Probability and Stochastic Processes - Lecture 12

Michael Evans
University of Toronto
http://www.utstat.utoronto.ca/mikevans/stac62/STAC622023.html

2023

- consider a random vector $\mathbf{X}$ with prob. measure $P_{\mathbf{X}}$ and suppose $\mathbf{Y} = \mathbf{y}$, where $\mathbf{Y} = T(\mathbf{X})$, is observed

- we want the conditional distribution of $\mathbf{X}$ given $T(\mathbf{X}) = \mathbf{y}$

discrete case

- so suppose $\mathbf{X}$ has a discrete distribution with probability function $p_{\mathbf{X}}$

- then the conditional probability function of $\mathbf{X}$ given $T(\mathbf{X}) = \mathbf{y}$ is 0 when $T(\mathbf{x}) \neq \mathbf{y}$ for no $\mathbf{x}$, where $p_{\mathbf{X}}(\mathbf{x}) > \mathbf{0}$ and otherwise

$$\begin{aligned} p_{\mathbf{X}\,|\,\mathbf{Y}}(\mathbf{x}\,|\,\mathbf{y}) &= P_{\mathbf{X}\,|\,\mathbf{Y}}(\{\mathbf{x}\}\,|\,T^{-1}\{\mathbf{y}\}) \\ &= \frac{P_{\mathbf{X}}(\{\mathbf{x}\} \cap T^{-1}\{\mathbf{y}\})}{P_{\mathbf{X}}(T^{-1}\{\mathbf{y}\})} \\ &= \frac{p_{\mathbf{X}}(\mathbf{x})}{\sum_{\mathbf{z} \in T^{-1}\{\mathbf{y}\}} p_{\mathbf{X}}(\mathbf{z})} = \frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{y})} \end{aligned}$$

Example II.8.1 *Conditioning the multinomial*$(n, p_1, \ldots, p_k)$

- suppose $Y = T(X_1, \ldots, X_k) = X_1 \sim$ binomial(n, p_1)

- now we want the conditional probability function of $(X_1, \ldots, X_k) \mid X_1 = x_1$ or equivalently $(X_2, \ldots, X_k) \mid X_1 = x_1$ so, for $x_2, \ldots, x_k \in \{0, \ldots, n - x_1\}, x_2 + \cdots + x_k = n - x_1$

$$\begin{aligned} p_{(X_2,\ldots,X_k)\,|\,X_1}(x_2, \ldots, x_k \mid x_1) &= \frac{\binom{n}{x_1\, x_2 \ldots x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}}{\binom{n}{x_1} p_1^{x_1} (1 - p_1)^{n - x_1}} \\ &= \frac{(n - x_1)!}{x_2! \cdots x_k!} \left(\frac{p_2}{1 - p_1}\right)^{x_2} \cdots \left(\frac{p_k}{1 - p_1}\right)^{x_k} \end{aligned}$$

- therefore, $(X_2, \ldots, X_k) \mid X_1 = x_1 \sim$ multinomial$\left(n - x_1, \frac{p_2}{1 - p_1}, \ldots, \frac{p_k}{1 - p_1}\right)$

■

Exercise II.8.1 If $\mathbf{X} \sim$ multinomial$(n, p_1, \ldots, p_k)$ and $Y = X_1 + \cdots + X_l$ for some $l \le k$, then determine the conditional distribution of $\mathbf{X}$ given $Y = y$.

absolutely continuous case

- so suppose $\mathbf{X}$ has a.c. distribution with density function $f_{\mathbf{X}}$ and $T : R^k \to R^l$ is smooth $(k \geq l)$

- then the conditional density function of $\mathbf{X}$ given $T(\mathbf{X}) = \mathbf{y}$ is, when $\mathbf{x} \in T^{-1}\{\mathbf{y}\}$

$$f_{\mathbf{X}\,|\,\mathbf{Y}}(\mathbf{x}\,|\,\mathbf{y}) = \lim_{\delta_1 \downarrow 0, \delta_2 \downarrow 0} \left\{ \frac{P_{\mathbf{X}}(B_{\delta_1}(\mathbf{x}) \cap T^{-1}B_{\delta_2}(\mathbf{y}))}{Vol(B_{\delta_1}(\mathbf{x}) \cap T^{-1}B_{\delta_2}(\mathbf{y}))} \Big/ \frac{P_{\mathbf{Y}}(B_{\delta_2}(\mathbf{y}))}{Vol(B_{\delta_2}(\mathbf{y}))} \right\}$$
$$\overset{\text{fact}}{=} \frac{f_{\mathbf{X}}(\mathbf{x})J_T(x)}{f_{\mathbf{Y}}(\mathbf{y})}$$

where

$$J_T(\mathbf{x}) = \left| \det \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial T_l(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_l(\mathbf{x})}{\partial x_k} \end{pmatrix} \begin{pmatrix} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial T_l(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_l(\mathbf{x})}{\partial x_k} \end{pmatrix}' \right|^{-1/2}$$

Example II.8.2 *Projections*

- if $T(x_1,\ldots,x_k) = (x_1,x_2)$ then $l = 2$

$$J_T(\mathbf{x}) = \left|\det\left(\begin{array}{ccc} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\mathbf{x})}{\partial x_k} \\ \frac{\partial T_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial T2(\mathbf{x})}{\partial x_k} \end{array}\right)\left(\begin{array}{cc} \frac{\partial T_1(\mathbf{x})}{\partial x_1} & \frac{\partial T_2(\mathbf{x})}{\partial x_1} \\ \vdots & \vdots \\ \frac{\partial T_1(\mathbf{x})}{\partial x_k} & \frac{\partial T_2(\mathbf{x})}{\partial x_k} \end{array}\right)\right|^{-1/2}$$

$$= \left|\det\left(\begin{array}{cccc} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \end{array}\right)\left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 0 \end{array}\right)\right|^{-1/2} = \left|\det\left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right)\right|^{-1/2} = 1$$

- also $f_{(X_1,X_2)}(x_1,x_2) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{(X_1,\ldots,X_k)}(x_1,\ldots,x_k)\,dx_3\cdots dx_k$ so

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

Exercise II.8.2 Repeat Example II.8.2 when $T(x_1,\ldots,x_k) = x_1$.

Example II.8.3 *Projection conditionals of the* $N_k(\boldsymbol{\mu}, \Sigma)$

- suppose $\mathbf{X} \sim N_k(\boldsymbol{\mu}, \Sigma)$ and $\mathbf{X}_1 = T(\mathbf{X}) = (X_1, \ldots, X_l)'$ for $l \leq k$

- partition $\boldsymbol{\mu}$ and Σ as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \text{ where } \boldsymbol{\mu}_1 \in R^l, \boldsymbol{\mu}_2 \in R^{k-l}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \text{ where } \begin{array}{ll} \Sigma_{11} \in R^{l \times l}, & \Sigma_{12} \in R^{l \times (k-l)}, \\ & \Sigma_{22} \in R^{(k-l) \times (k-l)} \end{array}$$

Exercise II.8.3 Prove that Σ_{11} and Σ_{22} are p.d. when Σ is p.d.

- we need to obtain the distribution of $\mathbf{X}_1$ and for this we need another matrix decomposition

Gram-Schmidt (QR) decomposition

- consider a matrix $A = (\mathbf{a}_1 \cdots \mathbf{a}_k) \in R^{k\times k}$ of rank k (so nonsingular)

- so the columns of A form a basis for R^k : $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly independent ($c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k = \mathbf{0}$ iff $c_1 = \ldots = c_k = 0$) and $L\{\mathbf{a}_1, \ldots, \mathbf{a}_k\} = \{c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k : c_1, \ldots, c_k \in R^1\} = R^k$

- applying the Gram-Schmidt process to $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ we obtain an orthonormal basis $\{\mathbf{q}_1, \ldots, \mathbf{q}_k\}$ for R^k

- then $Q = (\mathbf{q}_1 \cdots \mathbf{q}_k) \in R^{k\times k}$ is an orthogonal matrix and

$$
\begin{aligned}
A &= QR = (\mathbf{q}_1 \cdots \mathbf{q}_k)\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ 0 & r_{22} & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{kk} \end{pmatrix} \text{ and} \\
\mathbf{q}_1 &= \mathbf{a}_1/||\mathbf{a}_1||, r_{11} = ||\mathbf{a}_1|| > 0 \\
\mathbf{q}_2 &= \frac{\mathbf{a}_2 - (\mathbf{q}_1'\mathbf{a}_2)\mathbf{q}_1}{||\mathbf{a}_2 - \mathbf{q}_1'\mathbf{a}_2||}, r_{12} = \mathbf{q}_1'\mathbf{a}_2, r_{22} = ||\mathbf{a}_2 - (\mathbf{q}_1'\mathbf{a}_2)\mathbf{q}_1|| > 0 \\
&\ \vdots
\end{aligned}
$$

- so R is an upper triangular matrix with positive diagonals and it is unique

Exercise II.8.4 Prove that if $A = QR$ as just described, then R is unique given Q.

- applying the QR decomposition to $\Sigma^{1/2} = QR$ gives

$$\Sigma = \Sigma^{1/2}\Sigma^{1/2} = (\Sigma^{1/2})'\Sigma^{1/2} = R'R$$

and this is known as the *Cholesky decomposition* of Σ

Exercise II.8.5 Restrict to 2×2 matrices. (i) Prove that the product of two upper triangular 2×2 matrices with positive diagonals is upper triangular with positive diagonals. (ii) Prove that an upper triangular matrix with positive diagonal is nonsingular and its inverse is upper triangular with positive diagonal equal to the inverse of the diagonal elements of the original matrix. (iii) Show that the upper triangular matrix in the Cholesky decomposition of p.d. Σ is unique. (iv) (Challenge) Generalize (i), (ii) and (iii) to $k \times k$ upper triangular matrices.

- now

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} R_{11}' & 0 \\ R_{12}' & R_{22} \end{pmatrix}' \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}$$

- let $\mathbf{Z} \sim N_k(\mathbf{0}, I)$ and then using the fact that $Z_1, \ldots, Z_k$ are mut. stat. ind.

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \text{ where } \mathbf{Z}_1 \sim N_l(\mathbf{0}, I) \text{ stat. ind. of } \mathbf{Z}_2 \sim N_{k-l}(\mathbf{0}, I)$$

- now (recall $\mathbf{a} + A\mathbf{Z} \sim N_k(\mathbf{a}, AA')$)

$$\begin{aligned} \mathbf{X} &= \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} + \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}' \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} + \begin{pmatrix} R_{11}' & 0 \\ R_{12}' & R_{22}' \end{pmatrix} \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\mu}_1 + R_{11}'\mathbf{Z}_1 \\ R_{12}'\mathbf{Z}_1 + R_{22}'\mathbf{Z}_2 \end{pmatrix} \end{aligned}$$

- therefore $\mathbf{X}_1 = \boldsymbol{\mu}_1 + R_{11}'\mathbf{Z}_1 \sim N_l(\boldsymbol{\mu}_1, R_{11}'R_{11}) = N_l(\boldsymbol{\mu}_1, \Sigma_{11})$ and we have proved

Proposition II.8.1 If

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}\right)$$

where $\mathbf{X}_1 \in R^l$, then $\mathbf{X}_1 \sim N_l(\boldsymbol{\mu}_1, \Sigma_{11})$.

Exercise II.8.4 If I_r denotes an $r \times r$ identity matrix, then use a matrix of the form

$$C = \begin{pmatrix} 0 & I_{k-l} \\ I_l & 0 \end{pmatrix}$$

to determine the distribution of $\mathbf{X}_2$ in Proposition II.8.1. In general a permutation matrix A has a single 1 in each row and column with all the remaining entries 0. Use such a matrix to determine the marginal distribution of any subvector of $\mathbf{X}$.

Proposition II.8.2 If

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \right)$$

where $\mathbf{X}_1 \in R^l$, then

$$\mathbf{Y} = \mathbf{X}_2 - \Sigma_{12}'\Sigma_{11}^{-1}\mathbf{X}_1 \sim N_{k-l}(\boldsymbol{\mu}_2 - \Sigma_{12}'\Sigma_{11}^{-1}\boldsymbol{\mu}_1, \Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})$$

stat. ind. of $\mathbf{X}_1 \sim N_l(\boldsymbol{\mu}_1, \Sigma_{11})$.

Proof: We have

$$\begin{aligned} \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{Y} \end{pmatrix} &= \begin{pmatrix} I & 0 \\ -\Sigma_{12}'\Sigma_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} = A\mathbf{X} \\ &\sim N_k(A\boldsymbol{\mu}, A\Sigma A') \text{ where } (\textbf{Exercise II.8.6}) \\ A\boldsymbol{\mu} &= \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 - \Sigma_{12}'\Sigma_{11}^{-1}\boldsymbol{\mu}_1 \end{pmatrix} \\ A\Sigma A' &= \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix} \end{aligned}$$

which proves the first part.

Now observe in general, if

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right),$$

then

$$\begin{aligned} & f_{\mathbf{W}}(\mathbf{w}_1, \mathbf{w}_2) = (2\pi)^{-k/2}\left(\det\begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}\right)^{-1/2} \times \\ & \exp\left(-\frac{1}{2}\begin{pmatrix} \mathbf{w}_1 - \boldsymbol{\nu}_1 \\ \mathbf{w}_2 - \boldsymbol{\nu}_2 \end{pmatrix}' \begin{pmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 - \boldsymbol{\nu}_1 \\ \mathbf{w}_2 - \boldsymbol{\nu}_2 \end{pmatrix}\right) \\ = \; & (2\pi)^{-l/2}(\det \Sigma_{11})^{-1/2} \exp(-(\mathbf{w}_1 - \boldsymbol{\nu}_1)'\Sigma_{11}^{-1}(\mathbf{w}_1 - \boldsymbol{\nu}_1)/2) \times \\ & (2\pi)^{-(k-l)/2}(\det \Sigma_{22})^{-1/2} \exp(-(\mathbf{w}_2 - \boldsymbol{\nu}_2)'\Sigma_{22}^{-1}(\mathbf{w}_2 - \boldsymbol{\nu}_2)/2) \end{aligned}$$

and so $\mathbf{W}_1$ and $\mathbf{W}_2$ are statistically independent and this proves the second part. ■

Corollary II.8.2

$$\mathbf{X}_2 \mid \mathbf{X}_1 = \mathbf{x}_1 \sim N_{k-l}(\mu_2 + \Sigma_{12}'\Sigma_{11}^{-1}(\mathbf{x}_1 - \mu_1), \Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})$$

Proof: So the joint density of $\mathbf{X}_1$ and $\mathbf{Y}$ is $f_{\mathbf{X}_1}(\mathbf{x}_1)f_{\mathbf{Y}}(\mathbf{y})$. Now make the transformation

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = T\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{y}+\Sigma_{12}'\Sigma_{11}^{-1}\mathbf{x}_1 \end{pmatrix}$$

which has

$$J_T(\mathbf{x}_1, \mathbf{y}) = \left|\det\begin{pmatrix} I_l & 0 \\ stuff & I_{k-l} \end{pmatrix}\right|^{-1} = 1.$$

So by the change of variable

$$f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2) = f_{\mathbf{X}_1}(\mathbf{x}_1)f_{\mathbf{Y}}(\mathbf{x}_2 - \Sigma_{12}'\Sigma_{11}^{-1}\mathbf{x}_1)$$

and therefore by conditioning on projections

$$f_{\mathbf{X}_2 \mid \mathbf{X}_1}(\mathbf{x}_2 \mid \mathbf{x}_1) = \frac{f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2)}{f_{\mathbf{X}_1}(\mathbf{x}_1)} = f_{\mathbf{Y}}(\mathbf{x}_2 - \Sigma_{12}'\Sigma_{11}^{-1}\mathbf{x}_1)$$

and

$$
\begin{aligned}
& f_{\mathbf{Y}}(\mathbf{x}_2 - \Sigma_{12}'\Sigma_{11}^{-1}\mathbf{x}_1) \\
= \; & (2\pi)^{-(k-l)/2}(\det(\Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12}))^{-1/2} \times \\
& \exp\left(-\frac{(\mathbf{x}_2 - (\boldsymbol{\mu}_2 + \Sigma_{12}'\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)))'(\Sigma_{22} - \Sigma_{12}'\Sigma_{11}^{-1}\Sigma_{12})^{-1}(\cdot)}{2}\right)
\end{aligned}
$$

which establishes the result. ■

- the linear function $\boldsymbol{\mu}_2 + \Sigma_{12}'\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ is called the *regression* of $\mathbf{X}_2$ on $\mathbf{X}_1$

note *Monte Carlo estimates of probability contents*

- suppose we want to compute $P_{\mathbf{X}}(A)$

- sometimes this can be computed exactly but typically we need to resort to Monte Carlo simulation and estimate $P_{\mathbf{X}}(A)$

- suppose then we have an algorithm that allows us to generate $\mathbf{X} \sim P_{\mathbf{X}}$ and recall $I_A(\mathbf{X}) \sim \text{Bernoulli}(P_{\mathbf{X}}(A))$

- then generate $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim P_{\mathbf{X}}$ and estimate $P_{\mathbf{X}}(A)$ by

$$\hat{P}_{\mathbf{X}}(A) = \frac{1}{n}\sum_{i=1}^{n} I_A(\mathbf{X}_i) = \text{proportion of sampled values falling in } A$$

with standard error

$$\sqrt{\hat{P}_{\mathbf{X}}(A)(1-\hat{P}_{\mathbf{X}}(A))/n}$$

and the interval

$$\left[\hat{P}_{\mathbf{X}}(A) - 3\sqrt{\hat{P}_{\mathbf{X}}(A)(1-\hat{P}_{\mathbf{X}}(A))/n}, \hat{P}_{\mathbf{X}}(A) + 3\sqrt{\hat{P}_{\mathbf{X}}(A)(1-\hat{P}_{\mathbf{X}}(A))/n}\right]$$

contains the value $P_{\mathbf{X}}(A)$ with virtual certainty provided n is large enough

Exercise II.8.7 Suppose

$$\Sigma = \begin{pmatrix} 21 & 26 & 24 \\ 26 & 34 & 30 \\ 24 & 30 & 36 \end{pmatrix}.$$

(a) Using the R software compute $\Sigma^{1/2}$ (command eigen). Verify $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$ numerically (up to small rounding errors).
(b) Using the R software compute the Cholesky factor R. (command chol). Verify $\Sigma = R'R$ numerically (up to small rounding errors).

Exercise II.8.8 Suppose $\boldsymbol{\mu} = (0, 1, 2)'$ and Σ is as in Exercise II.8.7.
(a) Using the R software and the representation $\mathbf{X} = \boldsymbol{\mu} + \Sigma^{1/2}\mathbf{Z}$, where $\mathbf{Z} \sim N_3(\mathbf{0}, I)$, generate a sample of $n = 10^3$ from the $N_3(\boldsymbol{\mu}, \Sigma)$ distribution and based on this sample estimate $P(||\mathbf{X}|| \leq 10)$ and provide the interval containing the exact value with virtual certainty.
(b) Using the R software and the representation $\mathbf{X} = \boldsymbol{\mu} + R'\mathbf{Z}$, where $\mathbf{Z} \sim N_3(\mathbf{0}, I)$, generate a sample of $n = 10^3$ from the $N_3(\boldsymbol{\mu}, \Sigma)$ distribution and based on this sample estimate $P(||\mathbf{X}|| \leq 10)$ and provide the interval containing the exact value with virtual certainty.
(c) Compare the two estimates.

Exercise II.8.9 Suppose

$$\mathbf{X} \sim N_2\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 5/2 & 1/2 \\ 1/2 & 5/2 \end{pmatrix}\right).$$

(a) Determine the conditional distribution $X_2 \mid X_1 = 2$.
(b) Using the conditional distribution in (a) compute the conditional probability of $A = \{(x_1, x_2) : x_1^2 + x_2^2 \leq 5\}$.
(c) Estimate the unconditional probability of A.