Mathematical Statistics 1

Chapter 3. Some Special Distributions

3.5. The Multivariate Normal Distribution—Proofs of Theorems

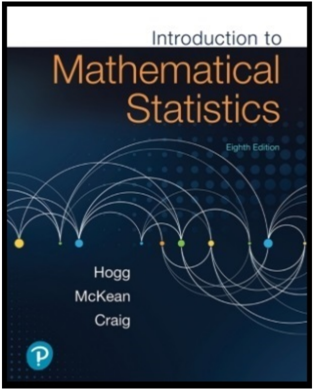

Table of contents

Lemma 3.5.A

Lemma 3.5.A. Let random vector (X, Y) have the bivariate normal distribution. Then X and Y are independent if and only if they are uncorrelated (that is, $\rho = 0$).

Proof. The joint moment generating function of (X, Y) is (by Note 3.5.B)

$$M_{(X,Y)}(t_1, t_2) = \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(t_1^2\sigma_1^2 + 2t_1t_2\rho\sigma_1\sigma_2 + t_2^2\sigma_2^2)\right).$$

Lemma 3.5.A

Lemma 3.5.A. Let random vector (X, Y) have the bivariate normal distribution. Then X and Y are independent if and only if they are uncorrelated (that is, $\rho = 0$).

Proof. The joint moment generating function of (X, Y) is (by Note 3.5.B)

$$M_{(X,Y)}(t_1, t_2) = \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(t_1^2\sigma_1^2 + 2t_1t_2\rho\sigma_1\sigma_2 + t_2^2\sigma_2^2)\right).$$

If $\rho = 0$ then the joint moment generating function becomes

$$M_{(X,Y)}(t_1, t_2) = \exp\left(t_1\mu_1 + t_2\mu_2 + t_1^2\sigma_1^2/2 + t_2^2\sigma_2^2/2\right)$$

$$= \exp\left(t_1\mu_1 + t_1^2\sigma_2^2/2\right)\exp\left(t_2\mu_2 + t_2\sigma_2^2/2\right) = M_{(X,Y)}(t_1, 0)M_{(X,Y)}(0, t_2).$$

So by Theorem 2.4.5, X and Y are independent.

Lemma 3.5.A

Lemma 3.5.A. Let random vector (X, Y) have the bivariate normal distribution. Then X and Y are independent if and only if they are uncorrelated (that is, $\rho = 0$).

Proof. The joint moment generating function of (X, Y) is (by Note 3.5.B)

$$M_{(X,Y)}(t_1, t_2) = \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(t_1^2\sigma_1^2 + 2t_1t_2\rho\sigma_1\sigma_2 + t_2^2\sigma_2^2)\right).$$

If $\rho = 0$ then the joint moment generating function becomes

$$M_{(X,Y)}(t_1, t_2) = \exp\left(t_1\mu_1 + t_2\mu_2 + t_1^2\sigma_1^2/2 + t_2^2\sigma_2^2/2\right)$$

$$= \exp\left(t_1\mu_1 + t_1^2\sigma_2^2/2\right)\exp\left(t_2\mu_2 + t_2\sigma_2^2/2\right) = M_{(X,Y)}(t_1, 0)M_{(X,Y)}(0, t_2).$$

So by Theorem 2.4.5, X and Y are independent.

Conversely, Suppose X and Y are independent. The by Theorem 2.4.5, $M_{(X,Y)}(t_1, t_2) = M_{(X,Y)}(t_1, 0)M_{(X,Y)}(0, t_2)$ and so the form of the joint moment generating function $M_{(X,Y)}(t_1, t_2)$ given above, we must have $\rho = 0$, as claimed. □

Lemma 3.5.A

Lemma 3.5.A. Let random vector (X, Y) have the bivariate normal distribution. Then X and Y are independent if and only if they are uncorrelated (that is, $\rho = 0$).

Proof. The joint moment generating function of (X, Y) is (by Note 3.5.B)

$$M_{(X,Y)}(t_1, t_2) = \exp\left(t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(t_1^2\sigma_1^2 + 2t_1t_2\rho\sigma_1\sigma_2 + t_2^2\sigma_2^2)\right).$$

If $\rho = 0$ then the joint moment generating function becomes

$$M_{(X,Y)}(t_1, t_2) = \exp\left(t_1\mu_1 + t_2\mu_2 + t_1^2\sigma_1^2/2 + t_2^2\sigma_2^2/2\right)$$

$$= \exp\left(t_1\mu_1 + t_1^2\sigma_2^2/2\right)\exp\left(t_2\mu_2 + t_2\sigma_2^2/2\right) = M_{(X,Y)}(t_1, 0)M_{(X,Y)}(0, t_2).$$

So by Theorem 2.4.5, X and Y are independent.

Conversely, Suppose X and Y are independent. The by Theorem 2.4.5, $M_{(X,Y)}(t_1, t_2) = M_{(X,Y)}(t_1, 0)M_{(X,Y)}(0, t_2)$ and so the form of the joint moment generating function $M_{(X,Y)}(t_1, t_2)$ given above, we must have $\rho = 0$, as claimed. □

Theorem 3.5.1

Theorem 3.5.1. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma}$ is positive definite. Then the random variable $Y = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}(\mathbf{X} - \boldsymbol{\mu})$ has a $\chi^2(n)$ distribution.

Proof. Since $\mathbf{X} = \Sigma^{1/2}\mathbf{Z} + \mu$ then

$$
\begin{aligned}
Y &= (\mathbf{X} - \mu)'\Sigma^{-1}(\mathbf{X} - \mu) = (\Sigma^{1/2}\mathbf{Z})'\Sigma^{-1}(\Sigma^{1/2}\mathbf{Z}) \\
&= \mathbf{Z}'\Sigma^{1/2}\Sigma^{-1}\Sigma^{1/2}\mathbf{Z} \text{ since } \Sigma^{1/2} \text{ is symmetric} \\
&= \mathbf{Z}'\mathbf{Z} = \sum_{i=1}^{n} Z_i^2.
\end{aligned}
$$

Now Z_i^2 has a χ^2 distribution by Theorem 2.4.1. So by Corollary 3.3.1, $Y = \sum_{i=1}^{n} Z_i^2$ has a $\chi^2(n)$ distribution, as claimed. □

Theorem 3.5.1

Theorem 3.5.1. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma}$ is positive definite. Then the random variable $Y = (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}(\mathbf{X} - \boldsymbol{\mu})$ has a $\chi^2(n)$ distribution.

Proof. Since $\mathbf{X} = \boldsymbol{\Sigma}^{1/2}\mathbf{Z} + \boldsymbol{\mu}$ then

$$\begin{aligned} Y &= (\mathbf{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu}) = (\boldsymbol{\Sigma}^{1/2}\mathbf{Z})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Sigma}^{1/2}\mathbf{Z}) \\ &= \mathbf{Z}'\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{1/2}\mathbf{Z} \text{ since } \boldsymbol{\Sigma}^{1/2} \text{ is symmetric} \\ &= \mathbf{Z}'\mathbf{Z} = \sum_{i=1}^{n} Z_i^2. \end{aligned}$$

Now Z_i^2 has a χ^2 distribution by Theorem 2.4.1. So by Corollary 3.3.1, $Y = \sum_{i=1}^{n} Z_i^2$ has a $\chi^2(n)$ distribution, as claimed. □

Theorem 3.5.2

Theorem 3.5.2. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{Y}$ has a $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ distribution.

Proof. The moment generating function of $\mathbf{Y}$ is

$$\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= E[\exp(\mathbf{t}'\mathbf{Y})] = E[\exp(\mathbf{t}'((A)\mathbf{X} + \mathbf{b}))] \\
&= E[\exp(\mathbf{t}'\mathbf{AX} + \mathbf{t}'\mathbf{b})] = E[\exp(\mathbf{t}'\mathbf{b})\exp(\mathbf{t}'\mathbf{AX})] \\
&= \exp(\mathbf{t}'\mathbf{b})E[\exp(\mathbf{t}'\mathbf{AX})] = \exp(\mathbf{t}'\mathbf{b})E[\exp((\mathbf{A}'\mathbf{t})'\mathbf{X})] \\
&= \exp(\mathbf{t}'\mathbf{b})\exp\left((\mathbf{A}'\mathbf{t})'\mu + \frac{1}{2}(\mathbf{A}'\mathbf{t})'\Sigma(\mathbf{A}'\mathbf{t})\right) \text{ by Definition 3.5.1} \\
&= \exp\left((\mathbf{t}'\mathbf{b}) + \mathbf{t}'\mathbf{A}\mu + \frac{1}{2}\mathbf{t}'\mathbf{A}\Sigma\mathbf{A}'\mathbf{t}\right) \\
&= \exp\left(\mathbf{t}'(\mathbf{A}\mu + \mathbf{b}) + \frac{1}{2}\mathbf{t}'\mathbf{A}\Sigma\mathbf{A}'\mathbf{t}\right) \ldots
\end{aligned}$$

Theorem 3.5.2

Theorem 3.5.2. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{Y}$ has a $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A\Sigma A}')$ distribution.

Proof. The moment generating function of $\mathbf{Y}$ is

$$\begin{aligned}
M_{\mathbf{Y}}(\mathbf{t}) &= E[\exp(\mathbf{t}'\mathbf{Y})] = E[\exp(\mathbf{t}'((A)\mathbf{X} + \mathbf{b}))] \\
&= E[\exp(\mathbf{t}'\mathbf{AX} + \mathbf{t}'\mathbf{b})] = E[\exp(\mathbf{t}'\mathbf{b})\exp(\mathbf{t}'\mathbf{AX})] \\
&= \exp(\mathbf{t}'\mathbf{b})E[\exp(\mathbf{t}'\mathbf{AX})] = \exp(\mathbf{t}'\mathbf{b})E[\exp((\mathbf{A}'\mathbf{t})'\mathbf{X})] \\
&= \exp(\mathbf{t}'\mathbf{b})\exp\left((\mathbf{A}'\mathbf{t})'\boldsymbol{\mu} + \frac{1}{2}(\mathbf{A}'\mathbf{t})'\Sigma(\mathbf{A}'\mathbf{t})\right) \text{ by Definition 3.5.1} \\
&= \exp\left((\mathbf{t}'\mathbf{b}) + \mathbf{t}'\mathbf{A}\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\mathbf{A}\Sigma\mathbf{A}'\mathbf{t}\right) \\
&= \exp\left(\mathbf{t}'(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}) + \frac{1}{2}\mathbf{t}'\mathbf{A}\Sigma\mathbf{A}'\mathbf{t}\right) \ldots
\end{aligned}$$

Theorem 3.5.2 (continued)

Theorem 3.5.2. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$. Then $\mathbf{Y}$ has a $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ distribution.

Proof. ...

$$M_{\mathbf{Y}}(\mathbf{t}) = \exp\left(\mathbf{t}'(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}) + \frac{1}{2}\mathbf{t}'\mathbf{A}\Sigma\mathbf{A}'\mathbf{t}\right),$$

which is the moment generating function of an $N_m(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}')$ distribution, as claimed. □

Corollary 3.5.1

Corollary 3.5.1. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution partitioned as

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

where $\mathbf{X}_1$ and $\boldsymbol{\mu}_1$ are m dimensional and $\boldsymbol{\Sigma}_{\mathbf{11}}$ is $m \times m$. Then $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution.

Proof. Define $m \times (m+p)$ matrix $\mathbf{A} = [\mathbf{I}_m \; \mathbf{0}_{mp}]$ where $\mathbf{0}_{mp}$ is an $m \times p$ matrix of zeros. Then $\mathbf{X}_1 = \mathbf{AX}$ (notice that $\mathbf{A}$ is $m \times (m+p)$ and $\mathbf{X}$ is $(m+p) \times 1$ so $\mathbf{X}_1 = m \times 1$). So with $\mathbf{b} = \mathbf{0}$, we have by Theorem 3.5.2 that $\mathbf{X}_1$ has a $N_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}')$ distribution. Now $\mathbf{A}\boldsymbol{\mu} = \boldsymbol{\mu}_1$ and writing $\mathbf{A}\Sigma\mathbf{A}'$ in terms of partitioned matrices gives

$$\mathbf{A}\Sigma\mathbf{A}' = [\mathbf{I}_m \; \mathbf{0}_{mp}] \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{mp} \end{bmatrix} = \Sigma_{11}$$

(notice that Σ_{11} is a matrix itself so we do not write $[\Sigma_{11}]$).

Corollary 3.5.1

Corollary 3.5.1. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution partitioned as

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

where $\mathbf{X}_1$ and $\boldsymbol{\mu}_1$ are m dimensional and $\boldsymbol{\Sigma}_{11}$ is $m \times m$. Then $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution.

Proof. Define $m \times (m+p)$ matrix $\mathbf{A} = [\mathbf{I}_m \ \mathbf{0}_{mp}]$ where $\mathbf{0}_{mp}$ is an $m \times p$ matrix of zeros. Then $\mathbf{X}_1 = \mathbf{AX}$ (notice that $\mathbf{A}$ is $m \times (m+p)$ and $\mathbf{X}$ is $(m+p) \times 1$ so $\mathbf{X}_1 = m \times 1$). So with $\mathbf{b} = \mathbf{0}$, we have by Theorem 3.5.2 that $\mathbf{X}_1$ has a $N_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ distribution. Now $\mathbf{A}\boldsymbol{\mu} = \boldsymbol{\mu}_1$ and writing $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$ in terms of partitioned matrices gives

$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = [\mathbf{I}_m \ \mathbf{0}_{mp}] \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{mp} \end{bmatrix} = \boldsymbol{\Sigma}_{11}$$

(notice that $\boldsymbol{\Sigma}_{11}$ is a matrix itself so we do not write $[\boldsymbol{\Sigma}_{11}]$).

Corollary 3.5.1 (continued)

Corollary 3.5.1. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution partitioned as

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

where $\mathbf{X}_1$ and $\boldsymbol{\mu}_1$ are m dimensional and $\boldsymbol{\Sigma}_{\mathbf{11}}$ is $m \times m$. Then $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution.

Proof. So $\mathbf{A}\boldsymbol{\mu} = \boldsymbol{\mu}_1$ and $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' = \boldsymbol{\Sigma}_{11}$. Hence $\mathbf{X}_1$ has a $N_m(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}') = N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution, as claimed. □

Theorem 3.5.3

Theorem 3.5.3. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

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

Then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if the covariance satisfies $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Proof. Since $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$ then $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$. By Definition 3.5.1, the moment generating function of $\mathbf{X}$ is

$$M_{\mathbf{X}}(\mathbf{t}) = \exp(\mathbf{t}'\boldsymbol{\mu} + (1/2)\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}) \text{ for } \mathbb{R}^n.$$

Since $\mathbf{t}' = [\mathbf{t}_1' \ \mathbf{t}_2']$ and $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ then $\mathbf{t}'\boldsymbol{\mu} = \mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2$.

Theorem 3.5.3

Theorem 3.5.3. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

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

Then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if the covariance satisfies $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Proof. Since $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$ then $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$. By Definition 3.5.1, the moment generating function of $\mathbf{X}$ is

$$M_{\mathbf{X}}(\mathbf{t}) = \exp(\mathbf{t}'\boldsymbol{\mu} + (1/2)\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}) \text{ for } \mathbb{R}^n.$$

Since $\mathbf{t}' = [\mathbf{t}_1' \ \mathbf{t}_2']$ and $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ then $\mathbf{t}'\boldsymbol{\mu} = \mathbf{t}_1'\boldsymbol{\mu}_1 + \mathbf{t}_2'\boldsymbol{\mu}_2$.

Theorem 3.5.3 (continued 1)

Proof (continued). Also,

$$\begin{aligned}
\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t} &= [\mathbf{t}_1' \ \mathbf{t}_2'] \left[\begin{array}{cc} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{array} \right] \left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{array} \right] \\
&= [\mathbf{t}_1'\boldsymbol{\Sigma}_{11} + \mathbf{t}_2\boldsymbol{\Sigma}_{21} \ \ \mathbf{t}_1'\boldsymbol{\Sigma}_{12} + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}] \left[\begin{array}{c} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{array} \right] \\
&= \mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}_2\boldsymbol{\Sigma}_{21}\mathbf{t}_1 + \mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2 + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2.
\end{aligned}$$

By Corollary 3.5.1, $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution and (similarly) $\mathbf{X}_2$ has a $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ distribution. So by Definition 3.5.1, the marginal distribution functions are $M_{\mathbf{X}_1}(\mathbf{t}_1) = \exp(\mathbf{t}_1'\boldsymbol{\mu}_1 + (1/2)\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1)$ and $M_{\mathbf{X}_2}(\mathbf{t}_2) = \exp(\mathbf{t}_2'\boldsymbol{\mu}_2 + (1/2)\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2)$ for $[\mathbf{t}_1' \ \mathbf{t}_2'] \in \mathbb{R}^n$. By Note 2.6.C (and its observation that Theorem 2.4.5 can be extended to several random variables) we have that $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}_1}(\mathbf{t}_1)M_{\mathbf{X}_2}(\mathbf{t}_2)$.

Theorem 3.5.3 (continued 1)

Proof (continued). Also,

$$\begin{aligned}
\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t} &= [\mathbf{t}_1' \ \mathbf{t}_2'] \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \\
&= [\mathbf{t}_1'\boldsymbol{\Sigma}_{11} + \mathbf{t}_2\boldsymbol{\Sigma}_{21} \ \ \mathbf{t}_1'\boldsymbol{\Sigma}_{12} + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}] \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \\
&= \mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}_2\boldsymbol{\Sigma}_{21}\mathbf{t}_1 + \mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2 + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2.
\end{aligned}$$

By Corollary 3.5.1, $\mathbf{X}_1$ has a $N_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ distribution and (similarly) $\mathbf{X}_2$ has a $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ distribution. So by Definition 3.5.1, the marginal distribution functions are $M_{\mathbf{X}_1}(\mathbf{t}_1) = \exp(\mathbf{t}_1'\boldsymbol{\mu}_1 + (1/2)\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1)$ and $M_{\mathbf{X}_2}(\mathbf{t}_2) = \exp(\mathbf{t}_2'\boldsymbol{\mu}_2 + (1/2)\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2)$ for $[\mathbf{t}_1' \ \mathbf{t}_2'] \in \mathbb{R}^n$. By Note 2.6.C (and its observation that Theorem 2.4.5 can be extended to several random variables) we have that $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}_1}(\mathbf{t}_1)M_{\mathbf{X}_2}(\mathbf{t}_2)$.

Theorem 3.5.3 (continued 2)

Theorem 3.5.3. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

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

Then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if the covariance satisfies $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Proof (continued). If $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, so that $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}' = \mathbf{0}'$, then $M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}_1}(\mathbf{t}_1)M_{\mathbf{X}_2}(\mathbf{t}_2)$ and so by Note 2.6.C $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, as claimed. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then by Note 2.6.C $M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}_1}(\mathbf{t}_1)M_{\mathbf{X}_2}(\mathbf{t}_2)$ and so $\mathbf{t}_2'\boldsymbol{\Sigma}_{21}\mathbf{t}_1 = 0 = \mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2$ for all $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \in \mathbb{R}^n$. So we must have $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{21} = \mathbf{0}'$, as claimed. □

Theorem 3.5.3 (continued 2)

Theorem 3.5.3. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

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

Then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if the covariance satisfies $\boldsymbol{\Sigma}_{12} = \mathbf{0}$.

Proof (continued). If $\boldsymbol{\Sigma}_{12} = \mathbf{0}$, so that $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}' = \mathbf{0}'$, then $M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}_1}(\mathbf{t}_1) M_{\mathbf{X}_2}(\mathbf{t}_2)$ and so by Note 2.6.C $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, as claimed. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then by Note 2.6.C $M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}_1}(\mathbf{t}_1) M_{\mathbf{X}_2}(\mathbf{t}_2)$ and so $\mathbf{t}_2' \boldsymbol{\Sigma}_{21} \mathbf{t}_1 = 0 = \mathbf{t}_1' \boldsymbol{\Sigma}_{12} \mathbf{t}_2$ for all $\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \in \mathbb{R}^n$. So we must have $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{21} = \mathbf{0}'$, as claimed. □

Theorem 3.5.4

Theorem 3.5.4. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

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

Assume that $\boldsymbol{\Sigma}$ is positive definite. Then the conditional distribution of $\mathbf{X}_1 \mid \mathbf{X}_2$ is

$$N_m(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Proof. Define random variable $\mathbf{W} = \mathbf{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2$. Then

$$\begin{bmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

Theorem 3.5.4

Theorem 3.5.4. Suppose $\mathbf{X}$ has a $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, partitioned as

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

Assume that $\boldsymbol{\Sigma}$ is positive definite. Then the conditional distribution of $\mathbf{X}_1 \mid \mathbf{X}_2$ is

$$N_m(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}).$$

Proof. Define random variable $\mathbf{W} = \mathbf{X}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2$. Then

$$\begin{bmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.$$

Theorem 3.5.4 (continued 1)

Proof (continued). By Theorem 3.5.2 (with $\mathbf{A} = \begin{bmatrix} \mathbf{I}_m & -\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}$ and $\mathbf{b} = \mathbf{0}$) we have that $\begin{bmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{bmatrix}$ has a multivariate normal distribution $N_n(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\Sigma\mathbf{A}')$ where

$$\mathbf{A}' = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}' \\ -(\mathbf{\Sigma}_{22}^{-1})'\mathbf{\Sigma}_{12}' & \mathbf{I}_p \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ -\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} & \mathbf{I}_p \end{bmatrix}$$

since $(M^{-1})' = (M')^{-1}$ (see Theorem 3.3.7 in my online notes for Theory of Matrices [MATH 5090] on Section 3.3. Matrix Rank and the Inverse of a Full Rank Matrix). Since

$$\mathbf{A}\mu = \begin{bmatrix} \mathbf{I}_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{bmatrix},$$

then the means are $E[\mathbf{W}] = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2$ and $E[\mathbf{X}_2] = \mu_2$.

Theorem 3.5.4 (continued 1)

Proof (continued). By Theorem 3.5.2 (with $\mathbf{A} = \begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix}$ and $\mathbf{b} = \mathbf{0}$) we have that $\begin{bmatrix} \mathbf{W} \\ \mathbf{X}_2 \end{bmatrix}$ has a multivariate normal distribution $N_n(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$ where

$$\mathbf{A}' = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}' \\ -(\boldsymbol{\Sigma}_{22}^{-1})'\boldsymbol{\Sigma}_{12}' & \mathbf{I}_p \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_p \end{bmatrix}$$

since $(M^{-1})' = (M')^{-1}$ (see Theorem 3.3.7 in my online notes for Theory of Matrices [MATH 5090] on Section 3.3. Matrix Rank and the Inverse of a Full Rank Matrix). Since

$$\mathbf{A}\boldsymbol{\mu} = \begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 \end{bmatrix},$$

then the means are $E[\mathbf{W}] = \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2$ and $E[\mathbf{X}_2] = \boldsymbol{\mu}_2$.

Theorem 3.5.4 (continued 2)

Proof (continued). The covariance matrix is

$$\mathbf{A}\Sigma\mathbf{A}' = \begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}' \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_p \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}' \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_p \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Since we have a matrix of all 0's in the upper right, then by Theorem 3.5.3 the random vectors $\mathbf{W}$ and $\mathbf{X}_2$ are independent. By Note 2.4.1, if the joint probability density function of $\mathbf{W}$ and $\mathbf{X}_2$ is $f(\mathbf{w}, \mathbf{x}_2)$ then the conditional probability density functions are $f_{\mathbf{W}|\mathbf{X}_2}(\mathbf{w} \mid \mathbf{x}_2) = f(\mathbf{w}, \mathbf{x}_2)/f(\mathbf{x}_2)$ and $f_{\mathbf{X}_2|\mathbf{W}}(\mathbf{w} \mid \mathbf{x}_2) = f(\mathbf{w}, \mathbf{x}_2)/f_1(\mathbf{w})$ where the marginal distributions are $f_1(\mathbf{w})$ and $f_2(\mathbf{x}_2)$. By Definition 2.4.1, since $\mathbf{W}$ and $\mathbf{X}_2$ are independent, then $f_{\mathbf{X}_2|\mathbf{W}}(\mathbf{w} \mid \mathbf{x}_2) = f_1(\mathbf{w})f_2(\mathbf{x}_2)$ (though Note 2.4.1 and Definition 2.4.1 deal with single random variables instead of random vectors).

Theorem 3.5.4 (continued 2)

Proof (continued). The covariance matrix is

$$\mathbf{A}\Sigma\mathbf{A}' = \begin{bmatrix} \mathbf{I}_m & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \mathbf{0} & \mathbf{I}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}' \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_p \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_m & \mathbf{0}' \\ -\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{I}_p \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \mathbf{0}' \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}.$$

Since we have a matrix of all 0's in the upper right, then by Theorem 3.5.3 the random vectors $\mathbf{W}$ and $\mathbf{X}_2$ are independent. By Note 2.4.1, if the joint probability density function of $\mathbf{W}$ and $\mathbf{X}_2$ is $f(\mathbf{w}, \mathbf{x}_2)$ then the conditional probability density functions are $f_{\mathbf{W}|\mathbf{X}_2}(\mathbf{w} \mid \mathbf{x}_2) = f(\mathbf{w}, \mathbf{x}_2)/f(\mathbf{x}_2)$ and $f_{\mathbf{X}_2|\mathbf{W}}(\mathbf{w} \mid \mathbf{x}_2) = f(\mathbf{w}, \mathbf{x}_2)/f_1(\mathbf{w})$ where the marginal distributions are $f_1(\mathbf{w})$ and $f_2(\mathbf{x}_2)$. By Definition 2.4.1, since $\mathbf{W}$ and $\mathbf{X}_2$ are independent, then $f_{\mathbf{X}_2|\mathbf{W}}(\mathbf{w} \mid \mathbf{x}_2) = f_1(\mathbf{w})f_2(\mathbf{x}_2)$ (though Note 2.4.1 and Definition 2.4.1 deal with single random variables instead of random vectors).

Theorem 3.5.4 (continued 3)

Proof (continued). So the conditional probability density function of $\mathbf{W} \mid \mathbf{X}_2$ is equal to the marginal density function:

$$f_{\mathbf{W}|\mathbf{X}_2}(\mathbf{x}_2 \mid \mathbf{w}) = \frac{f(\mathbf{w}, \mathbf{x}_2)}{f_2(\mathbf{x}_2)} = \frac{f_1(\mathbf{w}) f_2(\mathbf{x}_2)}{f_2(\mathbf{x}_2)} = f_1(\mathbf{w}).$$

Since $E[\mathbf{W}] = \boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2$ and the variance of $\mathbf{W}$ is $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, then the marginal distribution of $\mathbf{W}$ (and also the conditional distribution of $\mathbf{W} \mid \mathbf{X}_2$) is $N_m(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$. Now $\mathbf{X}_1 = \mathbf{W} + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2$ and so (again by the independence) the distribution of $\mathbf{X}_1 \mid \mathbf{X}_2$ is $N_m(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2 + \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$, as claimed. □

Theorem 3.5.4 (continued 3)

Proof (continued). So the conditional probability density function of $\mathbf{W} \mid \mathbf{X}_2$ is equal to the marginal density function:

$$f_{\mathbf{W}|\mathbf{X}_2}(\mathbf{x}_2 \mid \mathbf{w}) = \frac{f(\mathbf{w}, \mathbf{x}_2)}{f_2(\mathbf{x}_2)} = \frac{f_1(\mathbf{w}) f_2(\mathbf{x}_2)}{f_2(\mathbf{x}_2)} = f_1(\mathbf{w}).$$

Since $E[\mathbf{W}] = \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2$ and the variance of $\mathbf{W}$ is $\boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$, then the marginal distribution of $\mathbf{W}$ (and also the conditional distribution of $\mathbf{W} \mid \mathbf{X}_2$) is $N_m(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$. Now $\mathbf{X}_1 = \mathbf{W} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2$ and so (again by the independence) the distribution of $\mathbf{X}_1 \mid \mathbf{X}_2$ is $N_m(\boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{X}_2, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21})$, as claimed. □

Exercise 3.5.8

Exercise 3.5.8. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 20$, $\mu_2 = 40$, $\sigma_1^2 = 9$, $\sigma_2^2 = 4$, and $\rho = 0.6$. Find the shortest interval for which 0.90 is the conditional probability that Y is in the interval, given that $x = 22$.

Solution. As seen in Example 3.5.A, the conditional distribution of Y gives $X = 22$ is

$$N(\mu_2 + (\rho\sigma_1/\sigma_2)(x - \mu_1), \sigma_2^2(1 - \rho^2))$$

$$= N((40) + ((0.6)(3)/(2)((22) - (20)), (4)(1 - (0.6)^2)) = N(41.8, 2.56).$$

So the mean is 41.2 and the standard deviation is $\sqrt{1.56} = 1.6$. To get a ("two-sided") interval centered at 41.8 that contains 0.90 of the distribution, we take the Z-value of $Z = 1.645$ and the interval is

$$((41.8) - (1.645)(1.6), (41.8) + (1.645)(1.6)) = (39.168, 44.432). \qquad \square$$

Exercise 3.5.8

Exercise 3.5.8. Let X and Y have a bivariate normal distribution with parameters $\mu_1 = 20$, $\mu_2 = 40$, $\sigma_1^2 = 9$, $\sigma_2^2 = 4$, and $\rho = 0.6$. Find the shortest interval for which 0.90 is the conditional probability that Y is in the interval, given that $x = 22$.

Solution. As seen in Example 3.5.A, the conditional distribution of Y gives $X = 22$ is

$$N(\mu_2 + (\rho\sigma_1/\sigma_2)(x - \mu_1), \sigma_2^2(1 - \rho^2))$$

$$= N((40) + ((0.6)(3)/(2)((22) - (20)), (4)(1 - (0.6)^2)) = N(41.8, 2.56).$$

So the mean is 41.2 and the standard deviation is $\sqrt{1.56} = 1.6$. To get a ("two-sided") interval centered at 41.8 that contains 0.90 of the distribution, we take the Z-value of $Z = 1.645$ and the interval is

$$((41.8) - (1.645)(1.6), (41.8) + (1.645)(1.6)) = (39.168, 44.432). \qquad \square$$

Lemma 3.5.B

Lemma 3.5.B. Consider random vector $\mathbf{X}$ with multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ where $\boldsymbol{\Gamma}$ is an orthogonal positive definite matrix. Then for any $\mathbf{a} \in \mathbb{R}^n$ with $\|\mathbf{a}\| = 1$, we have $\mathrm{Var}(\mathbf{a}'\mathbf{X}) \le \mathrm{Var}(Y_1)$. That is, Y_1 has the maximum variance of any linear combination $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})$ where $\|\mathbf{a}\| = \|\mathbf{a}'\| = 1$.

Proof. The first component of $\mathbf{Y}$ is given by $Y_1 = \mathbf{v}_1'(\mathbf{X} - \boldsymbol{\mu})$ where $\mathbf{v}_1$ is the first column of $\boldsymbol{\Gamma}'$ (and hence the first row of $\boldsymbol{\Gamma}$); since $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}'$ are orthogonal, then $\|\mathbf{v}_1\|^2 = \sum_{j=1}^n v1j^2 = 1$. For $\mathbf{a} \in \mathbb{R}^n$ with $\|\mathbf{a}\| = 1$, we have $\mathbf{a} = \sum_{j=1}^n a_j \mathbf{v}_j$ where $\mathbf{v}_j$ is the jth column of $\boldsymbol{\Gamma}'$ (since $\boldsymbol{\Gamma}'$ is orthogonal and so its columns for an orthonormal set of n vectors in $\mathbb{R}^n$; i.e., $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$).

Lemma 3.5.B

Lemma 3.5.B. Consider random vector $\mathbf{X}$ with multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ where $\boldsymbol{\Gamma}$ is an orthogonal positive definite matrix. Then for any $\mathbf{a} \in \mathbb{R}^n$ with $\|\mathbf{a}\| = 1$, we have $\mathrm{Var}(\mathbf{a}'\mathbf{X}) \leq \mathrm{Var}(Y_1)$. That is, Y_1 has the maximum variance of any linear combination $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})$ where $\|\mathbf{a}\| = \|\mathbf{a}'\| = 1$.

Proof. The first component of $\mathbf{Y}$ is given by $Y_1 = \mathbf{v}_1'(\mathbf{X} - \boldsymbol{\mu})$ where $\mathbf{v}_1$ is the first column of $\boldsymbol{\Gamma}'$ (and hence the first row of $\boldsymbol{\Gamma}$); since $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}'$ are orthogonal, then $\|\mathbf{v}_1\|^2 = \sum_{j=1}^n v1j^2 = 1$. For $\mathbf{a} \in \mathbb{R}^n$ with $\|\mathbf{a}\| = 1$, we have $\mathbf{a} = \sum_{j=1}^n a_j \mathbf{v}_j$ where $\mathbf{v}_j$ is the jth column of $\boldsymbol{\Gamma}'$ (since $\boldsymbol{\Gamma}'$ is orthogonal and so its columns for an orthonormal set of n vectors in $\mathbb{R}^n$; i.e., $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$).

Lemma 3.5.B (continued 1)

Proof (continued). Since $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma} = \sum_{i=1}^n \lambda_i \mathbf{v}_i\mathbf{v}_i'$ (see Note 3.5.D and Exercise 3.5.19) then

$$\begin{aligned}
\mathrm{Var}(\mathbf{a}'\mathbf{X}) &= \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \text{ by Theorem 3.5.2} \\
&= \mathbf{a}'\boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{a} \text{ since } \boldsymbol{\Sigma} = \boldsymbol{\Gamma}'\boldsymbol{\Lambda}\boldsymbol{\Gamma} \\
&= \left(\sum_{i=1}^n a_i\mathbf{v}_i\right)\boldsymbol{\Lambda}\left(\sum_{j=1}^n a_j\mathbf{v}_j'\right) \text{ since } \mathbf{a}'\boldsymbol{\Gamma}' \text{ is a linear} \\
&\quad \text{combination of the columns of } \boldsymbol{\Gamma}' \text{ with scalars } a_i, \\
&\quad \text{and } \boldsymbol{\Gamma}\mathbf{a} \text{ is a linear combination of the rows of } \boldsymbol{\Gamma} \\
&\quad \text{with scalars } a_i \text{ (notice that the rows of } \boldsymbol{\Gamma} \text{ are the} \\
&\quad \text{columns of } \boldsymbol{\Gamma}' \text{ transposed)} \\
&= \left(\sum_{i=1}^n \lambda_i a_i\mathbf{v}_i\right)\left(\sum_{j=1}^n a_j\mathbf{v}_j'\right) \text{ since } \boldsymbol{\Lambda} \text{ is a diagonal matrix} \ldots
\end{aligned}$$

Lemma 3.5.B (continued 2)

Proof (continued). ...

$$\begin{aligned}
\mathrm{Var}(\mathbf{a}'\mathbf{X}) &= \left(\sum_{i=1}^{n} \lambda_i a_i \mathbf{v}_i\right)\left(\sum_{j=1}^{n} a_j \mathbf{v}_j'\right) \text{ since } \boldsymbol{\Lambda} \text{ is a diagonal matrix} \\
&= \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} a_i a_j \mathbf{v}_i \mathbf{v}_j' = \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} a_i a_j (\mathbf{v}_i \cdot \mathbf{v}_j) \\
&= \sum_{i=1}^{n} \lambda_i a_i^2 \text{ since } \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \text{ is an orthonormal set} \\
&\leq \lambda_1 \sum_{i=1}^{n} a_i^2 \text{ since } \lambda_1 \text{ is the greatest eigenvalue} \\
&= \lambda_1 \text{ since} \sum_{i=1}^{n} a_i^2 = \|\mathbf{a}\|^1 = 1 \\
&= \mathrm{Var}(Y_1).
\end{aligned}$$

Lemma 3.5.B (continued 3)

Lemma 3.5.B. Consider random vector $\mathbf{X}$ with multivariate normal distribution $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{Y} = \boldsymbol{\Gamma}(\mathbf{X} - \boldsymbol{\mu})$ where $\boldsymbol{\Gamma}$ is an orthogonal positive definite matrix. Then for any $\mathbf{a} \in \mathbb{R}^n$ with $\|\mathbf{a}\| = 1$, we have $\mathrm{Var}(\mathbf{a}'\mathbf{X}) \leq \mathrm{Var}(Y_1)$. That is, Y_1 has the maximum variance of any linear combination $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})$ where $\|\mathbf{a}\| = \|\mathbf{a}'\| = 1$.

Proof (continued). $\ldots \mathrm{Var}(\mathbf{a}'\mathbf{X}) \leq \mathrm{Var}(Y_1)$. So $\mathrm{Var}(Y_1) \geq \mathrm{Var}(\mathbf{a}'\mathbf{X})$ and hence Y_1 has the maximum variance of any linear combination $\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})$ where $\|\mathbf{a}'\| = 1$, as claimed. $\square$