $A \in \mathfrak{L}(V, W_1)$, $B \in \mathfrak{L}(V, W_2)$ where $(W_1, [\cdot, \cdot]_1)$ and $(W_2, [\cdot, \cdot]_2)$ are inner product spaces. AX and BX are independent iff $A\Sigma B' = 0$.

Proof. We apply the previous proposition to $X_1 = AX$ and $X_2 = BX$. That $\{X_1, X_2\}$ has a normal distribution on $W_1 \oplus W_2$ follows from

$$[w_1, X_1]_1 + [w_2, X_2]_2 = (A'w_1, X) + (B'w_2, X) = (A'w_1 + B'w_2, X)$$

and the normality of (x, X) for all $x \in V$. However,

$$\begin{aligned} \operatorname{cov}\{[w_1, X_1]_1, [w_2, X_2]_2\} &= \operatorname{cov}\{(A'w_1, X), (B'w_2, X)\} \\ &= (A'w_1, \Sigma B'w_2) \\ &= [w_1, A\Sigma B'w_2]_1. \end{aligned}$$

Thus $X_1 = AX$ and $X_2 = BX$ are uncorrelated iff $A\Sigma B' = 0$. Since $\{X_1, X_2\}$ has a normal distribution, the condition $A\Sigma B' = 0$ is equivalent to the independence of X_1 and X_2. □

One special case of Proposition 3.4 is worthy of mention. If $\mathfrak{L}(X) = N(\mu, I)$ on $(V, (\cdot, \cdot))$ and P is an orthogonal projection in $\mathfrak{L}(V, V)$, then PX and $(I - P)X$ are independent since $P(I - P) = 0$. Also, it should be mentioned that the result of Proposition 3.3 extends to the case of k random vectors—that is, if $\{X_1, X_2, \ldots, X_k\}$ has a normal distribution on the direct sum space $V_1 \oplus V_2 \oplus \cdots \oplus V_k$, then $X_1, X_2, \ldots, X_k$ are independent iff X_i and X_j are uncorrelated for all $i \neq j$. The proof of this is essentially the same as that given for the case of $k = 2$ and is left to the reader.

A particularly useful result for the multivariate normal distribution is the following.

Proposition 3.5. Suppose $\mathcal{L}(X) = N(\mu, \Sigma)$ on the n-dimensional vector space $(V, (\cdot, \cdot))$. Write $\Sigma = \sum_1^n \lambda_i x_i \square x_i$ in spectral form, and let $X_i = (x_i, X)$, $i = 1, \ldots, n$. Then $X_1, \ldots, X_n$ are independent random variables that have a normal distribution on R with $\mathcal{E}X_i = (x_i, \mu)$ and $\text{var}(X_i) = \lambda_i$, $i = 1, \ldots, n$. In particular, if $\Sigma = I$, then for any orthonormal basis $\{x_1, \ldots, x_n\}$ for V, the random variables $X_i = (x_i, X)$ are independent and normal with $\mathcal{E}X_i = (x_i, \mu)$ and $\text{var}(X_i) = 1$.

Proof. For any scalars $\alpha_1, \ldots, \alpha_n$ in R, $\sum_1^n \alpha_i X_i = \sum_1^n \alpha_i (x_i, X) = (\sum_1^n \alpha_i x_i, X)$, which has a normal distribution. Thus the random vector $\tilde{X} \in R^n$ with coordinates $X_1, \ldots, X_n$ has a normal distribution in the coordinate vector space R^n. Thus $X_1, \ldots, X_n$ are independent iff they are uncorrelated. However,

$$\text{cov}\{X_j, X_k\} = \text{cov}\{(x_j, X), (x_k, X)\} = (x_j, \Sigma x_k)$$

$$= \left(x_j, (\textstyle\sum_1^n \lambda_i x_i \square x_i) x_k\right) = \lambda_j \delta_{jk}.$$

Thus independence follows. It is clear that each X_i is normal with $\mathcal{E}X_i = (x_i, \mu)$ and $\text{var}(X_i) = \lambda_i$, $i = 1, \ldots, n$. When $\Sigma = I$, then $\sum_1^n x_i \square x_i = I$ for any orthonormal basis $x_1, \ldots, x_n$. This completes the proof. □

The following is a technical discussion having to do with representations of the normal distribution that are useful when establishing properties of the normal distribution. It seems preferable to dispose of the issues here rather than repeat the same argument in a variety of contexts later. Suppose $X \in (V, (\cdot, \cdot))$ has a normal distribution, say $\mathcal{L}(X) = N(\mu, \Sigma)$, and let Q be the probability distribution of X on $(V, (\cdot, \cdot))$. If we are interested in the distribution of some function of X, say $f(X) \in (W, [\cdot, \cdot])$, then the underlying space on which X is defined is irrelevant since the distribution Q determines the distribution of $f(X)$—that is, if $B \in \mathcal{B}(W)$, then

$$P\{f(X) \in B\} = P\{X \in f^{-1}(B)\} = Q\left(f^{-1}(B)\right).$$

Therefore, if Y is another random vector in $(V, (\cdot, \cdot))$ with $\mathcal{L}(X) = \mathcal{L}(Y)$, then $f(X)$ and $f(Y)$ have the same distribution. At times, it is convenient to represent $\mathcal{L}(X)$ by $\mathcal{L}(CZ + \mu)$ where $\mathcal{L}(Z) = N(0, \text{I})$ and $CC' = \Sigma$. Thus

$\mathcal{L}(X) = \mathcal{L}(CZ + \mu)$ so $f(X)$ and $f(CZ + \mu)$ have the same distribution. A slightly more subtle point arises when we discuss the independence of two functions of X, say $f_1(X)$ and $f_2(X)$, taking values in $(W_1, [\cdot, \cdot]_1)$ and $(W_2, [\cdot, \cdot]_2)$. To show that independence of $f_1(X)$ and $f_2(X)$ depends only on Q, consider $B_i \in \mathcal{B}(W_i)$ for $i = 1, 2$. Then independence is equivalent to

$$P\{f_1(X) \in B_1, f_2(X) \in B_2\} = P\{f_1(X) \in B_1\}P\{f_2(X) \in B_2\}.$$

But both of these probabilities can be calculated from Q:

$$\begin{aligned} P\{f_1(X) \in B_1, f_2(X) \in B_2\} &= P\{X \in f_1^{-1}(B_1) \cap f_2^{-1}(B_2)\} \\ &= Q\left(f_1^{-1}(B_1) \cap f_2^{-1}(B_2)\right) \end{aligned}$$

and

$$P\{f_i(X) \in B_i\} = Q\left(f_i^{-1}(B_i)\right), \qquad i = 1, 2.$$

Again, if $\mathcal{L}(Y) = \mathcal{L}(X)$, then $f_1(X)$ and $f_2(X)$ are independent iff $f_1(Y)$ and $f_2(Y)$ are independent. More generally, if we are trying to prove something about the random vector X, $\mathcal{L}(X) = N(\mu, \Sigma)$, and if what we are trying to prove depends only on the distribution Q_1 of X, then we can represent X by any other random vector Y as long as $\mathcal{L}(Y) = \mathcal{L}(X)$. In particular, we can take $Y = CZ + \mu$ where $\mathcal{L}(Z) = N(0, I)$ and $CC' = \Sigma$. This representation of X is often used in what follows.

3.2. QUADRATIC FORMS

The problem in this section is to derive, or at least describe, the distribution of (X, AX) where $X \in (V, (\cdot, \cdot))$, A is self-adjoint in $\mathcal{L}(V, V)$ and $\mathcal{L}(X) = N(\mu, \Sigma)$. First, consider the special case of $\Sigma = I$, and by the spectral theorem, write $A = \Sigma_1^n \lambda_i x_i \square x_i$. Thus

$$(X, AX) = \left(X, (\Sigma_1^n \lambda_i x_i \square x_i) X\right) = \Sigma_1^n \lambda_i (x_i, X)^2.$$

But $X_i \equiv (x_i, X)$, $i = 1, \ldots, n$, are independent since $\Sigma = I$ (Proposition 3.5) and $\mathcal{L}(X_i) = N((x_i, \mu), 1)$. Thus our first task is to derive the distribution of X_i^2 when $\mathcal{L}(X_i) = N((x_i, \mu), 1)$.

Recall that a random variable Z has a chi-square distribution with m degrees of freedom, written $\mathcal{L}(Z) = \chi_m^2$, if Z has a density on $(0, \infty)$ given by

$$p_m(z) = \frac{z^{(m/2)-1}}{\Gamma(m/2)2^{m/2}}\exp[-\tfrac{1}{2}z], \qquad z > 0.$$

Here m is a positive integer and $\Gamma(\cdot)$ is the gamma function. The characteristic function of a χ_m^2 random variable is easily shown to be

$$\mathcal{E}e^{it\chi_m^2} = (1-2it)^{-m/2}, \qquad t \in R^1.$$

Thus, if $\mathcal{L}(Z_1) = \chi_m^2$, $\mathcal{L}(Z_2) = \chi_n^2$, and Z_1 and Z_2 are independent, then

$$\mathcal{E}\exp[it(Z_1 + Z_2)] = \mathcal{E}\exp[itZ_1]\mathcal{E}\exp[itZ_2]$$
$$= (1-2it)^{-m/2}(1-2it)^{-n/2} = (1-2it)^{-(m+n)/2}.$$

Therefore, $\mathcal{L}(Z_1 + Z_2) = \chi_{m+n}^2$. This argument clearly extends to more than two factors. In particular, if $\mathcal{L}(Z) = \chi_m^2$, then, for independent random variables $Z_1, \ldots, Z_m$ with $\mathcal{L}(Z_i) = \chi_1^2$, $\mathcal{L}(\Sigma_1^m Z_i) = \mathcal{L}(Z)$. It is not difficult to show that if $\mathcal{L}(X) = N(0,1)$ on R, then $\mathcal{L}(X^2) = \chi_1^2$. However, if $\mathcal{L}(X) = N(\alpha, 1)$ on R, the distribution of X^2 is a bit harder to derive. To this end, we make the following definition.

Definition 3.2. Let p_m, $m = 1, 2, \ldots,$ be the density of a χ_m^2 random variable and, for $\lambda \geqslant 0$, let

$$q_j = \exp\left[-\frac{\lambda}{2}\right]\frac{1}{j!}\left(\frac{\lambda}{2}\right)^j.$$

For $\lambda = 0$, $q_0 = 1$ and $q_j = 0$ for $j > 0$. A random variable with density

$$h(z) = \sum_{j=0}^{\infty} q_j p_{m+2j}(z), \qquad z > 0$$

is said to have a *noncentral chi-square distribution* with m degrees of freedom and noncentrality parameter λ. If Z has such a distribution, we write $\mathcal{L}(Z) = \chi_m^2(\lambda)$.

When $\lambda = 0$, it is clear that $\mathcal{L}(\chi_m^2(0)) = \chi_m^2$. The weights $q_j, j = 0, 1, \ldots,$ are Poisson probabilities with parameter $\lambda/2$ (the reason for the 2 becomes clear in a bit). The characteristic function of a $\chi_m^2(\lambda)$ random variable is calculated as follows:

$$\begin{aligned}
\mathcal{E}\exp\left[it\chi_m^2(\lambda)\right] &= \sum_{j=0}^{\infty} q_j \int_0^{\infty} \exp(itx) p_{m+2j}(x)\,dx \\
&= \sum_{j=0}^{\infty} q_j (1-2it)^{-(m/2+j)} \\
&= (1-2it)^{-m/2} \sum_{j=0}^{\infty} q_j (1-2it)^{-j} \\
&= (1-2it)^{-m/2} \exp(-\lambda/2) \sum_{j=0}^{\infty} \left(\frac{\lambda}{2}\right)^j \frac{(1-2it)^{-j}}{j!} \\
&= (1-2it)^{-m/2} \exp\left[-\frac{\lambda}{2} + \frac{\lambda}{2}\frac{1}{1-2it}\right] \\
&= (1-2it)^{-m/2} \exp\frac{\lambda}{2}\left[\frac{2it}{1-2it}\right].
\end{aligned}$$

From this expression for the characteristic function, it follows that if $\mathcal{L}(Z_i) = \chi_{m_i}^2(\lambda_i)$, $i = 1, 2$, with Z_1 and Z_2 independent, then $\mathcal{L}(Z_1 + Z_2) = \chi_{m_1+m_2}^2(\lambda_1 + \lambda_2)$. This result clearly extends to the sum of k independent noncentral chi-square variables. The reason for introducing the noncentral chi-square distribution is provided in the next result.

Proposition 3.6. Suppose $\mathcal{L}(X) = N(\alpha, 1)$ on R. Then $\mathcal{L}(X^2) = \chi_1^2(\alpha^2)$.

Proof. The proof consists of calculating the characteristic function of X^2. A justification of the change of variable in the calculation below can be given using contour integration. The characteristic function of X^2 is

$$\begin{aligned}
\mathcal{E}\exp(itX^2) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[itx^2 - \tfrac{1}{2}(x-\alpha)^2\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}(1-2it)x^2 + \alpha x - \tfrac{1}{2}\alpha^2\right] dx \\
&= \frac{(1-2it)^{-1/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}w^2 + \alpha(1-2it)^{-1/2} \right. \\
&\qquad \left. \times w - \tfrac{1}{2}\alpha^2\right] dw
\end{aligned}$$

$$= \frac{(1-2it)^{-1/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}\left(w - \alpha(1-2it)^{-1/2}\right)^2 + \frac{\alpha^2}{2}\left(\frac{2it}{1-2it}\right)\right] dw$$

$$= (1-2it)^{-1/2} \exp \frac{\alpha^2}{2}\left(\frac{2it}{1-2it}\right).$$

By the uniqueness of characteristic functions, $\mathcal{L}(X^2) = \chi_1^2(\alpha^2)$. □

Proposition 3.7. Suppose the random vector X in $(V, (\cdot,\cdot))$ has a $N(\mu, I)$ distribution. If $A \in \mathcal{L}(V, V)$ is an orthogonal projection of rank k, then $\mathcal{L}((X, AX)) = \chi_k^2((\mu, A\mu))$.

Proof. Let $\{x_1, \dots, x_k\}$ be an orthonormal basis for the range of A. Thus $A = \Sigma_1^k x_i \square x_i$ and

$$(X, AX) = \Sigma_1^k (x_i, X)^2.$$

But the random variables $(x_i, X)^2$, $i = 1, \dots, k$, are independent (Proposition 3.5) and, by Proposition 3.6, $\mathcal{L}(X_i^2) = \chi_1^2((x_i, \mu)^2)$. From the additive property of independent noncentral chi-square variables,

$$\mathcal{L}\left(\Sigma_1^k (x_i, X)^2\right) = \chi_k^2\left(\Sigma_1^k (x_i, \mu)^2\right).$$

Noting that $(\mu, A\mu) = \Sigma_1^k (x_i, \mu)^2$, the proof is complete. □

When $\mathcal{L}(X) = N(\mu, \Sigma)$, the distribution of the quadratic form (X, AX), with A self-adjoint, is reasonably complicated, but there is something that can be said. Let B be the positive semidefinite square root of Σ and assume that $\mu \in \mathfrak{R}(\Sigma)$. Thus $\mu \in \mathfrak{R}(B)$ since $\mathfrak{R}(B) = \mathfrak{R}(\Sigma)$. Therefore, for some vector $\tau \in V$, $\mu = B\tau$. Thus $\mathcal{L}(X) = \mathcal{L}(BY)$ where $\mathcal{L}(Y) = N(\tau, I)$ and it suffices to describe the distribution of $(BY, ABY) = (Y, BABY)$. Since A and B are self-adjoint, BAB is self-adjoint. Write BAB in spectral form:

$$BAB = \Sigma_1^n \lambda_i x_i \square x_i$$

where $\{x_1, \dots, x_n\}$ is an orthonormal basis for $(V, (\cdot,\cdot))$. Then

$$(Y, BABY) = \Sigma_1^n \lambda_i (x_i, Y)^2$$

and the random variables (x_i, Y), $i = 1,\ldots,n$, are independent with $\mathcal{L}((x_i, Y)^2) = \chi_1^2((x_i, \tau)^2)$. It follows that the quadratic form $(Y, BABY)$ has the same distribution as a linear combination of independent noncentral chi-square random variables. Symbolically,

$$\mathcal{L}((Y, BABY)) = \mathcal{L}\left(\Sigma_1^n \lambda_i \chi_{1,i}^2\left((x_i, \tau)^2\right)\right).$$

In general not much more can be said about this distribution without some assumptions concerning the eigenvalues $\lambda_1,\ldots,\lambda_n$. However, when BAB is an orthogonal projection of rank k, then Proposition 3.7 is applicable and

$$\mathcal{L}((Y, BABY)) = \chi_k^2((\tau, BAB\tau)) = \chi_k^2((B\tau, AB\tau)) = \chi_k^2((\mu, A\mu)).$$

In summary, we have the following.

Proposition 3.8. Suppose $\mathcal{L}(X) = N(\mu, \Sigma)$ where $\mu \in \mathcal{R}(\Sigma)$, and let B be the positive semidefinite square root of Σ. If A is self-adjoint and BAB is a rank k orthogonal projection, then

$$\mathcal{L}((X, AX)) = \chi_k^2((\mu, A\mu)).$$

We can use a slightly different set of assumptions and reach the same conclusion as Proposition 3.8, as follows.

Proposition 3.9. Suppose $\mathcal{L}(X) = N(\mu, \Sigma)$ and let B be the positive semidefinite square root of Σ. Write $\mu = \mu_1 + \mu_2$ where $\mu_1 \in \mathcal{R}(\Sigma)$ and $\mu_2 \in \mathcal{N}(\Sigma)$. If A is a self-adjoint such that $A\mu_2 = 0$ and BAB is a rank k orthogonal projection, then

$$\mathcal{L}((X, AX)) = \chi_k^2((\mu, A\mu)).$$

Proof. Since $A\mu_2 = 0$, $(X, AX) = (X - \mu_2, A(X - \mu_2))$. Let $Y = X - \mu_2$ so $\mathcal{L}(Y) = N(\mu_1, \Sigma)$ and $\mathcal{L}((X, AX)) = \mathcal{L}((Y, AY))$. Since $\mu_1 \in \mathcal{R}(\Sigma)$, Proposition 3.8 shows that

$$\mathcal{L}((Y, AY)) = \chi_k^2((\mu_1, A\mu_1)).$$

However, $(\mu, A\mu) = (\mu_1, A\mu_1)$ as $A\mu_2 = 0$. □

3.3. INDEPENDENCE OF QUADRATIC FORMS

Thus far, necessary and sufficient conditions for the independence of different linear transformations of a normal random vector have been given and the distribution of a quadratic form in a normal random vector has been described. In this section, we give sufficient conditions for the independence of different quadratic forms in normal random vectors.

Suppose $X \in (V, (\cdot, \cdot))$ has an $N(\mu, \Sigma)$ distribution and consider two self-adjoint linear transformations, A_i, $i = 1, 2$, on V to V. To discuss the independence of $(X, A_1 X)$ and $(X, A_2 X)$, it is convenient to first reduce the discussion to the case when $\mu = 0$ and $\Sigma = I$. Let B be the positive semidefinite square root of Σ so if $\mathcal{L}(Y) = N(0, I)$, then $\mathcal{L}(X) = \mathcal{L}(BY + \mu)$. Thus it suffices to discuss the independence of $(BY + \mu, A_1(BY + \mu))$ and $(BY + \mu, A_2(BY + \mu))$ when $\mathcal{L}(Y) = N(0, I)$. However,

$$(BY + \mu, A_i(BY + \mu)) = (Y, BA_iBY) + 2(BA_i\mu, Y) + (\mu, A_i\mu)$$

for $i = 1, 2$. Let $C_i = BA_iB$, $i = 1, 2$, and let $x_i = 2BA_i\mu$. Then we want to know conditions under which $(Y, C_1Y) + (x_1, Y)$ and $(Y, C_2Y) + (x_2, Y)$ are independent when $\mathcal{L}(Y) = N(0, I)$. Clearly, the constants $(\mu, A_i\mu)$, $i = 1, 2$, do not affect the independence of the two quadratic forms. It is this problem, in reduced form, that is treated now. Before stating the principal result, the following technical proposition is needed.

Proposition 3.10. For self-adjoint linear transformations A_1 and A_2 on $(V, (\cdot, \cdot))$ to $(V, (\cdot, \cdot))$, the following are equivalent:

(i) $A_1A_2 = 0$.
(ii) $\mathfrak{R}(A_1) \perp \mathfrak{R}(A_2)$.

Proof. If $A_1A_2 = 0$, then $A_1A_2x = 0$ for all $x \in V$ so $\mathfrak{R}(A_2) \subseteq \mathfrak{N}(A_1)$. Since $\mathfrak{N}(A_1) \perp \mathfrak{R}(A_1)$, $\mathfrak{R}(A_2) \perp \mathfrak{R}(A_1)$. Conversely, if $\mathfrak{R}(A_1) \perp \mathfrak{R}(A_2)$, then $\mathfrak{R}(A_2) \subseteq \mathfrak{R}(A_1)^\perp = \mathfrak{N}(A_1)$ and this implies that $A_1A_2x = 0$ for all $x \in V$. Therefore, $A_1A_2 = 0$. □

Proposition 3.11. Let $Y \in (V, (\cdot, \cdot))$ have a $N(0, I)$ distribution and suppose $Z_i = (Y, A_iY) + (x_i, Y)$ where A_i is self-adjoint and $x_i \in V$, $i = 1, 2$. If $A_1A_2 = 0$, $A_1x_2 = 0$, $A_2x_1 = 0$, and $(x_1, x_2) = 0$, then Z_1 and Z_2 are independent random variables.

Proof. The idea of the proof is to show that Z_1 and Z_2 are functions of two different independent random vectors. To this end, let P_i be the orthogonal projection onto $\mathfrak{R}(A_i)$ for $i = 1, 2$. It is clear that $P_iA_iP_i = A_i$ for $i = 1, 2$. Thus $Z_i = (P_iY, A_iP_iY) + (x_i, Y)$ for $i = 1, 2$. The random vector $\{P_1Y, (x_1, Y)\}$ takes values in the direct sum $V \oplus R$ and Z_1 is a function of this vector. Also, $\{P_2Y, (x_2, Y)\}$ takes values in $V \oplus R$ and Z_2 is a function of this vector. The remainder of the proof is devoted to showing that $\{P_1Y, (x_1, Y)\}$ and $\{P_2Y, (x_2, Y)\}$ are independent random vectors. This is done by verifying that the random vectors are jointly normal and that they are uncorrelated. Let $[\cdot, \cdot]$ denote the induced inner product on the direct sum $V \oplus R$. The inner product of the vector $\{\{y_1, \alpha_1\}, \{y_2, \alpha_2\}\}$ in $(V \oplus R) \oplus (V \oplus R)$ with $\{\{P_1Y, (x_1, Y)\}, \{P_2Y, (x_2, Y)\}\}$ is

$$(y_1, P_1Y) + \alpha_1(x_1, Y) + (y_2, P_2Y) + \alpha_2(x_2, Y)$$

$$= (P_1y_1 + \alpha_1 x + P_2y_2 + \alpha_2 x_2, Y),$$

which has a normal distribution since Y is normal. Thus $\{\{P_1Y, (x_1, Y)\}, \{P_2Y, (x_2, Y)\}\}$ has a normal distribution. The independence of these two vectors follows from the calculation below, which shows the vectors are uncorrelated. For $\{y_1, \alpha_1\} \in V \oplus R$ and $\{y_2, \alpha_2\} \in V \oplus R$,

$$\begin{aligned}
&\operatorname{cov}\{[\{y_1, \alpha_1\}, \{P_1Y, (x_1, Y)\}], [\{y_2, \alpha_2\}, \{P_2Y, (x_2, Y)\}]\} \\
&\quad = \operatorname{cov}\{(y_1, P_1Y) + \alpha_1(x_1, Y), (y_2, P_2Y) + \alpha_2(x_2, Y)\} \\
&\quad = \operatorname{cov}\{(P_1y_1, Y), (P_2y_2, Y)\} + \alpha_1 \operatorname{cov}\{(x_1, Y), (P_2y_2, Y)\} \\
&\qquad + \alpha_2 \operatorname{cov}\{(P_1y_1, Y), (x_2, Y)\} + \alpha_1\alpha_2 \operatorname{cov}\{(x_1, Y), (x_2, Y)\} \\
&\quad = (P_1y_1, P_2y_2) + \alpha_1(x_1, P_2y_2) + \alpha_2(x_2, P_1y_1) + \alpha_1\alpha_2(x_1, x_2) \\
&\quad = (y_1, P_1P_2y_2) + \alpha_1(P_2x_1, y_2) + \alpha_2(P_1x_2, y_1) + \alpha_1\alpha_2(x_1, x_2).
\end{aligned}$$

However, $P_1P_2 = 0$ since $\mathcal{R}(A_1) \perp \mathcal{R}(A_2)$. Also, $P_2x_1 = 0$ as $x_1 \in \mathcal{N}(A_2)$ and, similarly, $P_1x_2 = 0$. Further, $(x_1, x_2) = 0$ by assumption. Thus the above covariance is zero so Z_1 and Z_2 are independent. □

A useful consequence of Proposition 3.11 is Proposition 3.12.

Proposition 3.12. Suppose $\mathcal{L}(X) = N(\mu, \Sigma)$ on $(V, (\cdot, \cdot))$ and let C_i, $i = 1, 2$, be self-adjoint linear transformations. If $C_1\Sigma C_2 = 0$, then (X, C_1X) and (X, C_2X) are independent.

Proof. Let B denote the positive semidefinite square root of Σ, and suppose $\mathcal{L}(Y) = N(0, I)$. It suffices to show that $Z_1 \equiv (BY + \mu, C_1(BY +$

μ)) is independent of $Z_2 \equiv (BY + \mu, C_2(BY + \mu))$ since $\mathcal{L}(X) = \mathcal{L}(BY + \mu)$. But

$$Z_i = (Y, BC_iBY) + 2(BC_i\mu, Y) + (\mu, C_i\mu)$$

for $i = 1, 2$. Proposition 3.11 can now be applied with $A_i = BC_iB$ and $x_i = 2BC_i\mu$ for $i = 1, 2$. Since $\Sigma = BB$, $A_1A_2 = BC_1BBC_2B = BC_1\Sigma C_2B = 0$ as $C_1\Sigma C_2 = 0$ by assumption. Also, $A_1x_2 = 2BC_1BBC_2\mu = 2BC_1\Sigma C_2\mu = 0$. Similarly, $A_2x_1 = 0$ and $(x_1, x_2) = 4(BC_1\mu, BC_2\mu) = 4(\mu, C_1\Sigma C_2\mu) = 0$. Thus $(Y, BC_1BY) + 2(BC_1\mu, Y)$ and $(Y, BC_2BY) + 2(BC_2\mu, Y)$ are independent. Hence Z_1 and Z_2 are independent. □

The results of this section are general enough to handle most situations that arise when dealing with quadratic forms. However, in some cases we need a sufficient condition for the independence of k quadratic forms. An examination of the proof of Proposition 3.11 shows that when $\mathcal{L}(Y) = N(0, I)$, the quadratic forms $Z_i = (Y, A_iY) + (x_i, Y)$, $i = 1, \dots, k$, are mutually independent if, for each $i \neq j$, $A_iA_j = 0$, $A_ix_j = 0$, $A_jx_i = 0$, and $(x_i, x_j) = 0$. The details of this verification are left to the reader.

3.4. CONDITIONAL DISTRIBUTIONS

The basic result of this section gives the conditional distribution of one normal random vector given another normal random vector. It is this result that underlies many of the important distributional and independence properties of the normal and related distributions that are established in later chapters.

Consider random vectors $X_i \in (V_i, (\cdot, \cdot)_i)$, $i = 1, 2$, and assume that the random vector $\{X_1, X_2\}$ in the direct sum $V_1 \oplus V_2$ has a normal distribution with mean vector $\{\mu_1, \mu_2\} \in V_1 \oplus V_2$ and covariance given by

$$\text{Cov}(X) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}.$$

Thus $\mathcal{L}(X_i) = N(\mu_i, \Sigma_{ii})$ on $(V_i, (\cdot, \cdot)_i)$ for $i = 1, 2$. The conditional distribution of X_1 given $X_2 = x_2 \in V_2$ is described in the next result.

Proposition 3.13. Let $\mathcal{L}(X_1 | X_2 = x_2)$ denote the conditional distribution of X_1 given $X_2 = x_2$. Then, under the above normality assumptions,

$$\mathcal{L}(X_1 | X_2 = x_2) = N\big(\mu_1 + \Sigma_{12}\Sigma_{22}^{-}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{12}'\big).$$

Here, Σ_{22}^{-} denotes the generalized inverse of Σ_{22}.

Proof. The proof consists of calculating the conditional characteristic function of X_1 given $X_2 = x_2$. To do this, first note that $X_1 - \Sigma_{12}\Sigma_{22}^- X_2$ and X_2 are jointly normal on $V_1 \oplus V_2$ and are uncorrelated by Proposition 2.17. Thus $X_1 - \Sigma_{12}\Sigma_{22}^- X_2$ and X_2 are independent. Therefore, for $x \in V_1$,

$$\begin{aligned}
\phi(x) &\equiv \mathcal{E}\left(\exp[i(x, X_1)_1] | X_2 = x_2\right) \\
&= \mathcal{E}\left(\exp[i(x, X_1)_1 - i(x, \Sigma_{12}\Sigma_{22}^- X_2)_1 + i(x, \Sigma_{12}\Sigma_{22}^- X_2)_1] | X_2 = x_2\right) \\
&= \exp[i(x, \Sigma_{12}\Sigma_{22}^- x_2)_1]\,\mathcal{E}\left(\exp[i(x, X_1 - \Sigma_{12}\Sigma_{22}^- X_2)_1] | X_2 = x_2\right) \\
&= \exp[i(x, \Sigma_{12}\Sigma_{22}^- x_2)_1]\,\mathcal{E}\exp[i(x, X_1 - \Sigma_{12}\Sigma_{22}^- X_2)_1]
\end{aligned}$$

where the last equality follows from the independence of X_2 and $X_1 - \Sigma_{12}\Sigma_{22}^- X_2$. However, it is clear that

$$\mathcal{L}(X_1 - \Sigma_{12}\Sigma_{22}^- X_2) = N(\mu_1 - \Sigma_{12}\Sigma_{22}^-\mu_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{12}')$$

as $X_1 - \Sigma_{12}\Sigma_{22}^- X_2$ is normal on V_1 and has the given mean vector and covariance (Proposition 2.17). Thus

$$\begin{aligned}
\phi(x) &= \exp[i(x, \Sigma_{12}\Sigma_{22}^- x_2)_1]\exp[i(x, \mu_1 - \Sigma_{12}\Sigma_{22}^-\mu_2)_1] \\
&\qquad \times\exp[-\tfrac{1}{2}(x, (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{12}')_1 x)] \\
&= \exp[i(x, \mu_1 + \Sigma_{12}\Sigma_{22}^-(x_2 - \mu_2))_1 - \tfrac{1}{2}(x, (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{12}')x)_1].
\end{aligned}$$

The uniqueness of characteristic functions yields the desired conclusion. □

For normal random vectors, $X_i \in (V_i, (\cdot, \cdot)_i)$, $i = 1, 2$, Proposition 3.13 shows that the conditional mean of X_1 given $X_2 = x_2$ is an affine function of x_2 (affine means a linear transformation, plus a constant vector so zero does not necessarily get mapped into zero). In other words,

$$\mathcal{E}(X_1 | X_2 = x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^-(x_2 - \mu_2).$$

Further, the conditional covariance of X_1 does not depend on the value of X_2. Also, this conditional covariance is the same as the unconditional covariance of the normal random vector $X_1 - \Sigma_{12}\Sigma_{22}^- X_2$. Of course, the specification of the conditional mean vector and covariance specifies the conditional distribution of X_1 given $X_2 = x_2$ as this conditional distribution is normal.

◆ **Example 3.1.** Let $W_1, \ldots, W_n$ be independent coordinate random vectors in R^p where R^p has the usual inner product. Assume that $\mathcal{L}(W_i) = N(\mu, \Sigma)$ so $\mu \in R^p$ is the coordinate mean vector of each W_i and Σ is the $p \times p$ covariance matrix of each W_i. Form the random matrix $X \in \mathcal{L}_{p,n}$ with rows W_i', $i = 1, \ldots, n$. We know that

$$\mathcal{E}X = e\mu'$$

and

$$\operatorname{Cov}(X) = I_n \otimes \Sigma$$

where $e \in R^n$ is the vector of ones. To show X has a normal distribution on the inner product space $(\mathcal{L}_{p,n}, \langle \cdot, \cdot \rangle)$, it must be verified that for each $A \in \mathcal{L}_{p,n}$, $\langle A, X \rangle$ has a normal distribution. To do this, let the rows of A be $a_1', \ldots, a_n'$, $a_i \in R^p$. Then

$$\langle A, X \rangle = \operatorname{tr} AX' = \sum_1^n a_i' W_i.$$

However, $a_i'W_i$ has a normal distribution on R since $\mathcal{L}(W_i) = N(\mu, \Sigma)$ on R^p. Also, since $W_1, \ldots, W_n$ are independent, $a_1'W_1, \ldots, a_n'W_n$ are independent. Since a linear combination of independent normal random variables is normal, $\langle A, X \rangle$ has a normal distribution for each $A \in \mathcal{L}_{p,n}$. Thus

$$\mathcal{L}(X) = N(e\mu', I_n \otimes \Sigma)$$

on the inner product space $(\mathcal{L}_{p,n}, \langle \cdot, \cdot \rangle)$. We now want to describe the conditional distribution of the first q columns of X given the last r columns of X where $q + r = p$. After some relabeling and a bit of manipulation, this conditional distribution follows from Proposition 3.13. Partition each W_i into Y_i and Z_i where $Y_i \in R^q$ consists of the first q coordinates of W_i and $Z_i \in R^r$ consists of the last r coordinates of W_i. Let $X_1 \in \mathcal{L}_{q,n}$ have rows $Y_1', \ldots, Y_n'$ and let $X_2 \in \mathcal{L}_{r,n}$ have rows $Z_1', \ldots, Z_n'$. Also, partition μ into $\mu_1 \in R^q$ and $\mu_2 \in R^r$ so $\mathcal{E}Y_i = \mu_1$ and $\mathcal{E}Z_i = \mu_2$, $i = 1, \ldots, n$. Further, partition the covariance matrix Σ of each W_i so that

$$\operatorname{Cov}\{Y_i, Z_i\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{21} = \Sigma_{12}'$. From the independence of $W_1, \ldots, W_n$, it follows that

$$\mathcal{L}(X_1) = N(e\mu_1', I_n \otimes \Sigma_{11}),$$

$$\mathcal{L}(X_2) = N(e\mu_2', I_n \otimes \Sigma_{22})$$

and $\{X_1, X_2\}$ has a normal distribution on $\mathcal{L}_{q,n} \oplus \mathcal{L}_{r,n}$ with mean vector $\{e\mu_1', e\mu_2'\}$ and

$$\operatorname{Cov}\{X_1, X_2\} = \begin{pmatrix} I_n \otimes \Sigma_{11} & I_n \otimes \Sigma_{12} \\ I_n \otimes \Sigma_{21} & I_n \otimes \Sigma_{22} \end{pmatrix}.$$

Now, Proposition 3.13 is directly applicable to $\{X_1, X_2\}$ where we make the parameter correspondence

$$\mu_i \leftrightarrow e\mu_i', \qquad i = 1,2$$

and

$$\Sigma_{ij} \leftrightarrow I_n \otimes \Sigma_{ij}.$$

Therefore, the conditional distribution of X_1 given $X_2 = x_2 \in \mathcal{L}_{r,n}$ is normal with mean vector

$$\mathcal{E}(X_1 | X_2 = x_2) = e\mu_1' + (I_n \otimes \Sigma_{12})(I_n \otimes \Sigma_{22})^{-}(x_2 - e\mu_2')$$

and

$$\operatorname{Cov}(X_1 | X_2 = x_2)$$
$$= I_n \otimes \Sigma_{11} - (I_n \otimes \Sigma_{12})(I_n \otimes \Sigma_{22})^{-}(I_n \otimes \Sigma_{21}).$$

However, it is not difficult to show that $(I_n \otimes \Sigma_{22})^{-} = I_n \otimes \Sigma_{22}^{-}$. Using the manipulation rules for Kronecker products, we have

$$\mathcal{E}(X_1 | X_2 = x_2) = e\mu_1' + (x_2 - e\mu_2')\Sigma_{22}^{-}\Sigma_{21}$$

and

$$\operatorname{Cov}(X_1 | X_2 = x_2) = I_n \otimes (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-}\Sigma_{21}).$$

This result is used in a variety of contexts in later chapters. ◆

3.5. THE DENSITY OF THE NORMAL DISTRIBUTION

The problem considered here is how to define the density function of a nonsingular normal distribution on an inner product space $(V, (\cdot, \cdot))$. By nonsingular, we mean that the covariance of the distribution is nonsingular. To motivate the technical considerations given below, the density function of a nonsingular normal distribution is first given for the standard coordinate space R^n with the usual inner product.

Consider a random vector X in R^n with coordinates $X_1, \ldots, X_n$ and assume that $X_1, \ldots, X_n$ are independent with $\mathcal{L}(X_i) = N(0, 1)$. The symbol dx denotes Lebesgue measure on R^n. Since $X_1, \ldots, X_n$ are independent, the joint density of $X_1, \ldots, X_n$ in R^n is just the product of the marginal densities, that is, X has a density with respect to dx given by

$$p(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}x_i^2\right] = \frac{1}{(2\pi)^{n/2}} \exp\left[-\tfrac{1}{2}\Sigma_1^n x_i^2\right]$$

where $x \in R^n$ has coordinates $x_1, \ldots, x_n$. Thus

$$p(x) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}x'x\right]$$

and $x'x$ is just the inner product of x with x in R^n. To derive the density of an arbitrary nonsingular normal distribution in R^n, let A be an $n \times n$ nonsingular matrix and set $Y = AX + \mu$ where $\mu \in R^n$. Since $\mathcal{L}(X) = N(0, I_n)$, $\mathcal{L}(Y) = N(\mu, \Sigma)$ where $\Sigma = AA'$ is positive definite. Thus $X = A^{-1}(Y - \mu)$ and the Jacobian of the nonsingular linear transformation on R^n to R^n sending x into $A^{-1}(x - \mu)$ is $|\det(A^{-1})|$ where $|\cdot|$ denotes absolute value. Therefore, the density function of Y with respect to dy is

$$\begin{aligned} p_1(y) &= |\det(A^{-1})| p(A^{-1}(y - \mu)) = (\det \Sigma)^{-1/2}(2\pi)^{-n/2} \\ &\quad \times \exp\left[-\tfrac{1}{2}(y - \mu)'A'^{-1}A^{-1}(y - \mu)\right] \\ &= (\det \Sigma)^{-1/2}(2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu)\right]. \end{aligned}$$

Thus we have the density function with respect to dy of any nonsingular normal distribution on R^n. Of course, this expression makes no sense when Σ is singular.

Now, suppose Y is a random vector in an n-dimensional vector space $(V, (\cdot, \cdot))$ and $\mathcal{L}(Y) = N(\mu, \Sigma)$ where Σ is positive definite. The expression

$$p_2(y) = (2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp\left[-\tfrac{1}{2}((y - \mu), \Sigma^{-1}(y - \mu))\right],$$

for $y \in V$, certainly makes sense and it is tempting to call this the density function of $Y \in (V, (\cdot, \cdot))$. The problem is: What is the measure on $(V, (\cdot, \cdot))$ with respect to which p_2 is a density? In other words, what is the analog of Lebesgue measure on $(V, (\cdot, \cdot))$? To answer the question, we now show that there is a natural measure on $(V, (\cdot, \cdot))$, which is constructed from Lebesgue measure on R^n, and p_2 is the density function of Y with respect to this measure.

The details of the construction of "Lebesgue measure" on an n- dimensional inner product space $(V, (\cdot, \cdot))$ follow. First, we review some basic topological notions for $(V, (\cdot, \cdot))$. Recall that $S_r(x_0) \equiv \{x | \|x - x_0\| < r\}$ is called the open ball of radius r with center x_0. A set $B \subseteq V$ is called *open* if, for each $x_0 \in B$, there is an $r > 0$ such that $S_r(x_0) \subseteq B$. Since all inner products on V are related by positive definite linear transformations, the definition of open does not depend on the given inner product. A set is *closed* iff its complement is open and a set if *bounded* iff it is contained in $S_r(0)$ for some $r > 0$. Just as in R^n, a set is compact iff it is closed and bounded (see Rudin, 1953, for the definition and characterization of compact sets in R^n). As with openness, the definitions and characterizations of closedness, boundedness, and compactness do not depend on the particular inner product on V. Let l denote standard Lebesgue measure on R^n. To move l over to the space V, let $x_1, \dots, x_n$ be a fixed orthonormal basis in $(V, (\cdot, \cdot))$ and define the linear transformation T on R^n to V by

$$T(a) = \Sigma_1^n a_i x_i$$

where $a \in R^n$ has coordinates $a_1, \dots, a_n$. Clearly, T is one-to-one, onto, and maps open, closed, bounded, and compact sets of R^n into open, closed, bounded, and compact sets of V. Also, T^{-1} on V to R^n maps $x \in V$ into the vector with coordinates (x_i, x), $i = 1, \dots, n$. Now, define the measure ν_0 on Borel sets $B \in \mathfrak{B}(V)$ by

$$\nu_0(B) = l(T^{-1}(B)).$$

Notice that $\nu_0(B + x) = l(T^{-1}(B + x)) = l(T^{-1}(B) + T^{-1}x) = l(T^{-1}(B)) = \nu_0(B)$ since Lebesgue measure is invariant under translations. Also, $\nu_0(B) < +\infty$ if B is a compact set. This leads to the following definition.

Definition 3.3. A nonzero measure ν defined on the Borel sets $\mathfrak{B}(V)$ of $(V, (\cdot, \cdot))$ is *invariant* if:

(i) $\nu(B + x) = \nu(B)$ for $x \in V$ and $B \in \mathfrak{B}(V)$.

(ii) $\nu(B) < +\infty$ for all compact sets B.

The measure ν_0 defined above is invariant and it is shown that, if ν is any invariant measure on $\mathcal{B}(V)$, then $\nu = c\nu_0$ for some constant $c > 0$. Condition (ii) of Definition 3.3 relates the topology of V to the measure ν. The measure that counts the number of points in a set satisfies (i) but not (ii) of Definition 3.3 and this measure is not equal to a positive constant times ν_0.

Before characterizing the measure ν_0, it is now shown that ν_0 is a dominating measure for the density function of a nonsingular normal distribution on $(V, (\cdot, \cdot))$.

Proposition 3.14. Suppose $\mathcal{L}(Y) = N(\mu, \Sigma)$ on the inner product space $(V, (\cdot, \cdot))$ where Σ is nonsingular. The density function of Y with respect to the measure ν_0 is given by

$$p(y) = (2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp\left[-\tfrac{1}{2}(y - \mu, \Sigma^{-1}(y - \mu))\right]$$

for $y \in V$.

Proof. It must be shown that, for each Borel set B,

$$P\{Y \in B\} = \int I_B(y) p(y) \nu_0(dy),$$

where I_B is the indicator function of the set B. From the definition of the measure ν_0, it follows that (see Lehmann, 1959, p. 38)

$$\int I_B(y) p(y) \nu_0(dy) = \int I_B(T(a)) p(T(a)) l(da).$$

Let $X = T^{-1}(Y) \in R^n$ so X is a random vector with coordinates (x_i, Y), $i = 1, \ldots, n$. Thus X has a normal distribution in R^n with mean vector $T^{-1}(\mu)$ and covariance matrix $[\Sigma]$ where $[\Sigma]$ is the matrix of Σ in the given orthonormal basis $x_1, \ldots, x_n$. Therefore,

$$P\{Y \in B\} = P\{T^{-1}(Y) \in T^{-1}(B)\} = P\{X \in T^{-1}(B)\}$$

$$= \int I_{T^{-1}(B)}(a)(2\pi)^{-n/2}(\det[\Sigma])^{-1/2}$$

$$\times \exp\left[-\tfrac{1}{2}(a - T^{-1}(\mu))'[\Sigma]^{-1}(a - T^{-1}(\mu))\right] l(da)$$

$$= \int I_B(T(a)) p(T(a)) l(da).$$

The last equality follows since $I_{T^{-1}(B)}(a) = I_B(T(a))$ and

$$p(T(a)) = (2\pi)^{-n/2}(\det \Sigma)^{-1/2} \times \exp\left[-\tfrac{1}{2}(T(a)-\mu, \Sigma^{-1}(T(a)-\mu))\right]$$
$$= (2\pi)^{-n/2}(\det[\Sigma])^{-1/2} \times \exp\left[-\tfrac{1}{2}(a - T^{-1}(\mu))'[\Sigma]^{-1}(a - T^{-1}(\mu))\right].$$

Thus

$$P\{Y \in B\} = \int I_B(T(a))p(T(a))l(da)$$
$$= \int I_B(y)p(y)\nu_0(dy). \qquad \square$$

We now want to show that the measure ν_0, constructed from Lebesgue measure on R^n, is the unique translation invariant measure that satisfies

$$\int p(y)\nu_0(dy) = 1.$$

Let $\mathcal{K}^+$ be the collection of all bounded non-negative Borel measurable functions defined on V that satisfy the following: given $f \in \mathcal{K}^+$, there is a compact set B such that $f(v) = 0$ if $v \notin B$. If ν is any invariant measure on V and $f \in \mathcal{K}^+$, then $\int f(v)\nu(dv) < +\infty$ since f is bounded and the ν-measure of every compact set is finite. It is clear that, if ν_1 and ν_2 are invariant measures such that

$$\int f(v)\nu_1(dv) = \int f(v)\nu_2(dv) \qquad \text{for all } f \in \mathcal{K}^+,$$

then $\nu_1 = \nu_2$. From the definition of an invariant measure, we also have

$$\int f(v+x)\nu(dv) = \int f(v)\nu(dv)$$

for all $f \in \mathcal{K}^+$ and $x \in V$. Furthermore, the definition of ν_0 shows that

$$\int f(x)\nu_0(dx) = \int f(T(a))l(da) = \int f(T(-a))l(da)$$

$$= \int f(-T(a))l(da) = \int f(-x)\nu_0(dx)$$

for all $f \in \mathcal{K}^+$. Here, we have used the linearity of T and the invariance of Lebesgue measure under multiplication of the argument of integration by a minus one.

Proposition 3.15. If ν is an invariant measure on $\mathcal{B}(V)$, then there exists a positive constant c such that $\nu = c\nu_0$.

Proof. For $f, g \in \mathcal{K}^+$, we have

$$\int f(x)\nu(dx)\int g(y)\nu_0(dy) = \int\int f(x-y)g(y)\nu(dx)\nu_0(dy)$$

$$= \int\int f(-(y-x))g(y-x+x)\nu_0(dy)\nu(dx)$$

$$= \int\int f(-w)g(w+x)\nu_0(dw)\nu(dx)$$

$$= \int\int f(-w)g(w+x)\nu(dx)\nu_0(dw)$$

$$= \int f(-w)\nu_0(dw)\int g(x)\nu(dx)$$

$$= \int f(w)\nu_0(dw)\int g(x)\nu(dx).$$

Therefore,

$$\int f(x)\nu(dx)\int g(y)\nu_0(dy) = \int f(w)\nu_0(dw)\int g(y)\nu(dy)$$

for all $f, g \in \mathcal{K}^+$. Fix $f \in \mathcal{K}^+$ such that $\int f(w)\nu_0(dw) = 1$ and set $c = \int f(x)\nu(dx)$. Then

$$\int g(y)\nu(dy) = c\int g(y)\nu_0(dy)$$

for all $g \in \mathcal{K}^+$. The constant c cannot be zero as the measure ν is not zero. Thus $c > 0$ and $\nu = c\nu_0$. □

The measure ν_0 is called the Lebesgue measure on V and is henceforth denoted by dv or dx, as is the Lebesgue measure on R^n. It is possible to show that ν_0 does not depend on the particular orthonormal basis used to define it by using a Jacobian argument in R^n. However, the argument given above contains more information than this. In fact, some minor technical modifications of the proof of Proposition 3.15 yield the uniqueness (up to a positive constant) of invariant measures on locally compact topological groups. This topic is discussed in detail in Chapter 6.

An application of Proposition 3.14 to the situation treated in Example 3.1 follows.

◆ **Example 3.2.** For independent coordinate random vectors $W_i \in R^p$, $i = 1,\ldots, n$, with $\mathcal{L}(W_i) = N(\mu, \Sigma)$, form the random matrix $X \in \mathcal{L}_{p,n}$ with rows W_i', $i = 1,\ldots, n$. As shown in Example 3.1,

$$\mathcal{L}(X) = N(e\mu', I_n \otimes \Sigma)$$

on the inner product space $(\mathcal{L}_{p,n}, \langle\cdot,\cdot\rangle)$, where $e \in R^n$ is the vector of ones. Let dX denote Lebesgue measure on the vector space $\mathcal{L}_{p,n}$. If Σ is nonsingular, then $I_n \otimes \Sigma$ is nonsingular and $(I_n \otimes \Sigma)^{-1} = I_n \otimes \Sigma^{-1}$. Thus when Σ is nonsingular, the density of X with respect to dX is

$$(3.1)\quad p(X) = (\sqrt{2\pi})^{-np}(\det(I_n \otimes \Sigma))^{-1/2} \times \exp\left[-\tfrac{1}{2}\langle X - e\mu', (I_n \otimes \Sigma^{-1})(X - e\mu')\rangle\right].$$

It is shown in Chapter 5 that $\det(I_n \otimes \Sigma) = (\det \Sigma)^n$. Since the inner product $\langle\cdot,\cdot\rangle$ is given by the trace, the density p can be written

$$p(X) = (\sqrt{2\pi})^{-np}(\det \Sigma)^{-n/2} \times \exp\left[-\tfrac{1}{2}\operatorname{tr}(X - e\mu')'(X - e\mu')\Sigma^{-1}\right].$$

However, this form of the density is somewhat less revealing, from a statistical point of view, than (3.1). In order to make this statement more precise and to motivate some future statistical considerations, we now think of $\mu \in R^p$ and Σ as unknown parameters. Thus, we can write (3.1) as

$$(3.2)\quad p(X|\mu, \Sigma) = \left(\sqrt{2\pi}\right)^{-np} (\det \Sigma)^{-n/2} \times \exp\left[-\tfrac{1}{2}\langle X - e\mu', \left(I_n \otimes \Sigma^{-1}\right)(X - e\mu')\rangle\right]$$

where μ ranges over R^p and Σ ranges over all $p \times p$ positive definite matrices. Thus we have a parametric family of densities for the distribution of the random vector X. As a first step in analyzing this parametric family, let

$$M = \left\{x \in \mathcal{L}_{p,n} | x = e\mu', \mu \in R^p\right\}.$$

It is clear that M is a p-dimensional linear subspace of $\mathcal{L}_{p,n}$ and M is simply the space of possible values for the mean vector of X. Let $P_e = (1/n)ee'$ so P_e is the orthogonal projection onto span$\{e\} \subseteq R^n$. Thus $P_e \otimes I_p$ is an orthogonal projection and it is easily verified that the range of $P_e \otimes I_p$ is M. Therefore, the orthogonal projection onto M is $P_e \otimes I_p$. Let $Q_e = I_n - P_e$ so $Q_e \otimes I_p$ is the orthogonal projection onto $M^\perp$ and $(Q_e \otimes I_p)(P_e \otimes I_p) = 0$. We now decompose X into the part of X in M and the part of X in $M^\perp$—that is, write $X = (P_e \otimes I_p)X + (Q_e \otimes I_p)X$. Substituting this into the exponential part of (3.2) and using the relation $(P_e \otimes I_p)(I_n \otimes \Sigma)(Q_e \otimes I_p) = 0$, we have

$$\begin{aligned}
&\langle X - e\mu', \left(I_n \otimes \Sigma^{-1}\right)(X - e\mu')\rangle \\
&\quad = \left\langle P_e(X - e\mu'), \left(I_n \otimes \Sigma^{-1}\right)P_e(X - e\mu')\right\rangle \\
&\qquad + \left\langle Q_e X, \left(I_n \otimes \Sigma^{-1}\right)Q_e X\right\rangle \\
&\quad = \langle P_e X - e\mu', \left(I_n \otimes \Sigma^{-1}\right)(P_e X - e\mu')\rangle + \operatorname{tr} Q_e X \Sigma^{-1}(Q_e X)' \\
&\quad = \langle P_e X - e\mu', \left(I_n \otimes \Sigma^{-1}\right)(P_e X - e\mu')\rangle + \operatorname{tr} X' Q_e X \Sigma^{-1}.
\end{aligned}$$

Thus the density $p(X|\mu, \Sigma)$ is a function of the pair $P_e X$ and $X'Q_e X$ so $P_e X$ and $X'Q_e X$ is a sufficient statistic for the parametric family (3.2). Proposition 3.4 shows that $(P_e \otimes I_p)X$ and $(Q_e \otimes I_p)X$ are independent since $(P_e \otimes I_p)(I_n \otimes \Sigma)(Q_e \otimes I_p) = (P_e Q_e) \otimes \Sigma = 0$ as $P_e Q_e = 0$. Therefore, $P_e X$ and $X'Q_e X$ are independent since $P_e X = (P_e \otimes I_p)X$ and $X'Q_e X = ((Q_e \otimes I_p)X)'((Q_e \otimes I_p)X)$. To interpret the sufficient statistic in terms of the original random vectors $W_1, \dots, W_n$, first note that

$$P_e X = \frac{1}{n} ee'X = e\overline{W}'$$

where $\overline{W} = (1/n)\Sigma W_i$ is the sample mean. Also,

$$\begin{aligned} X'Q_e X &= (Q_e X)'(Q_e X) = ((I_n - P_e)X)'((I - P_e)X) \\ &= (X - e\overline{W}')'(X - e\overline{W}') = \Sigma_1^n (W_i - \overline{W})(W_i - \overline{W})'. \end{aligned}$$

The quantity $(1/n)X'Q_e X$ is often called the sample covariance matrix. Since $e\overline{W}'$ and $\overline{W}$ are one-to-one functions of each other, we have that the sample mean and sample covariance matrix form a sufficient statistic and they are independent. It is clear that

$$\mathcal{L}(\overline{W}) = N\left(\mu, \frac{1}{n}\Sigma\right).$$

The distribution of $X'Q_e X$, commonly called the Wishart distribution, is derived later. The procedure of decomposing X into the projection onto the mean space (the subspace M) and the projection onto the orthogonal complement of the mean space is fundamental in multivariate analysis as in univariate statistical analysis. In fact, this procedure is at the heart of analyzing linear models—a topic to be considered in the next chapter. ◆

PROBLEMS

1. Suppose $X_1, \dots, X_n$ are independent with values in $(V, (\cdot, \cdot))$ and $\mathcal{L}(X_i) = N(\mu_i, A_i)$, $i = 1, \dots, n$. Show that $\mathcal{L}(\Sigma X_i) = N(\Sigma\mu_i, \Sigma A_i)$.

2. Let X and Y be random vectors in R^n with a joint normal distribution given by

$$\mathcal{L}\begin{pmatrix} X \\ Y \end{pmatrix} = N\left(0, \begin{pmatrix} I_n & \rho I_n \\ \rho I_n & I_n \end{pmatrix}\right)$$

where ρ is a scalar. Show that $|\rho| \leqslant 1$ and the covariance is positive definite iff $|\rho| < 1$. Let $Q(Y) = I_n - (Y'Y)^{-1}YY'$. Prove that $W = X'Q(Y)X$ has the distribution of $(1 - \rho^2)\chi^2_{n-1}$ (the constant $1 - \rho^2$ times a chi-squared random variable with $n - 1$ degrees of freedom).

3. When $X \in R^n$ and $\mathcal{L}(X) = N(0, \Sigma)$ with Σ nonsingular, then $\mathcal{L}(X) = \mathcal{L}(CZ)$ where $\mathcal{L}(Z) = N(0, I_n)$ and $CC' = \Sigma$. Hence, $\mathcal{L}(C^{-1}X) = \mathcal{L}(Z)$ so C^{-1} transforms X into a vector of i.i.d. $N(0,1)$ random variables. There are many C^{-1}'s that do this. The problem at hand concerns the construction of one such C^{-1}. Given any $p \times p$ positive definite matrix A, $p \geqslant 2$, partition A as

$$A = \begin{pmatrix} a_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where $a_{11} \in R^1$, $A_{21} = A'_{12} \in R^{p-1}$. Define $T_p(A)$ by

$$T_p(A) = \begin{pmatrix} a_{11}^{-1/2} & 0 \\ -\dfrac{A_{21}}{a_{11}} & I_{p-1} \end{pmatrix}.$$

(i) Partition $\Sigma : n \times n$ as A is partitioned and set $X^{(1)} = T_n(\Sigma)X$. Show that

$$\text{Cov}(X^{(1)}) = \begin{pmatrix} 1 & 0 \\ 0 & \Sigma^{(1)} \end{pmatrix}$$

where $\Sigma^{(1)} = \Sigma_{22} - \Sigma_{21}\Sigma_{12}/\sigma_{11}$.

(ii) For $k = 1, 2, \ldots, n-2$, define $X^{(k+1)}$ by

$$X^{(k+1)} = \begin{pmatrix} I_k & 0 \\ 0 & T_{n-k}(\Sigma^{(k)}) \end{pmatrix} X^{(k)}.$$

Prove that

$$\text{Cov}(X^{(k+1)}) = \begin{pmatrix} I_{k+1} & 0 \\ 0 & \Sigma^{(k+1)} \end{pmatrix}$$

for some positive definite $\Sigma^{(k+1)}$.

(iii) For $k = 0, \ldots, n-2$, let

$$T^{(k)} = \begin{pmatrix} I_k & 0 \\ 0 & T_{n-k}(\Sigma^{(k)}) \end{pmatrix},$$

where $T^{(0)} = T_n(\Sigma)$. With $T = T^{(n-2)} \ldots T^{(0)}$, show that $X^{(n-1)} = TX$ and $\text{Cov}(X^{(n-1)}) = I_n$. Also, show that T is lower triangular and $\Sigma^{-1} = T'T$.

4. Suppose $X \in R^2$ has coordinates X_1 and X_2, and has a density

$$p(x) = \begin{cases} \dfrac{1}{\pi} \exp\left[-\frac{1}{2}(x_1^2 + x_2^2)\right] & \text{if } x_1 x_2 > 0 \\ 0 & \text{otherwise} \end{cases}$$

so p is zero in the second and fourth quadrants. Show X_1 and X_2 are both normal but X is not normal.

5. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables. Show that $U = \Sigma\Sigma(X_i - X_j)^2$ and $W = \Sigma X_i$ are independent. What is the distribution of U?

6. For $X \in (V, (\cdot, \cdot))$ with $\mathcal{L}(X) = N(0, I)$, suppose (X, AX) and (X, BX) are independent. If A and B are both positive semidefinite, prove that $AB = 0$. Hint: Show that $\operatorname{tr} AB = 0$ by using $\operatorname{cov}\{(X, AX), (X, BX)\} = 0$. Then use the positive semidefiniteness and $\operatorname{tr} AB = 0$ to conclude that $AB = 0$.

7. The method used to define the normal distribution on $(V, (\cdot, \cdot))$ consisted of three steps: (i) first, an $N(0, 1)$ distribution was defined on R^1; (ii) next, if $\mathcal{L}(Z) = N(0, 1)$, then W is $N(\mu, \sigma^2)$ if $\mathcal{L}(W) = \mathcal{L}(\sigma Z + \mu)$; and (iii) X with values in $(V, (\cdot, \cdot))$ is normal if (x, X) is normal on R^1 for each $x \in V$. It is natural to ask if this procedure can be used to define other types of distributions on $(V, (\cdot, \cdot))$. Here is an attempt for the Cauchy distribution. For $X \in R^1$, say Z is standard Cauchy (which we write as $\mathcal{L}(Z) = C(0, 1)$) if the density of Z is

$$p(z) = \frac{1}{\pi} \frac{1}{1 + z^2}, \qquad z \in R^1.$$

Say W has a Cauchy distribution on R^1 if $\mathcal{L}(W) = \mathcal{L}(\sigma Z + \mu)$ for some $\mu \in R^1$ and $\sigma \geqslant 0$—in this case write $\mathcal{L}(W) = C(\mu, \sigma)$. Finally, say $X \in (V, (\cdot, \cdot))$ is Cauchy if (x, X) is Cauchy on R^1.

(i) Let $W_1, \ldots, W_n$ be independent $C(\mu_j, \sigma_j)$, $j = 1, \ldots, n$. Show that $\mathcal{L}(\Sigma a_j W_j) = C(\Sigma a_j \mu_j, \Sigma |a_j| \sigma_j)$. Hint: The characteristic function of a $C(0, 1)$ distribution is $\exp[-|t|]$, $t \in R^1$.

(ii) Let $Z_1, \ldots, Z_n$ be i.i.d. $C(0, 1)$ and let $x_1, \ldots, x_n$ be any basis for $(V, (\cdot, \cdot))$. Show $X = \Sigma Z_j x_j$ has a Cauchy distribution on $(V, (\cdot, \cdot))$.

8. Consider a density on R^1 given by

$$f(u) = \int_0^\infty t^{-1} \phi(u/t) G(dt)$$

where ϕ is the density of an $N(0, 1)$ distribution and G is a distribution function with $G(0) = 0$. The distribution defined by f is called a *scale mixture of normals*.

(i) Let Z_0 be $N(0, 1)$ and let R be independent of Z_0 with $\mathcal{L}(R) = G$. Show that $U = RZ_0$ has f as its density function.

If $\mathcal{L}(Y) = \mathcal{L}(cU)$ for some $c > 0$, we can say that Y has a *type-f distribution*.

(ii) In $(V, (\cdot, \cdot))$, suppose $\mathcal{L}(Z) = N(0, I)$ and form $X = RZ$ where R and Z are independent and $\mathcal{L}(R) = G$. For each $x \in V$, show (x, X) has a type-f distribution.

Remark. The distribution of X in $(V, (\cdot, \cdot))$ provides a possible vector space generalization of a type-f distribution on R^1.

9. In the notation of Example 3.1, assume that $\mu = 0$ so $\mathcal{L}(X) = N(0, I_n \otimes \Sigma)$ on $(\mathcal{L}_{p,n}, \langle \cdot , \cdot \rangle)$. Also,

$$\mathcal{L}(X_1 | X_2 = x_2) = N\left(x_2 \Sigma_{22}^{-1} \Sigma_{21}, I_n \otimes \Sigma_{11\cdot 2}\right)$$

where $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. Show that the conditional distribution of $X_2'X_1$ given X_2 is the same as the conditional distribution of $X_2'X_1$ given $X_2'X_2$.

10. The map T of Section 3.5 has been defined on R^n to $(V, (\cdot, \cdot))$ by $Ta = \sum_1^n a_i x_i$ where $x_1, \ldots, x_n$ is an orthonormal basis for $(V, (\cdot, \cdot))$. Also, we have defined ν_0 by $\nu_0(B) = l(T^{-1}(B))$ for $B \in \mathcal{B}(V)$. Consider another orthonormal basis $y_1, \ldots, y_n$ for $(V(\cdot, \cdot))$ and define T_1 by $T_1 a = \sum_1^n a_i y_i$, $a \in R^n$. Define ν_1 by $\nu_1(B) = l(T_1^{-1}(B))$ for $B \in \mathcal{B}(V)$. Prove that $\nu_0 = \nu_1$.

11. The measure ν_0 in Problem 10 depends on the inner product $(\cdot, \cdot)$ on V. Suppose $[\cdot, \cdot]$ is another inner product given by $[x, y] = (x, Ay)$ where $A > 0$. Let ν_1 be the measure constructed on $(V, [\cdot, \cdot])$ in the same manner that ν_0 was constructed on $(V, (\cdot, \cdot))$. Show that $\nu_1 = c\nu_0$ where $c = (\det(A))^{1/2}$.

12. Consider the space $\mathcal{S}_p$ of $p \times p$ symmetric matrices with the inner product given by $\langle S_1, S_2 \rangle = \operatorname{tr} S_1 S_2$. Show that the density function of an $N(0, I)$ distribution on $(\mathcal{S}_p, \langle \cdot , \cdot \rangle)$ with respect to the measure ν_0 is

$$p(S) = (2\pi)^{-p(p+1)/4} \exp\left[-\tfrac{1}{2}\left(\sum_1^p s_{ii}^2 + 2\sum_{i<j}\sum s_{ij}^2\right)\right]$$

where $S = \{s_{ij}\}$, $i, j = 1, \ldots, p$. Explain your answer (what is ν_0)?

13. Consider $X_1,\ldots, X_n$, which are i.i.d. $N(\mu, \Sigma)$ on R^p. Let $X \in \mathcal{L}_{p,n}$ have rows $X_1',\ldots, X_n'$ so $\mathcal{L}(X) = N(e\mu', I_n \otimes \Sigma)$. Assume that Σ has the form

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

where $\sigma^2 > 0$ and $-1/(p-1) < \rho < 1$ so Σ is positive definite. Such a covariance matrix is said to have intraclass covariance structure.

(i) On R^p, let $A = (1/p)e_1e_1'$ where $e_1 \in R^p$ is the vector of ones. Show that a positive definite covariance matrix has intraclass covariance structure iff $\Sigma = \alpha A + \beta(I - A)$ for some positive scalars α and β. In this case $\Sigma^{-1} = \alpha^{-1}A + \beta^{-1}(I - A)$.

(ii) Using the notation and methods of Example 3.2, show that when (μ, σ^2, ρ) are unknown parameters, then $(\bar{X}, \operatorname{tr} AX'Q_eX, \operatorname{tr}(I - A)X'Q_eX)$ is a sufficient statistic.

NOTES AND REFERENCES

1. A coordinate treatment of the normal distribution similar to the treatment given here can be found in Muirhead (1982).
2. Examples 3.1 and 3.2 indicate some of the advantages of vector space techniques over coordinate techniques. For comparison, the reader may find it instructive to formulate coordinate versions of these examples.
3. The converse of Proposition 3.11 is true. The only proof I know involves characteristic functions. For a discussion of this, see Srivastava and Khatri (1979, p. 64).