Chapter 5 solutions

1. The joint density function of n variables $x_i, i = 1, \ldots, n$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = Ke^{-\mathbf{x}^T\mathbf{\Lambda}}u(\mathbf{x}) = Ke^{-\sum_{i=1}^{n} x_i\lambda_i}u(x_1)\ldots u(x_n).$$

Clearly, if K is a non-negative constant, the pdf also would be non-negative. In order that $f_{\mathbf{X}}(\mathbf{x})$ be a pdf, it should also integrate to 1. Therefore,

$$\int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x} = 1 = K\int_{0}^{\infty} e^{-x_1\lambda_1}\, dx_1 \ldots \int_{0}^{\infty} e^{-x_n\lambda_n}\, dx_n = \frac{K}{\prod_{i=1}^{n}\lambda_1}.$$

Hence only for $K = \prod_{i=1}^{n}\lambda_i$ is $f_{\mathbf{X}}(\mathbf{x})$ a valid pdf.

2. Since $B_i, i = 1, \ldots, n$ are exhaustive, we can write $\Omega = \cup_{i=1}^{n} B_i$, and since they are also disjoint, $\{B_i\}_{i=1}^{n}$ form a partition. So we can write the event $\{\mathbf{X} \leq \mathbf{x}\}$ as

$$\{\mathbf{X} \leq \mathbf{x}\} = \{\mathbf{X} \leq \mathbf{x}\} \cap \Omega = \{\mathbf{X} \leq \mathbf{x}\} \cap (\cup_{i=1}^{n} B_i) = \cup_{i=1}^{n} \{\mathbf{X} \leq \mathbf{x}, B_i\}.$$

Since B_i are disjoint, $\{\mathbf{X} \leq \mathbf{x}, B_i\}$ are also disjoint, and therefore

$$\begin{aligned}
F_{\mathbf{X}}(\mathbf{x}) &= P[\{\mathbf{X} \leq \mathbf{x}\}] \\
&= P\left[\cup_{i=1}^{n} \{\mathbf{X} \leq \mathbf{x}, B_i\}\right] \\
&= \sum_{i=1}^{n} P\left[\{\mathbf{X} \leq \mathbf{x}, B_i\}\right] \\
&= \sum_{i=1}^{n} P\left[\mathbf{X} \leq \mathbf{x} | B_i\right] P[B_i] \text{(By definition of conditional probability)} \\
&= \sum_{i=1}^{n} F_{\mathbf{X}}(\mathbf{x}|B_i)P[B_i]. \text{(Definition of conditional CDF)}
\end{aligned}$$

3. We note that this multi-dimensional Gaussian is factorable so that

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{x_1^2}{2\sigma_1^2}} \ldots \frac{1}{\sqrt{2\pi\sigma_n^2}}e^{-\frac{x_n^2}{2\sigma_n^2}}.$$

With $g(x, \sigma) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x^2}{2\sigma^2}}$ (the Gaussian pdf with mean 0 and variance σ^2), we write

$$f_{\mathbf{X}}(\mathbf{x}) = g(x_1, \sigma_1) \ldots g(x_n, \sigma_n),$$

where $\int_{-\infty}^{\infty} g(x_i, \sigma_1)\, dx_i = 1$. Any marginal pdf $g(x_j, \sigma_j)$ can be obtained by

$$\begin{aligned}
f_{X_j}(x_j) &= \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x})\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_n \\
&= g(x_j, \sigma_j)\int_{-\infty}^{\infty} g(x_1, \sigma_1)\, dx_1 \ldots \int_{-\infty}^{\infty} g(x_{j-1}, \sigma_{j-1})\, dx_{j-1} \\
&\quad \times \int_{-\infty}^{\infty} g(x_{j+1}, \sigma_{j+1})\, dx_{j+1} \ldots \int_{-\infty}^{\infty} g(x_n, \sigma_n)\, dx_n \\
&= g(x_j, \sigma_j).
\end{aligned}$$

4. From Equation 5.3-1 we have

$$
\begin{aligned}
f_{Y_1Y_2Y_3}(y_1,y_2,y_3) &= 3!f_{SN}(y_1)f_{SN}(y_2)f_{SN}(y_3) \\
&= \begin{cases} \frac{3!}{(2\pi)^{3/2}}e^{-\frac{1}{2}(y_1^2+y_2^2+y_3^2)}, & y_1<y_2<y_3, \\ 0, & \text{else.} \end{cases}
\end{aligned}
$$

5. From Problem 5.4 we have that $f_{Y_1Y_2Y_3}(y_1,y_2,y_3) = 3!f(y_1)f(y_2)f(y_3)$ for $y_1 < y_2 < y_3$ and 0 else.To get $f_{Y_1}(y_1)$ we integrate out with respect to y_2 and y_3. Thus

$$
\begin{aligned}
f_{Y_1}(y_1) &= \int\int_{y_1<y_2<y_3} f_{Y_1Y_2Y_3}(y_1,y_2,y_3)dy_2dy_3 \\
&= 3!f(y_1)\int\int_{y_1<y_2<y_3} f(y_2)f(y_3)dy_2dy_3 \\
&= 3!f(y_1)\int_{y_3}^{\infty} f(y_3)\left(\int_{y_1}^{y_3} f(y_2)dy_2\right)dy_3 \\
&= 3!f(y_1)\int_{y_3}^{\infty} f(y_3)\left(F(y_3)-F(y_1)\right)dy_3 \\
&= 3!f(y_1)\int_{y_3}^{\infty} (F(y_3)-F(y_1))dF(y_3) \\
&= 3!f(y_1)\frac{(1-F(y_1))^2}{2} \\
&= 3f(y_1)\left(1-F(y_1)\right)^2.
\end{aligned}
$$

To get $f_{Y_2}(y_2)$ we integrate out with respect to y_1 and y_3. Thus

$$
\begin{aligned}
f_{Y_2}(y_2) &= \int\int_{y_1<y_2<y_3} f_{Y_1Y_2Y_3}(y_1,y_2,y_3)dy_1dy_3 \\
&= 3!f(y_2)\int\int_{y_1<y_2<y_3} f(y_1)f(y_3)dy_1dy_3 \\
&= 3!f(y_2)\int_{y_2}^{\infty} f(y_3)\left(\int_{-\infty}^{y_2} f(y_1)dy_1\right)dy_3 \\
&= 3!f(y_2)\int_{y_2}^{\infty} f(y_3)F(y_2)dy_3 \\
&= 3!f(y_2)F(y_2)\int_{y_2}^{\infty} f(y_3)dy_3 \\
&= 3!f(y_2)F(y_2)\left(1-F(y_2)\right).
\end{aligned}
$$

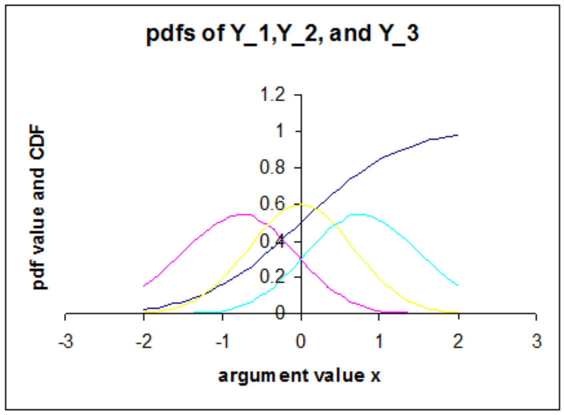


Figure 1:

Finally to get $f_{Y_3}(y_3)$, we integrate out with respect to y_1 and y_2, thus

$$\begin{aligned} f_{Y_3}(y_3) &= \int\int_{y_1<y_2<y_3} f_{Y_1Y_2Y_3}(y_1,y_2,y_3)dy_1dy_2 \\ &= 3!f(y_3)\int\int_{y_1<y_2<y_3} f(y_1)f(y_2)dy_1dy_2 \\ &= 3!f(y_3)\int_{-\infty}^{y_3} f(y_2)\left(\int_{-\infty}^{y_2} f(y_1)dy_1\right)dy_2 \\ &= 3!f(y_3)\int_{-\infty}^{y_3} f(y_2)F(y_2)dy_3 \\ &= 3!f(y_3)\int_{-\infty}^{y_3} F(y_2)dF(y_2) \\ &= 3f(y_3)F^2(y_3). \end{aligned}$$

In the figure above the pink curve is $f_{Y_1}(x)$, the yellow curve is $f_{Y_2}(x)$, and the blue curve is $f_{Y_3}(x)$. The CDF is shown in dark blue.

6. We are given $f_{Z_1Z_2\cdots Z_n}(z_1,z_2,\cdots,z_n) = n!z_1 < z_2 < \cdots < z_n$ and 0 else. To get $f_{Z_1Z_n}(z_1,z_n)$ we integrate out with respect to $z_2, z_3, ..., z_{n-1}$. Thus

$$f_{Z_1Z_n}(z_1,z_n) = n!\int_{z_{n-1}}^{z_n}\left(\cdots\left(\int_{z_1}^{z_3} dz_2\right)\cdots\right)dz_{n-1}.$$

Let's take the first integration and leave out the $n!$:$\int_{z_1}^{z_3} dz_2 = z_3 - z_1$. The second integration yields $\int_{z_1}^{z_4}(z_3-z_1)dz_3 = \frac{(z_4-z_1)^2}{2}$. The third integration yields $\int_{z_1}^{z_5}(z_4-z_1)^2/2dz_3 = \frac{(z_5-z_1)^3}{3\cdot 2}$. Thus after $n-2$ integrations we end up with $\int_{z_1}^{z_n}\frac{(z_{n-1}-z_1)^{n-3}}{n-3}dz_{n-1} = \frac{(z_n-z_1)^{n-2}}{(n-2)\cdots 3\cdot 2}$. Putting back the $n!$ then yields

$$f_{Z_1Z_n}(z_1,z_n) = \begin{cases} n(n-1)(z_n-z_1)^{n-2}, & 0<z_1<z_n<1, n\geq 2 \\ 0, & \text{else.} \end{cases}$$

7. Let $V_{1n} \triangleq Z_n - Z_1$ and define the auxiliary random variable $W \triangleq Z_1$. Consider the functional equations $v = z_n - z_1, w = z_1$. The Jacobian of this transformation is $\begin{vmatrix} \partial v/\partial z_1 & \partial v/\partial z_n \\ \partial w/\partial z_1 & \partial w/\partial z_n \end{vmatrix} = \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix}$, so that $|J| = 1$. Substituting into $f_{Z_1 Z_n}(z_1, z_n) = n(n-1)(z_n - z_1)^{n-2}$, we get

$$f_{V_{1n}W}(v,w) = \begin{cases} n(n-1)v^{n-2}, & 0 < w < 1-v,\ 0 < v < 1, \\ 0, & \text{else.} \end{cases}$$

8. We already know that $f_{V_{1n}W}(v,w) = n(n-1)v^{n-2}, 0 < w < 1-v, 0 < v < 1$. To get $f_{V_{1n}}(v)$, we integrate out with respect to w. This yields

$$\begin{aligned} f_{V_{1n}}(v) &= n(n-1)v^{n-2} \int_{w=0}^{1-v} dw \\ &= \begin{cases} n(n-1)(1-v)v^{n-2}, & 0 < v < 1, n \geq 2, \\ 0, & \text{else.} \end{cases} \end{aligned}$$

9. We need to show that the area under

$$f_{Z_1 Z_2 Z_3}(z_1, z_2, z_3) = \begin{cases} 3!, & 0 < z_1 < z_2 < z_3 < 1, \\ 0, & \text{else,} \end{cases}$$

is 1. A simple repeated integration yields

$$\begin{aligned} 6\int_0^1 \left(\int_0^{z_3} \left(\int_0^{z_2} dz_1 \right) dz_2 \right) dz_3 &= 6\int_0^1 \left(\int_0^{z_3} z_2 dz_2 \right) dz_3 \\ &= 6(\frac{1}{2}) \int_0^1 z_3^2 dz_3 \\ &= 6(\frac{1}{2})(\frac{1}{3}) \left(z_3^3 \big|_0^1 \right) \\ &= 1. \end{aligned}$$

10. The beta pdf is given

$$f(x;\alpha,\beta) = \begin{cases} \frac{(\alpha+\beta+1)}{\alpha!\beta!} x^\alpha (1-x)^\beta, & 0 < x < 1, \\ 0, & \text{else.} \end{cases}$$

So for $\beta = 0$ and $n = \alpha + 2 = 2$, which implies that $\alpha = 0$, we get $f(x;0,0) = 1, 0 < x < 1$. Hence the CDF is

$$F(x;0,0) = \begin{cases} 0, x \leq 0 \\ x, 0 < x \leq 1 \\ 1, x > 1 \end{cases}.$$

11. First we derive Equation 5.3-11. We begin with $f_{Z_1 Z_2 Z_3}(z_1 z_2 z_3) = 3!$ for $0 < z_1 < z_2 < z_3 < 1$ and 0, else. As shown in Section 5.3, $f_{Z_2 Z_3}(z_2 z_3) = 3! z_2$ for $0 < z_2 < z_3 < 1$. Now let $V_{23} \triangleq Z_3 - Z_2$ and $W \triangleq Z_2$, with the functional equations $v = z_2 - z_3$ and $w = z_2$. We find $\partial v/\partial z_2 = -1; \partial v/\partial z_3 = 1; \partial w/\partial z_2 = 1; \partial w/\partial z_3 = 0$. Hence $|J| = 1$ and

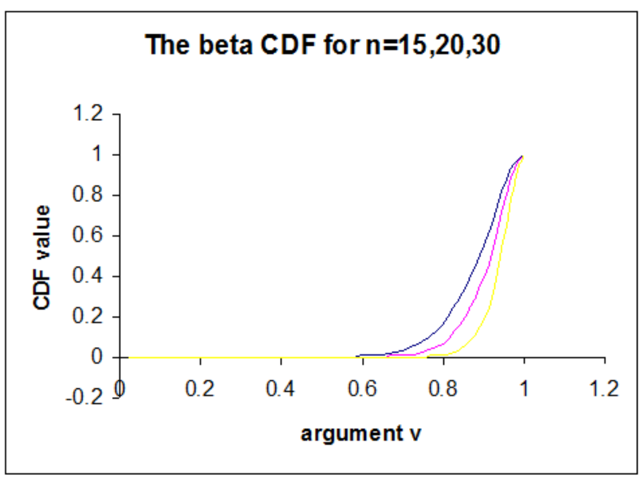
$$f_{V_{23}W}(v,w) = \begin{cases} 3!w, & 0 < w < 1-v, 0 < v < 1 \\ 0, & \text{else.} \end{cases}.$$

Finally

$$\begin{aligned} f_{V_{23}}(v) &= \int_0^{1-v} 3!wdw \\ &= \begin{cases} 3!(1-v)^2/2, & 0 < v < 1, \\ 0, & \text{else.} \end{cases} \end{aligned}$$

Next we derive Equation 5.3-12. We compute $f_{Z_1Z_2}(z_1, z_2) = 3! \int_{z_2}^1 dz_3 = 3!(1-z_2)$ for $0 < z_1 < z_2 < 1$. Let $V_{12} \triangleq Z_2 - Z_1$ and $W \triangleq Z_1$, with the functional equations $v = z_2 - z_1$ and $w = z_1$. Once again we find that $|J| = 1$ and $f_{V_{12}W}(v, w) = 3!(1 - w - v)$ for $0 < w < 1 - v, 0 < v < 1$. Integrating out with respect to w yields $3! \int_0^{1-v} wdw = 3!(1-v)^2/2$, the same as before! To derive Equation 5.3-13 takes more work but follows the same procedure as in deriving Equations 5.3-11 and 5.3-12. We let $V_{lm} \triangleq Z_m - Z_l$ and $W \triangleq Z_l$, with functional equations $v = z_m - z_l$ and $w = z_l$. The magnitude of the Jacobian of this transformation is 1. To get $f_{V_{lm}W}(v, w)$ we have to integrate out with respect to $z_1, z_2, ..., z_{l-1}$; then integrate out with respect to $z_{l+1}, ..., z_{m-1}$; and finally integrate out with respect to $z_{m+1}, ..., z_n$. When we are done with these integrations we integrate out with respect to w to obtain Equation 5.3-13.

12. The beta CDF for $\beta = 1$ *and* $n =$ is shown in the figure below. Recall that $V_{1n} \triangleq \int_{Y_1}^{Y_n} f_X(x)dx$ and therefore V_{1n} is the probability area between the smallest and largest of the ordered values $Y_1, Y_2, \cdots, Y_n$ obtained from the unordered i.i.d. observations on X.The beta CDF is $F_{V_{1n}}(v) = P[\int_{Y_1}^{Y_n} f_X(x)dx \leq v]$. We note that for large $nF_{V_{1n}}(v)$is exceedingly small except for argument values approaching 1. Why is that? For large n it is exceedingly unlikely that the smallest and largest value of the observations on Xwill be near each other. If they were the probability area bounded by them would be very small. It is much more likely that for large n Y_1 will be very small and $Y_n$will be very large. This would cause the probability area between them to be near 1. In effect this is what the curves in the figure below show.



13. There are several ways to do this problem. Here is one: Write

$$E[Z_1] = n! \int_0^1 \int_0^{z_n} \cdots \int_0^{z_2} z_1 dz_1 \cdots dz_{n-1} dz_n$$

Now, let's do the integrations one-by-one, leaving out temporarily the factor $n!$

$$\begin{aligned}
\int_0^{z_2} z_1 dz_1 &= \frac{z_2^2}{2} \\
\int_0^{z_3} (z_2^2/2) dz_1 &= \frac{z_3^3}{3\cdot 2} \\
&\vdots \\
\int_0^{z_n} \frac{z_{n-1}^{n-1}}{(n-1)!} dz_{n-1} &= \frac{z_n^n}{n!}, \\
\text{and finally, } \int_0^1 \frac{z_n^n}{n!} dz_n &= \frac{1}{(n+1)!}
\end{aligned}$$

Hence

$$\begin{aligned}
E[Z_1] &= n! \int_0^1 \int_0^{z_n} \cdots \int_0^{z_2} z_1 dz_1 \cdots dz_{n-1} dz_n \\
&= \frac{n!}{(n+1)!} = \frac{1}{n+1}.
\end{aligned}$$

In a similar way we could compute

$$\begin{aligned}
E[Z_2] &= \frac{2}{n+1} \\
E[Z_3] &= \frac{3}{n+1} \\
&\vdots \\
\text{and } E[Z_n] &= \frac{n}{n+1}.
\end{aligned}$$

Here is a better way: Write

$$\begin{aligned}
Z_i &= (Z_i - Z_1) + Z_1 \\
E[Z_i] &= E[Z_i - Z_1] + E[Z_1] \\
&= E[V_{1i}] + \frac{1}{n+1} \\
&= \frac{n!}{(i-2)!(n-i+1)!} \int_0^1 v \cdot \left(v^{i-2}(1-v)^{n-i+1} \right) dv + \frac{1}{n+1},
\end{aligned}$$

by Equation5.3-13. The integral $\int_0^1 v^{i-1}(1-v)^{n-i+1} dv$ is evaluated in any table of definite available on the internet, for example, formula 7 at sosmath.com/tables/integral/integ41/integ41/html as

$$\int_0^1 \left(v^{i-1}(1-v)^{n-i+1} \right) dv = \frac{(i-1)!(n-i-1)!}{(n+1)!}$$

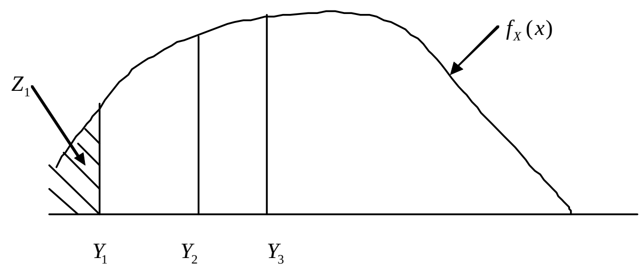
Hence

$$\begin{aligned}
E[Z_i] &= \frac{i-1}{n+1} + \frac{1}{n+1} \\
&= \frac{i}{n+1}.
\end{aligned}$$

14. Consider the n ordered RVs $Y_1 < Y_2 < \cdots < Y_n$ ordered from $X_1, X_2, \cdots, X_n$. With $Z_i = \int_{-\infty}^{Y_i} f_X(x)dx$, we found in the previous problem that $E[Z_i] = \frac{i}{n+1}$. Hence the average probability area between adjacent parts is

$$\begin{aligned} E[Z_{i+1}] - E[Z_i] &= \frac{i+1}{n+1} - \frac{i}{n+1} \\ &= \frac{1}{n+1}. \end{aligned}$$

Thus the number of equal parts is $1/\left(\frac{1}{n+1}\right) = n+1$.



---

15. (a) Let $\mathbf{x} = (x_1, x_2, ..., x_n)^T$, then

$$\begin{aligned} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} &= \begin{bmatrix} x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ x_2x_1 & x_2x_2 & \cdots & x_2x_n \\ \vdots & \vdots & \vdots & \vdots \\ x_nx_1 & x_nx_2 & & x_nx_n \end{bmatrix} \\ &= \begin{bmatrix} x_1\mathbf{x}^T \\ x_2\mathbf{x}^T \\ \vdots \\ x_n\mathbf{x}^T \end{bmatrix}. \end{aligned}$$

Hence any row is obtained from any other row by a scalar multiplication. Therefore, there is at most one linearly independent row. Thus the rank is at most 1.

(b) The expectation operator tends to destroy the linear dependence, i.e.

$$E[\mathbf{X}\mathbf{X}^T] = \begin{bmatrix} E[X_1X_1] & \cdots & E[X_1X_n] \\ \vdots & \cdots & \vdots \\ E[X_nX_1] & \cdots & E[X_nX_n] \end{bmatrix},$$

and we cannot usually write any row as a linear combination of any other one. Indeed if

$$E[X_iX_j] \simeq \frac{1}{N}\sum_{k=1}^{N} x_i^{(k)} x_j^{(k)},$$

when N is large, the matrix can be full rank, provided that $N > n$.

16. We need to show that the CDF of $Y_n \triangleq \max(X_1, X_2, ..., X_n)$ is $F_{Y_n}(y) = F_X^n(y)$. Here the $X_1, X_2, ..., X_n$ are n i.i.d. observations on X with CDF $F_X(x)$. Clearly $F_{Y_n}(y) = P[\max(X_1, X_2, ..., X_n) \leq y]$. But if the max is less than y, all the other RVs must be less than y, hence

$$\begin{aligned} F_{Y_n}(y) &= P[\max(X_1, X_2, ..., X_n) \leq y] \\ &= P[X_1 \leq y]P[X_2 \leq y] \cdots P[X_n \leq y] \\ &= F_X^n(y). \end{aligned}$$

Note that in obtaining the last line we used that the Xs were i.i.d. with CDF $F_X(x)$.

17. Recalling that $Y_1 \triangleq \min(X_1, X_2, ..., X_n)$, we note that $P[Y_1 \leq y] = 1 - P[Y_1 > y]$. Now if the *smallest* i.e. Y_1, of the Xs is *greater* than y, then all the other Xs must be greater than y also. Hence

$$\begin{aligned} P[Y_1 &> y] = P[X_1 > y, X_2 > y ... X_n > y] \\ &= P[X_1 > y]P[X_2 > y] \cdots P[X_n > y] \\ &= (1 - F_{X_1}(y))(1 - F_{X_2}(y)) \cdots (1 - F_{X_n}(y)) \\ &= (1 - F_X(y))^n. \end{aligned}$$

In obtaining the last line, we used the fact that the $\{X_i, i = 1, ..., n\}$ are i.i.d. with CDF $F_X(x)$. Finally, because $P[Y_1 \leq y] = 1 - P[Y_1 > y]$, there results that

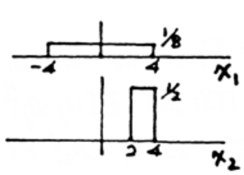
$$\begin{aligned} P[Y_1 \leq y] &= 1 - (1 - F_X(y))^n \\ &\triangleq F_{Y_1}(y). \end{aligned}$$

18. Given the ordered samples $Y_1, Y_2, ..., Y_r, ..., Y_n$ from the n i.i.d. observations $X_1, X_2, ..., X_n$ on X, we want to compute $P[Y_r \leq y]$. However, the event $\{Y_r \leq y\}$ implies that *at least* r of the unordered $\{X_i\}$ satisfy $\{X_i \leq y\}$. Now $P[X_i \leq y]$ is merely $F_X(y)$, and if we count it as a "success" when$\{X_i \leq y\}$, then the probability of at least r successes is the binomial probability

$$P[Y_r \leq y] = \sum\nolimits_{i=r}^{n} \binom{n}{i} F_X^i(y) (1 - F_X(y))^{n-1}.$$

19. We can write

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{8}[u(x_1 + 4) - u(x_1 - 4)] \\ f_{X_2}(x_2) &= \frac{1}{8}[u(x_2 - 2) - u(x_2 - 4)]. \end{aligned}$$



Then $f_{X_1}(x_1)f_{X_2}(x_2) = f_{X_1,X_2}(x_1,x_2)$ and the RVs X_1 and X_2 are independent. Hence, they are uncorrelated,

$$\begin{aligned} E[X_1X_2] &= E[X_1]E[X_2] \\ &= 0 \end{aligned}$$

since $E[X_1] = 0$. Thus, X_1 and X_2 are orthogonal.

20. Since $\mathbf{X}_i$ and $\mathbf{X}_j$ are mutually orthogonal, $E[\mathbf{X}_i^T\mathbf{X}_j] = 0$ for any $i \neq j, 1 \leq i,j \leq n$.

$$\begin{aligned} E\left[\left\|\sum_{i=1}^{n}\mathbf{X}_i\right\|^2\right] &= E\left[(\mathbf{X}_1 + \ldots + \mathbf{X}_n)^T(\mathbf{X}_1 + \ldots + \mathbf{X}_n)\right] \\ &= E\left[\mathbf{X}_1^T\mathbf{X}_1 + \ldots + \mathbf{X}_n^T\mathbf{X}_n + \sum_{i=1}^{n}\sum_{j\neq i}^{n}\mathbf{X}_i^T\mathbf{X}_j\right] \\ &= E\left[\mathbf{X}_1^T\mathbf{X}_1 + \ldots + \mathbf{X}_n^T\mathbf{X}_n\right] + E\left[\sum_{i=1}^{n}\sum_{j\neq i}^{n}\mathbf{X}_i^T\mathbf{X}_j\right] \\ &= \sum_{i=1}^{n}E[\mathbf{X}_i^T\mathbf{X}_i] + \sum_{i=1}^{n}\sum_{j\neq i}^{n}E[\mathbf{X}_i^T\mathbf{X}_j] \\ &= \sum_{i=1}^{n}E[\mathbf{X}_i^T\mathbf{X}_i] + 0 \\ &= \sum_{i=1}^{n}E[\mathbf{X}_i^T\mathbf{X}_i] \\ &= \sum_{i=1}^{n}E[\|\mathbf{X}_i\|]. \end{aligned}$$

21. Since the random vectors are mutually uncorrelated, $E[(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_j - \boldsymbol{\mu}_j)^T] = \mathbf{0}$ for all $i \neq j, 1 \leq i,j \leq n$.

$$\begin{aligned} E\left[\sum_{i=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu}_i)\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu}_j)^T\right] &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_j - \boldsymbol{\mu}_j)^T\right] \\ &= E\left[\sum_{i=1}^{n}(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_i - \boldsymbol{\mu}_i)^T + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_j - \boldsymbol{\mu}_j)\right] \\ &= \sum_{i=1}^{n}\mathbf{K}_{ii} + E\left[\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(\mathbf{X}_i - \boldsymbol{\mu}_i)(\mathbf{X}_j - \boldsymbol{\mu}_j)\right] \\ &= \sum_{i=1}^{n}\mathbf{K}_{ii} + \mathbf{0}. \end{aligned}$$

22.

$$\begin{aligned}\sum_{i=1}^{n}(\mathbf{X}_i-\boldsymbol{\mu}_i)\sum_{j=1}^{n}(\mathbf{X}_j-\boldsymbol{\mu}_j)^T &= \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{X}_i-\boldsymbol{\mu}_i)(\mathbf{X}_j-\boldsymbol{\mu}_j)^T\\ &= \sum_{i=1}^{n}(\mathbf{X}_i-\boldsymbol{\mu}_i)(\mathbf{X}_i-\boldsymbol{\mu}_i)^T+\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(\mathbf{X}_i-\boldsymbol{\mu}_i)(\mathbf{X}_j-\boldsymbol{\mu}_j).\end{aligned}$$

So, upon taking expectations, we have

$$\begin{aligned}E\left[\sum_{i=1}^{n}(\mathbf{X}_i-\boldsymbol{\mu}_i)\sum_{j=1}^{n}(\mathbf{X}_j-\boldsymbol{\mu}_j)^T\right] &= \sum_{i=1}^{n}\mathbf{K}_i+\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}E[(\mathbf{X}_i-\boldsymbol{\mu}_i)(\mathbf{X}_j-\boldsymbol{\mu}_j)]\\ &= \sum_{i=1}^{n}\mathbf{K}_i+\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}E[(\mathbf{X}_i-\boldsymbol{\mu}_i)]E[(\mathbf{X}_j-\boldsymbol{\mu}_j)]\\ &= \sum_{i=1}^{n}\mathbf{K}_i+\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(E[\mathbf{X}_i]-\boldsymbol{\mu}_i)](E[\mathbf{X}_j]-\boldsymbol{\mu}_j)\\ &= \sum_{i=1}^{n}\mathbf{K}_i+0.\end{aligned}$$

23. (a) From the Schwarz inequality,

$$\begin{aligned}\sigma_i^2\sigma_j^2 &= E[(X_i-\mu_i)^2]E[(X_j-\mu_j)^2]\\ &\geq \left|E[(X_i-\mu_i)(X_j-\mu_j)]\right|^2\\ &= |K_{ij}|^2.\end{aligned}$$

But in the given matrix, $\sigma_1^2\sigma_2^2 = 2\times 3 = 6$ and $|K_{12}|^2 = 16 > \sigma_1^2\sigma_2^2$. Thus violating the Schwarz inequality.

(b) Always $\sigma_{33}^2 = E[(X_3-\mu_3)^2] \geq 0$. In the given matrix, this value is -2.

(c) The covariance value $K_{12} = E[(X_1-\mu_1)(X_2-\mu_2)]$ must be real for a real valued random vector. In the given matrix, there are numbers with non-zero imaginary parts.

(d) Always for covariance matrices of real valued random vectors, we have the symmetry conditions

$$\begin{aligned}K_{ij} &= E[(X_i-\mu_i)(X_j-\mu_j)]\\ &= E[(X_j-\mu_j)(X_i-\mu_i)]\\ &= K_{ji}.\end{aligned}$$

But, in the given matrix, $K_{23} = 3, K_{32} = 12 \neq K_{23}$, thus violating the required symmetry.

24. (a)

$$\begin{aligned}\det(\mathbf{K}_{\mathbf{XX}}-\lambda\mathbf{I}) &= \det\begin{bmatrix}3-\lambda & \sqrt{2}\\ \sqrt{2} & 4-\lambda\end{bmatrix} = 0\\ \Rightarrow &\quad \lambda^2-7\lambda+10=0,\ \text{or}\quad \lambda_1=5,\quad \text{and}\quad \lambda_2=2.\end{aligned}$$

i) $\lambda_1 = 5:$ $\quad \mathbf{K_{XX}}\boldsymbol{\phi}_1 = 5\boldsymbol{\phi}_1$ or what is the same, $(\mathbf{K_{XX}} - 5\mathbf{I}) = \mathbf{0}$ leads to

$$\begin{aligned} -2\phi_{11} + \sqrt{2}\phi_{12} = &\ 0 \\ \sqrt{2}\phi_{11} - \phi_{12} = &\ 0 \\ \Rightarrow \quad \boldsymbol{\phi}_1 \ = &\ (\phi_{11}, \phi_{12})^T \\ = &\ \frac{1}{\sqrt{3}}(1, \sqrt{2})^T \quad . \end{aligned}$$

ii) $\lambda_2 = 2:$ $\quad \mathbf{K_{XX}}\boldsymbol{\phi}_1 = 2\boldsymbol{\phi}_1$ or what is the same, $(\mathbf{K_{XX}} - 2\mathbf{I}) = \mathbf{0}$ leads to

$$\begin{aligned} \phi_{21} + \sqrt{2}\phi_{22} = &\ 0 \\ \sqrt{2}\phi_{21} + 2\phi_{22} = &\ 0 \\ \Rightarrow \quad \boldsymbol{\phi}_2 \ = &\ (\phi_{21}, \phi_{22})^T \\ = &\ \sqrt{\frac{2}{3}}(1, -\frac{1}{\sqrt{2}})^T \quad . \end{aligned}$$

Thus

$$\begin{aligned} \mathbf{\Phi} &= \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{3}} & \sqrt{\frac{2}{3}} \\ \sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{3}} \end{bmatrix}. \end{aligned}$$

Set

$$\begin{aligned} \mathbf{\Lambda}^{-\frac{1}{2}} &\triangleq \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{5}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \end{aligned}$$

and the define

$$\mathbf{C} \triangleq \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Phi}^T.$$

Then with $\mathbf{Y} = \mathbf{CX}$, we have

$$\begin{aligned} E[\mathbf{YY}^T] &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Phi}^T E[\mathbf{XX}^T]\mathbf{\Phi}\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Phi}^T \mathbf{K_{XX}}\mathbf{\Phi}\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Phi}^T \left(\mathbf{\Phi}\mathbf{\Lambda}\right)\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}(\mathbf{\Phi}^T\mathbf{\Phi})\mathbf{\Lambda}\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{I}\mathbf{\Lambda}\mathbf{\Lambda}^{-\frac{1}{2}} \\ &= \mathbf{I}. \end{aligned}$$

Thus $\mathbf{Y} = \mathbf{CX} \;\Rightarrow\; \mathbf{K_{XX}} = \mathbf{I}$. Finally, since $\mathbf{C}^T \neq \mathbf{C}^{-1}$, $\mathbf{C}$ is not unitary.

(b) i)

$$\mathbf{AA}' = \begin{bmatrix} aa' + bb' & ab' + bc' \\ a'b + b'c & bb' + cc' \end{bmatrix} \quad \text{and} \quad \mathbf{A}'\mathbf{A} = \begin{bmatrix} aa' + bb' & a'b + b'c \\ ab' + c'b & bb' + cc' \end{bmatrix}.$$

Now

$$\begin{aligned} ab' + bc' &= a'b + b'c \quad \text{if} \quad a = c,\ a\prime = c\prime \\ a'b + b'c &= ab' + c'b \quad \text{if} \quad a = c,\ a\prime = c. \end{aligned}$$

ii) $(\mathbf{AA}')^T = \mathbf{A}'^T\mathbf{A}^T = \mathbf{A}'\mathbf{A}$ since $\mathbf{A}'^T = \mathbf{A}', \mathbf{A} = \mathbf{A}^T$. Hence if $\mathbf{A}'\mathbf{A} = \mathbf{AA}'$, then $(\mathbf{AA}')^T = \mathbf{AA}'$, and the product is a real symmetric matrix.

25. (a) We know $\mathbf{KA} = (\mathbf{A}^T)^{-1}$, and so

$$\begin{aligned} \mathbf{K}_1\mathbf{A} &= (\mathbf{A}^T)^{-1}\mathbf{\Lambda}^{(1)} \\ &= \mathbf{KA\Lambda}^{(1)}. \end{aligned}$$

Thus

$$\mathbf{K}^{-1}\mathbf{K}_1\mathbf{A} = \mathbf{A\Lambda}^{(1)}.$$

(b) Now

$$\begin{aligned} \mathbf{A}^T\mathbf{KA} &= \mathbf{I} \\ &= a_1\underbrace{\mathbf{A}^T\mathbf{K}_1\mathbf{A}}_{\mathbf{\Lambda}^{(1)}} + a_2\mathbf{A}^T\mathbf{K}_2\mathbf{A} \end{aligned}$$

So

$$\begin{aligned} \mathbf{A}^T\mathbf{K}_2\mathbf{A} &= \frac{1}{a_2}[\mathbf{I} - a_1\mathbf{\Lambda}^{(1)}], \quad \text{a difference of diagonal matrices,} \\ &= \mathbf{\Lambda}^{(2)}, \quad \text{a diagonal matrix.} \end{aligned}$$

(c) All diagonal matrices of the same size share the same eigenvectors:

$$\begin{aligned} \boldsymbol{\phi}_1 &= (1, 0, 0, \ldots, 0)^T, \\ \boldsymbol{\phi}_2 &= (0, 1, 0, \ldots, 0)^T, \\ &\ \ \vdots \qquad\qquad\qquad , \\ \boldsymbol{\phi}_n &= (0, \ldots, 0, 0, 1)^T. \end{aligned}$$

(d)

$$\mathbf{\Lambda}^{(2)} = \frac{1}{a_2}[\mathbf{I} - a_1\mathbf{\Lambda}^{(1)}] \quad \text{or} \quad \lambda_i^{(2)} = \frac{1}{a_2}(1 - a_1\lambda_i^{(1)}),$$

so $\max \lambda_i^{(1)}$ produces $\min \lambda_i^{(2)}$ if both $a_1 > 0$ and $a_2 > 0$.

26. (a) Write $\mathbf{W} = (\mathbf{X}_1 : \mathbf{X}_2 : \cdots : \mathbf{X}_m)$. Then

$$\begin{aligned} \mathbf{WW}^T &= [\mathbf{X}_1 : \mathbf{X}_2 : \cdots : \mathbf{X}_m] \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \vdots \\ \mathbf{X}_m^T \end{bmatrix} \\ &= \left[\mathbf{X}_1\mathbf{X}_1^T + \mathbf{X}_2\mathbf{X}_2^T + \ldots + \mathbf{X}_m\mathbf{X}_m^T\right]. \end{aligned}$$

Hence

$$\frac{1}{m}\mathbf{WW}^T = \frac{1}{m}\sum_{i=1}^{m}\mathbf{X}_i\mathbf{X}_i^T.$$

(b) Consider the case when $m = 2, n = 3$. Then with

$$\mathbf{X}_1 \triangleq (X_{11}, X_{21}, X_{31})^T \quad \text{and} \quad \mathbf{X}_2 \triangleq (X_{12}, X_{22}, X_{32})^T,$$

we get

$$[\mathbf{X}_1 : \mathbf{X}_2] \begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \end{bmatrix} = \begin{bmatrix} X_{11}\mathbf{X}_1^T + X_{12}\mathbf{X}_2^T \\ X_{21}\mathbf{X}_1^T + X_{22}\mathbf{X}_2^T \\ X_{31}\mathbf{X}_1^T + X_{32}\mathbf{X}_2^T \end{bmatrix}.$$

Thus there are three rows, but only two independent row vectors. So, any row can be written as a linear combination of the other two rows. Thus the rank is 2. More generally, we will have

$$\mathbf{S} = \begin{bmatrix} X_{11}\mathbf{X}_1^T + \ldots + X_{1m}\mathbf{X}_m^T \\ \vdots \\ X_{n1}\mathbf{X}_1^T + \ldots + X_{nm}\mathbf{X}_m^T \end{bmatrix}, \text{ with } \mathrm{rank}(\mathbf{S}) \le m.$$

(c) We have $\mathbf{W} = (n \times m)$ and so $\mathbf{W}^T = (m \times n)$. Hence

$$\mathbf{S}' = \frac{1}{m}\mathbf{W}^T\mathbf{W} = (m \times n).$$

Write

$$\begin{aligned} \mathbf{S\Phi} &= \mathbf{\Phi\Lambda} = \frac{1}{m}\mathbf{W}\mathbf{W}^T\mathbf{\Phi} \\ \mathbf{S}'\mathbf{\Phi}' &= \frac{1}{m}\mathbf{W}^T\mathbf{W}\mathbf{\Phi}' = \mathbf{\Phi}'\mathbf{\Lambda}'. \end{aligned}$$

Then

$$\begin{aligned} \mathbf{W}\mathbf{S}'\mathbf{\Phi}' &= \mathbf{W}\mathbf{\Phi}'\mathbf{\Lambda}' \\ &= \frac{1}{m}\mathbf{W}\mathbf{W}^T(\mathbf{W}\mathbf{\Phi}') \\ &= (\mathbf{W}\mathbf{\Phi}')\mathbf{\Lambda}'. \end{aligned}$$

Thus, the first m eigenvalues of $\mathbf{S}'$ are those of $\mathbf{S}$, and the remaining $n - m$ eigenvalues are zero. The first m eigenvalues of $\mathbf{S}'$ are $\mathbf{W}\mathbf{\Phi}' = (n \times m) \times (m \times n) = (n \times n)$.

(d) Only an $m \times m$ matrix is used for calculating m eigenvalues. In practice, very often $n >> m$, so this method saves a lot of work.

27. (a)

$$\begin{aligned} (\mathbf{K} + \Delta\mathbf{K})(\boldsymbol{\phi}_i + \Delta\boldsymbol{\phi}_i) &= (\lambda_i + \Delta\lambda_i)(\boldsymbol{\phi}_i + \Delta\boldsymbol{\phi}_i) \\ &\approx \mathbf{K}\boldsymbol{\phi}_i + \mathbf{K}\Delta\boldsymbol{\phi}_i + \Delta\mathbf{K}\boldsymbol{\phi}_i \\ &\approx \lambda_i\boldsymbol{\phi}_i + \lambda_i\Delta\boldsymbol{\phi}_i + \Delta\lambda_i\boldsymbol{\phi}_i, \end{aligned}$$

by keeping only first-order terms. Thus, to first order, we have:

$$\mathbf{K}\Delta\boldsymbol{\phi}_i + \Delta\mathbf{K}\boldsymbol{\phi}_i = \lambda_i\Delta\boldsymbol{\phi}_i + \Delta\lambda_i\boldsymbol{\phi}_i.$$

Then, multiplying by $\boldsymbol{\phi}_i^T$, we get

$$\boldsymbol{\phi}_i^T\mathbf{K}\Delta\boldsymbol{\phi}_i + \boldsymbol{\phi}_i^T\Delta\mathbf{K}\boldsymbol{\phi}_i = \lambda_i\boldsymbol{\phi}_i^T\Delta\boldsymbol{\phi}_i + \Delta\lambda_i\boldsymbol{\phi}_i^T\boldsymbol{\phi}_i.$$

But $\boldsymbol{\phi}_i^T \mathbf{K} = \lambda_i \boldsymbol{\phi}_i^T$, thus, it must be that

$$\begin{aligned} \boldsymbol{\phi}_i^T \Delta\mathbf{K}\boldsymbol{\phi}_i &= \Delta\lambda_i \boldsymbol{\phi}_i^T \boldsymbol{\phi}_i \\ &= \Delta\lambda_i, \end{aligned}$$

which is the desired result.

(b) First , write

$$\Delta\boldsymbol{\phi}_i = \sum_{j=1}^{n} b_{ij}\boldsymbol{\phi}_j.$$

Then, by orthogonality of the $\boldsymbol{\phi}_i$'s, we will have $b_{ij} = \boldsymbol{\phi}_j^T \Delta\boldsymbol{\phi}_i$. Then

$$\begin{aligned} \boldsymbol{\phi}_j^T(\mathbf{K}\Delta\boldsymbol{\phi}_i + \Delta\mathbf{K}\boldsymbol{\phi}_i) &= \lambda_i \boldsymbol{\phi}_j^T \Delta\boldsymbol{\phi}_i + \Delta\lambda_i \boldsymbol{\phi}_j^T \boldsymbol{\phi}_i \\ &= \lambda_i \boldsymbol{\phi}_j^T \Delta\boldsymbol{\phi}_i \quad \text{for} \quad j \neq i, \\ &= \lambda_i b_{ij}, \end{aligned}$$

or

$$\begin{aligned} \lambda_i b_{ij} &= \boldsymbol{\phi}_j^T \mathbf{K}\Delta\boldsymbol{\phi}_i + \boldsymbol{\phi}_j^T \Delta\mathbf{K}\boldsymbol{\phi}_i \\ &= \lambda_j \boldsymbol{\phi}_j^T \Delta\boldsymbol{\phi}_i + \boldsymbol{\phi}_j^T \Delta\mathbf{K}\boldsymbol{\phi}_i, \quad \text{by K-L equation,} \\ &= \lambda_j b_{ij} + \boldsymbol{\phi}_j^T \Delta\mathbf{K}\boldsymbol{\phi}_i, \quad \text{by result above,} \end{aligned}$$

thus

$$b_{ij} = \frac{\boldsymbol{\phi}_j^T \Delta\mathbf{K}\boldsymbol{\phi}_i}{\lambda_i - \lambda_j} \quad \text{for} \quad i \neq j.$$

28. (a) Let $\boldsymbol{\phi}_i, i = 1, ..., n$ be a complete set of eigenvectors of $n \times n$ matrix $\mathbf{M}$, i.e.

$$\mathbf{M}\boldsymbol{\phi}_i = \lambda_i \boldsymbol{\phi}_i \qquad i = 1, ..., n.$$

Then write $\mathbf{u} = \sum_{j=1}^{n} c_j \boldsymbol{\phi}_j$, which must be possible since the $\boldsymbol{\phi}_i$'s form a basis for R^n. Then

$$c_j = \boldsymbol{\phi}_j^T \mathbf{u} = \mathbf{u}^T \boldsymbol{\phi}_j$$

Now

$$\begin{aligned} \|\mathbf{u}\|^2 &= \|\mathbf{u}\| \\ &= 1 \\ &= \sum_{i,j} c_i c_j \boldsymbol{\phi}_i^T \boldsymbol{\phi}_j \\ &= \sum_{i=1}^{n} c_i^2, \end{aligned}$$

Thus $\sum_{i=1}^{n} c_i^2 = 1$ and we are given that $c_1 = c_2 = \cdots = c_{i-1} = 0$. Then,

$$\begin{aligned} \mathbf{u}^T \mathbf{M}\mathbf{u} &= \left(\sum_{j=i}^{n} c_j \boldsymbol{\phi}_j^T\right)\left(\sum_{k=i}^{n} c_k \lambda_k \boldsymbol{\phi}_k\right) \\ &= \sum_{j=i}^{n}\sum_{k=i}^{n} c_j c_k \lambda_k \boldsymbol{\phi}_j^T \boldsymbol{\phi}_k \\ &= \sum_{j=i}^{n} \lambda_j c_j^2. \end{aligned}$$

Now

$$\begin{aligned} \lambda_i c_i^2 + \lambda_{i+1} c_{i+1}^2 + \cdots \lambda_n c_n^2 &\leq \lambda_i \left(\sum_{j=1}^{n} c_j^2 \right) \\ &= \lambda_i. \end{aligned}$$

But λ_i is reachable as an upper bound if $c_i = 1$ and $c_{i+1} = \cdots = c_n = 0$, i.e. if $\mathbf{u} = \boldsymbol{\phi}_i$.

(b) Define

$$\mathbf{u} \triangleq \frac{\mathbf{x}}{\|\mathbf{x}\|},$$

then $\|\mathbf{u}\| = 1$ and

$$\frac{\mathbf{x}^T \mathbf{M} \mathbf{x}}{\|\mathbf{x}\|^2} = \mathbf{u}^T \mathbf{M} \mathbf{u},$$

which has a maximum value when $\mathbf{u} = \boldsymbol{\phi}_1$. Hence,

$$\boldsymbol{\phi}_1 = \frac{\mathbf{x}}{\|\mathbf{x}\|} \quad \text{or} \quad \mathbf{x} = \|\mathbf{x}\| \, \boldsymbol{\phi}_1 = \mathbf{K} \boldsymbol{\phi}_1,$$

where $\|\mathbf{x}\|$ is any real constant.

29. The mean of $\mathbf{Y}$ is given by

$$E[\mathbf{Y}] = E[\mathbf{A}^T \mathbf{X} + B] = \mathbf{A}^T \boldsymbol{\mu} + B = \begin{pmatrix} 2 & -1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -5 \\ 6 \end{pmatrix} + 5 = 32.$$

Let $E[\mathbf{Y}] = \boldsymbol{\mu}_\mathbf{1}$. Then

$$\begin{aligned} \sigma_Y^2 &= E\left[(\mathbf{Y} - \boldsymbol{\mu}_\mathbf{1})^T (\mathbf{Y} - \boldsymbol{\mu}_\mathbf{1}) \right] \\ &= E\left[\left(\mathbf{A}^T (\mathbf{X} - \boldsymbol{\mu}) \right)^T \left(\mathbf{A}^T (\mathbf{X} - \boldsymbol{\mu}) \right) \right] \\ &= E\left[\left((\mathbf{X} - \boldsymbol{\mu})^T \mathbf{A} \right) \left(\mathbf{A}^T (\mathbf{X} - \boldsymbol{\mu}) \right) \right] \\ &= E\left[\left(\mathbf{A}^T (\mathbf{X} - \boldsymbol{\mu}) \right) \left((\mathbf{X} - \boldsymbol{\mu})^T \mathbf{A} \right) \right] \\ &= \mathbf{A}^T E\left[(\mathbf{X} - \boldsymbol{\mu})^T (\mathbf{X} - \boldsymbol{\mu}) \right] \mathbf{A} \\ &= \mathbf{A}^T \mathbf{K}_\mathbf{X} \mathbf{A} \\ &= \begin{pmatrix} 2 & -1 & 2 \end{pmatrix} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 5 & 0 \\ -1 & 0 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix} = 25. \end{aligned}$$

30. From the given expression, we conclude that $\sigma_1^2 = \sigma_2^2 = 1$ and that $-2\rho = 3/2$ implying $\rho = -3/4$. Next, write

$$x_1^2 + \frac{3}{2} x_1 x_2 + x_2^2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & \frac{3}{4} \\ \frac{3}{4} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Thus we need to diagonalize (find the eigenvalues and eigenvectors) of

$$\begin{bmatrix} 1 & \frac{3}{4} \\ \frac{3}{4} & 1 \end{bmatrix} = \mathbf{K}^{-1}.$$

We obtain

$$\lambda_1 = \frac{7}{4}, \lambda_2 = \frac{1}{4} \text{ and } \boldsymbol{\phi}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \boldsymbol{\phi}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \text{ (unnormalized).}$$

Hence

$$\mathbf{y} = \mathbf{\Phi}^T \mathbf{x} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

and the transformation is

$$\begin{aligned} y_1 &= x_1 + x_2 \\ y_2 &= x_1 - x_2, \end{aligned}$$

with inverse

$$\begin{aligned} x_1 &= \frac{1}{2}(y_1 + y_2) \\ x_2 &= \frac{1}{2}(y_1 - y_2). \end{aligned}$$

Thus

$$\begin{aligned} f_{\mathbf{Y}}(y_1, y_2) &= f_{\mathbf{X}}\left(\frac{1}{2}(y_1+y_2), \frac{1}{2}(y_1-y_2)\right) \cdot \frac{1}{2} \\ &= \frac{1}{\pi\sqrt{7}} \exp\left\{-\frac{8}{7}\left[\left(\frac{y_1+y_2}{2}\right)^2 + \frac{3}{2}\left(\frac{y_1+y_2}{2}\right)\left(\frac{y_1-y_2}{2}\right) + \left(\frac{y_1-y_2}{2}\right)^2\right]\right\} \\ &= \frac{1}{\pi\sqrt{7}} \exp\left\{-\left[y_1^2 + \frac{1}{7}y_2^2\right]\right\} \\ &= \frac{1}{2\pi(1/\sqrt{2})(\sqrt{7/2})} \exp -\frac{1}{2}\left[\left(\frac{y_1}{1/\sqrt{2}}\right)^2 + \left(\frac{y_2}{\sqrt{7/2}}\right)^2\right] \end{aligned}$$

31. The mean of Y is given as

$$E[Y] = E[\sum_{i=1}^{n} p_i X_i] = \sum_{i=1}^{n} p_i E[X_i] = \sum_{i=1}^{n} p_i \mu_i.$$

The variance of Y is given by

$$\begin{aligned} \sigma_Y^2 &= E[(Y - \mu_Y)^2] \\ &= E\left[\sum_{i=1}^{n} p_i (X_i - \mu_i)\right]^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j E[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j K_{ij}. \end{aligned}$$

32. The characteristic function of X_i is

$$\begin{aligned}
\Phi_i(\omega) &= \int_{-\infty}^{+\infty} \frac{\alpha}{\pi(x^2+\alpha^2)} e^{j\omega x} dx, \quad ,\alpha > 0, \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\frac{\alpha}{\alpha + j\omega} + \frac{\alpha}{\alpha - j\omega} \right) e^{j\omega x} dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\alpha}{\alpha + j\omega} e^{j\omega x} dx + \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\alpha}{\alpha - j\omega} e^{j\omega x} dx \\
&= e^{-\alpha\omega} u(\omega) + e^{+\alpha\omega} u(-\omega) \\
&= \exp -\alpha|\omega| \ \ , -\infty < \omega < \infty.
\end{aligned}$$

Now

$$\begin{aligned}
\Phi_Y(\omega) &= [\Phi_i(\omega)]^n \\
&= \exp -n\alpha|\omega|,
\end{aligned}$$

thus

$$f_Y(y) = \frac{n\alpha}{\pi(x^2 + n^2\alpha^2)}.$$

33. Let r and p be the parameters of the binomial distribution, then for each X_i we have

$$\begin{aligned}
\Phi_{X_i}(\omega) &= \sum_{k=0}^{r} \binom{r}{k} p^k q^{r-k} e^{+j\omega k} \\
&= \left(pe^{j\omega} + q\right)^r \quad \text{with} \quad q \triangleq 1 - p.
\end{aligned}$$

Considering the vector $\mathbf{X} = (X_1, \ldots, X_n)^T$, we have

$$\begin{aligned}
\Phi_{\mathbf{X}}(\boldsymbol{\omega}) &= E[e^{+j\boldsymbol{\omega}^T \mathbf{X}}] \\
&= \prod_{i=1}^{n} E[e^{+j\omega_i X_i}] \\
&= \prod_{i=1}^{n} \left(pe^{j\omega_i} + q\right)^r, \qquad \text{by the i.i.d. assumption.}
\end{aligned}$$

Then, with $Y = \sum_{i=1}^{n} X_i = \mathbf{1}^T \mathbf{X}$, where is a column vector of 1's, we have

$$\begin{aligned}
\Phi_Y(\omega) &= E[e^{+j\omega Y}] \\
&= E[e^{+j\omega \mathbf{1}^T \mathbf{X}}] \\
&= \Phi_{\mathbf{X}}(\omega \mathbf{1}) \\
&= \left(pe^{j\omega} + q\right)^{nr},
\end{aligned}$$

which we can recognize as binomial with parameters nr and p. Hence

$$\begin{aligned}
P_Y(k) &= P[Y = k] \\
&= \binom{nr}{k} p^k q^{nr-k} \qquad k = 0, 1, ..., nr.
\end{aligned}$$

The mean of Y is nr, and the variance of Y is $nrpq$.

34. For covariance matrix $\mathbf{K}$, upon setting $\boldsymbol{\omega} \triangleq (\omega_1, \omega_2, \omega_3, \omega_4)^T$, we have

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) = \exp -\frac{1}{2}\boldsymbol{\omega}^T \mathbf{K} \boldsymbol{\omega} \ .$$

Now, by definition

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) \triangleq \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{j(\omega_1 x_1+\omega_2 x_2+\omega_3 x_3+\omega_4 x_4)} f_{\mathbf{X}}(x_1,x_2,x_3,x_4)dx_1dx_2dx_3dx_4,$$

so that

$$E[X_1X_2X_3X_4] = \left.\frac{\partial^4 \Phi_{\mathbf{X}}(\boldsymbol{\omega})}{\partial\omega_1\partial\omega_2\partial\omega_3\partial\omega_4}\right|_{\boldsymbol{\omega}=\mathbf{0}} \ .$$

In calculating this 4-th order partial derivative, it is helpful to write out

$$\boldsymbol{\omega}^T \mathbf{K} \boldsymbol{\omega} = \sum_{i=1}^{4}\sum_{l=1}^{4} K_{il}\omega_i\omega_l,$$

then

$$\begin{aligned}
\frac{\partial \Phi_{\mathbf{X}}}{\partial\omega_1} &= -\Phi_{\mathbf{X}}\sum_{l=1}^{4} K_{1l}\omega_l \quad \text{at} \quad \omega_1 = 0,\\
\frac{\partial^2 \Phi_{\mathbf{X}}}{\partial\omega_1\partial\omega_2} &= -\Phi_{\mathbf{X}}\left[K_{12} - \sum_{l=1}^{4}\sum_{k=1}^{4} K_{1l}K_{2k}\omega_l\omega_k\right] \quad \text{at} \quad \omega_1=\omega_2 = 0,\\
\frac{\partial^3 \Phi_{\mathbf{X}}(\boldsymbol{\omega})}{\partial\omega_1\partial\omega_2\partial\omega_3} &= \Phi_{\mathbf{X}}\left[2K_{13}K_{23} + K_{23}\sum_{l=1}^{4} K_{1l}\omega_l + K_{13}\sum_{l=1}^{4} K_{2l}\omega_l\right]\\
&+ \Phi_{\mathbf{X}}\left[\left(K_{12} - \sum_{l=1}^{4}\sum_{k=1}^{4} K_{1l}K_{2k}\omega_l\omega_k\right)\sum_{l=1}^{4} K_{3l}\omega_l\right] \quad \text{at} \quad \omega_1=\omega_2=\omega_3 = 0.
\end{aligned}$$

Finally

$$\left.\frac{\partial^4 \Phi_{\mathbf{X}}(\boldsymbol{\omega})}{\partial\omega_1\partial\omega_2\partial\omega_3\partial\omega_4}\right|_{\boldsymbol{\omega}=\mathbf{0}} = K_{12}K_{34} + K_{13}K_{24} + K_{23}K_{14}.$$

35. The first order pdf f_{X_1} is computed from

$$\begin{aligned}
f_{X_1}(x) &= \frac{2}{3}\int_0^1\int_0^1 (x_1+x_2+x_3)dx_2dx_3\\
&= \begin{cases} \frac{2}{3}(x+1), & 0 < x \le 1,\\ 0, & \text{else.}\end{cases}
\end{aligned}$$

By symmetry, $f_{X_1} = f_{X_2} = f_{X_3}$. The second-order pdf is calculated as

$$\begin{aligned}
f_{X_1,X_2}(x,y) &= \int_0^1 (x+y+x_3)dx_3\\
&= \begin{cases} \frac{2}{3}(x+y+\frac{1}{2}), & 0 < x, y \le 1,\\ 0, & \text{else.}\end{cases}
\end{aligned}$$

By symmetry, $f_{X_1,X_2} = f_{X_1,X_3} = f_{X_2,X_2}$. Likewise for the first moment, we get

$$\begin{aligned} E[X_1] &= \int_0^1 x\frac{2}{3}(x+1)dx \\ &= \frac{5}{9} = \mu_1 \\ &= \mu_2 = \mu_3. \end{aligned}$$

Then the variance is given as

$$\begin{aligned} \mathrm{Var}[X_i] &= E[X_i^2] - \mu_i^2 \\ &= \frac{2}{3}\int_0^1 x^2(x+1)dx - (\frac{5}{9})^2 \\ &\doteq 0.08 = \sigma^2. \end{aligned}$$

Also, $K_{12} = K_{13} = K_{23} \doteq -0.003$. Then, upon setting $\rho_{12} = K_{12}/\sigma^2 = \rho_{13} = \rho_{23}$, we have

$$\mathbf{K} \doteq 0.08 \begin{bmatrix} 1 & -0.04 & -0.04 \\ -0.04 & 1 & -0.04 \\ -0.04 & -0.04 & 1 \end{bmatrix}.$$

Since $\rho_{12} = \rho_{13} = \rho_{23} \simeq 0$, the random variables X_1, X_2, and X_3 are almost uncorrelated.

36. The first step is to find the eigenvalues and eigenvectors. Hence

$$\begin{aligned} (\mathbf{K} - \lambda\mathbf{I})\,\boldsymbol{\phi} &= \mathbf{0} \\ \Rightarrow \quad \det(\mathbf{K} - \lambda\mathbf{I}) &= 0 \\ &= \begin{vmatrix} 2-\lambda & -1.5 \\ -1.5 & 2-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 1.75. \end{aligned}$$

We get: $\lambda_1 = 3.5$ and $\lambda_2 = 0.5$. For $\lambda_1 = 3.5$, we compute a normalized $\boldsymbol{\phi}_1$ as the solution to $(\mathbf{K} - 3.5\mathbf{I})\,\boldsymbol{\phi}_1 = \mathbf{0}$

$$\begin{bmatrix} -1.5 & -1.5 \\ -1.5 & -1.5 \end{bmatrix} \begin{bmatrix} \phi_{11} \\ \phi_{12} \end{bmatrix} = \begin{matrix} 0 \\ 0 \end{matrix},$$

as $\boldsymbol{\phi}_1 = \frac{1}{\sqrt{2}}(1,-1)^T$. For $\lambda_2 = 0.5$, $(\mathbf{K} - 0.5\mathbf{I})\,\boldsymbol{\phi}_2 = \mathbf{0}$ yields a solution $\boldsymbol{\phi}_2 = \frac{1}{\sqrt{2}}(1,1)^T$. Thus

$$\boldsymbol{\Phi} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Lambda} = \begin{bmatrix} 3.5 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

Hence $\mathbf{Y} = \boldsymbol{\Phi}^T\mathbf{X}$ will produce $E[Y_1Y_2] = 0$. To whiten, let

$$\mathbf{Z} \triangleq \boldsymbol{\Lambda}^{-1/2}\mathbf{Y},$$

yielding

$$\begin{aligned} E[\mathbf{Z}\mathbf{Z}^T] &= \boldsymbol{\Lambda}^{-1/2}\mathbf{K}_{\mathbf{Y}}\boldsymbol{\Lambda}^{-1/2} \\ &= \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\Phi}^T\mathbf{K}_X\boldsymbol{\Phi}\boldsymbol{\Lambda}^{-1/2} \\ &= \mathbf{I}, \quad \text{the identity matrix.} \end{aligned}$$

Hence $\mathbf{Z} = \mathbf{\Lambda}^{-1/2}\mathbf{\Phi}^T\mathbf{X}$ is the whitening transformation. Since it is easy to generate samples from $N(0,1)$, the simplest way to obtain a scatter diagram for correlated random variables is the following: Let

$$\begin{aligned} Y_1 &\triangleq aX_1 + bX_2 \\ Y_2 &\triangleq cX_1 + dX_2 \end{aligned}$$

with both $X_i : N(0,1)$ and independent. We then need to satisfy $\text{Var}(Y_1) = a^2 + b^2 = 2 = \text{Var}(Y_2) = c^2 + d^2$ and covariance $\text{Cov}(Y_1, Y_2) = ac + bd = -1.5$. Since there are 3 equations and 4 unknowns, let $c = 0$, then $d = \pm\sqrt{2}$, $b = \mp 3/\sqrt{8}$, and $a = \pm\sqrt{7/8}$. Now, the signs of a and d are arbitrary. Thus, we obtain

$$\begin{aligned} Y_1 &= \sqrt{\frac{7}{8}}X_1 + \sqrt{\frac{9}{8}}X_2, \\ Y_2 &= -\sqrt{2}X_2, \end{aligned}$$

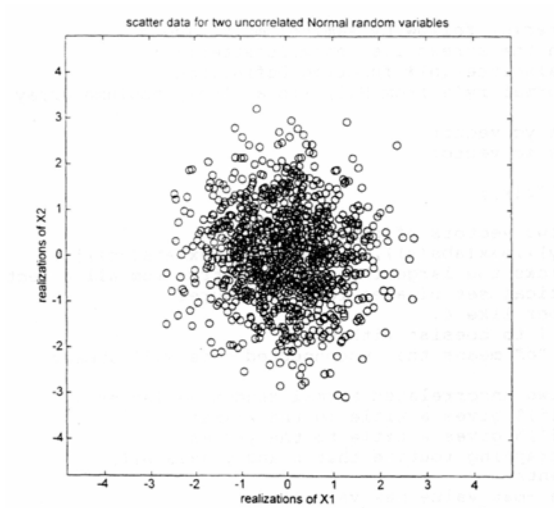
or $Y_1 = 0.9354X_1 + 1.067X_2$ and $Y_2 = -1.414X_2$.

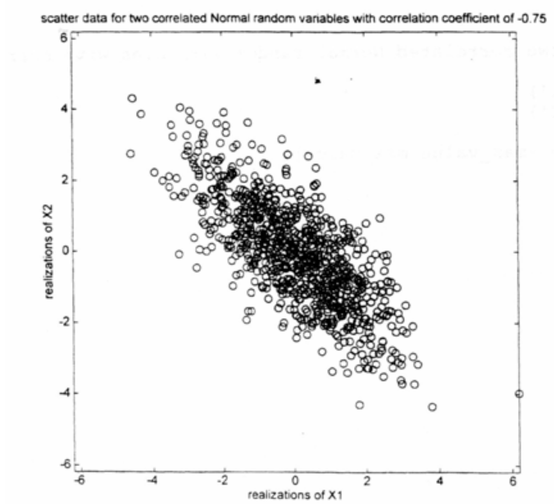
The MATLAB program can be given as:

```
% when calling normalscatter(n) follow by semicolumn to avoid
%seeing all the numbers on the screen i.e.,normalscatter(n);
function [y,z,yc,zc]=normalscatter(n)% function Definition
y=randn(1,n);%generates Normal rv's from N(0,1) in a 1row, ncolumn array
z=randn(1,n);
yc=zeros(1,n);%initializes yc vector
zx=zeros(1,n);%initializes zc vector
for k=1:n
yc(k)=0.9354*y(k)+1.067*z(k);
zc(k)=-1.414*z(k);
end %this loop generates two vectors of correlated RVs
max_value = max([max(abs(y)),max(abs(z)),max(abs(yc)),max(abs(zc))]);
% the above instruction picks the largest magnitude entry from all 4 vectors
% in order to set an identical set of axes for two figures.
% max_value will be a number like 4.
figure(1)% will allow Fig.1 to coesist with Fig.2
plot(y, z, 'o' )% the litlle "o" means that unconnected dote will appear
%in the figures
title(['scatter data for two uncorrelated Normal random variables'])
xlabel('realizations of X1')% gives a title to the x-axis
ylabel('realizations of X2')% gives a title to the y-axes
axis('square')%instructs graphing routine that x and y axis will
%cover the same displacements
axis([-max_value max_value -max_value max_value])
% the horizontal and vertical axes will go from
%-max_value to +max_value on both figures
figure(2)
plot(yc, zc, 'o' )
title(['scatter data for two correlated Normal random variables with correlation
coefficient of -0.75'])
```

xlabel('realizations of Xl')
ylabel('realizations of X2')
axis('square')
axis([-max_value max_value -max_value max_value])

with scatter plots:





37. Let $g_k(\mathbf{x}) \triangleq \sum_{j=1}^{n} a_{kj}x_j, j = 1, ..., n$. Then the only solution to the set of equations

$$\begin{bmatrix} g_1(\mathbf{x}) \\ \vdots \\ g_n(\mathbf{x}) \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \mathbf{0},$$

is

$$\begin{aligned} \mathbf{x}^o &= \mathbf{A}^{-1}\mathbf{y} \\ &= \mathbf{B}\mathbf{y}, \text{ with } \mathbf{B} \triangleq \mathbf{A}^{-1}. \end{aligned}$$

From Eq. (5.2-9 ??),

$$f_{\mathbf{Y}}(y_1,...,y_n) = \frac{f_{\mathbf{X}}(x_1^o,...,x_n^o)}{|J|},$$

where

$$\begin{aligned}
|J| &= \text{mag}\left(\det\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial xn} \end{bmatrix}\right) \\
&= \text{mag}\left(\det\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}\right) \\
&= |\det \mathbf{A}|.
\end{aligned}$$

But, if $\mathbf{B} = \mathbf{A}^{-1}$, then $\det B = 1/\det \mathbf{A}$, and so

$$f_{\mathbf{Y}}(y_1,...,y_n) = |\det \mathbf{A}| f_{\mathbf{X}}(x_1^o,...,x_n^o).$$

38. This problem is a special case of problem 5.37 with an easily computable inverse. We cannot solve the original system using the direct method unless we use auxiliary variables

$$\begin{array}{lcllllllll}
Y_1 & = & X_1 & + & X_2 & + & \cdots & + & X_n & \\
Y_2 & = & & & X_2 & + & \cdots & + & X_n & \Leftarrow \text{ original system}
\end{array}$$

$$\begin{array}{ccccccc}
Y_3 & = & X_3 & + & \cdots & + & X_n \\
\vdots & \vdots & & \ddots & \ddots & \vdots & \vdots \\
Y_{n-1} & = & & & X_{n-1} & + & X_n \\
Y_n & = & & & & & X_n
\end{array} \quad \Leftarrow \text{ auxiliary variables}$$

Now, we have n equations in n unknowns: $\mathbf{y} = (y_1, ..., y_n)^T = \mathbf{A}\mathbf{x}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & 1 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

$\mathbf{A}^{-1}$ is easily calculated from the n individual equations

$$\begin{aligned}
x_n^o &= y_n \\
x_{n-1}^o &= y_{n-1} - y_n \\
&\vdots \\
x_k^o &= y_k - y_{k+1} \\
&\vdots \\
x_2^o &= y_2 - y_3 \\
x_1^o &= y_1 - y_2.
\end{aligned}$$

Thus

$$\mathbf{A}^{-1} = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 & -1 \\ 0 & \cdots & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Clearly $\det \mathbf{A} = \det\left(\mathbf{A}^{-1}\right) = 1$, and so

$$\begin{aligned} f_{\mathbf{Y}}(y_1, ..., y_n) &= |\det \mathbf{A}| f_{\mathbf{X}}(x_1^o, ..., x_n^o) \\ &= f_{\mathbf{X}}(y_1 - y_2, y_2 - y_3, y_3 - y_4, ..., y_n). \end{aligned}$$

To obtain $f_{\mathbf{Y}}(y_1, y_2)$, we would have to integrate out with respect to $y_3, y_4, ..., y_n$, i.e.

$$f_{\mathbf{Y}}(y_1, y_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(y_1 - y_2, y_2 - y_3, y_3 - y_4, ..., y_n) dy_3 dy_4 \cdots dy_n$$