2.4 Exercises answer key

(1)

a The marginal distribution of Y is

$$\Pr(Y = y_j) = \sum_{k=1}^{K} p_{jk},$$

and the marginal distribution of X is

$$\Pr(X = x_k) = \sum_{j=1}^{J} p_{jk}.$$

b The randomness of $E[Y \mid X]$ solely comes from the randomness of X. Given $X = x_k$, $E[Y \mid X = x_k]$ is constant,

$$E[Y \mid X = x_k] = \sum_{j=1}^{J} y_j \frac{p_{jk}}{\sum_{j=1}^{J} p_{jk}}.$$

Therefore, the random variable $E[Y \mid X]$ takes K values, with the probability $\Pr(X = x_k)$ for taking kth value.

c

$$\begin{aligned}
E[Y] &= \sum_{j=1}^{J} [y_j \Pr(Y = y_j)] \\
&= \sum_{j=1}^{J} \left[y_j \sum_{k=1}^{K} p_{jk} \right] \\
&= \sum_{j=1}^{J} \left[\sum_{k=1}^{K} y_j p_{jk} \right] \\
&= \sum_{k=1}^{K} \left[\sum_{j=1}^{J} y_j p_{jk} \right] \\
&= \sum_{k=1}^{K} \left[\sum_{j=1}^{J} y_j \frac{p_{jk}}{\sum_{j=1}^{J} p_{jk}} \sum_{j=1}^{J} p_{jk} \right] \\
&= \sum_{k=1}^{K} \left[\left(\sum_{j=1}^{J} y_j \frac{p_{jk}}{\sum_{j=1}^{J} p_{jk}} \right) \sum_{j=1}^{J} p_{jk} \right] \\
&= \sum_{k=1}^{K} [E[Y \mid X = x_k] \Pr(X = x_k)] \\
&= E[E[Y \mid X]]
\end{aligned}$$

2

$$
\begin{aligned}
Var(Y \mid X) \overset{def}{=} & E\left[(Y - E(Y \mid X))^2 \mid X\right] \\
= & E\left[(Y - E(Y \mid X, Z) + E(Y \mid X, Z) - E(Y \mid X))^2 \mid X\right] \\
= & E[(Y - E(Y \mid X, Z))^2 + (E(Y \mid X, Z) - E(Y \mid X))^2 \\
& - 2(Y - E(Y \mid X, Z))(E(Y \mid X, Z) - E(Y \mid X)) \mid X] \\
= & E\left[(Y - E(Y \mid X, Z))^2 \mid X\right] + E\left[(E(Y \mid X, Z) - E(Y \mid X))^2 \mid X\right] \\
& - 2E\left[(Y - E(Y \mid X, Z))(E(Y \mid X, Z) - E(Y \mid X)) \mid X\right] \\
= & E\left[E\left[(Y - E(Y \mid X, Z))^2 \mid X, Z\right] \mid X\right] + E\left[(E(Y \mid X, Z) - E\left[E(Y \mid X, Z)\right])^2 \mid X\right] \\
= & E\left[Var(Y \mid X, Z) \mid X\right] + Var\left[E\left[Y \mid X, Z\right] \mid X\right]
\end{aligned}
$$

3

$$
E(XX') = \begin{bmatrix} E[X_1X_1] & E[X_1X_2] & \cdots & E[X_1X_K] \\ E[X_2X_1] & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ E[X_KX_1] & \cdots & \cdots & E[X_KX_K] \end{bmatrix}
$$

$$
Var(X) = \begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_K) \\ Cov(X_2, X_1) & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ Cov(X_K, X_1) & \cdots & \cdots & Var(X_K) \end{bmatrix}
$$

$$
Cov(X, Y) = \begin{bmatrix} Cov(X_1, Y) \\ Cov(X_2, Y) \\ \vdots \\ Cov(X_K, Y) \end{bmatrix}
$$

4

Without loss of generality, assume the dimension of A is (I, J). Denote the (i, j) element of A as $A_{i,j}$. By the definition of transpose, $A_{i,j} = A'_{j,i}$, and $(A')'_{i,j} = A'_{j,i}$. Therefore, $A_{i,j} = (A')'_{i,j}$ $\forall i, j$, which implies $A = (A')'$.

5

$$
A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}
$$

$$
B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1n} \\ B_{21} & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mn} \end{bmatrix} = \begin{bmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_n \end{bmatrix}
$$

By the rule of matrix multiplication,

$$(AB)_{ij} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots A_{in}B_{nj} = \sum_{z=1}^{n} A_{iz}B_{zj}.$$

Note

$$(a_z b_z')_{ij} = A_{iz}B_{zj},$$

therefore

$$AB = \sum_{z=1}^{n} a_z b_z'.$$

Similar proof can be applied to BA.

6

First we show that $\left(a_j b_j'\right)' = b_j a_j'$.

$$\begin{aligned}
\left(a_j b_j'\right)' &= \left(\begin{bmatrix} A_{1j}B_{j1} & A_{1j}B_{j2} & \cdots & A_{1j}B_{jn} \\ A_{2j}B_{j2} & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ A_{nj}B_{jn} & \cdots & \cdots & A_{nj}B_{jn} \end{bmatrix}\right)' \\
&= \begin{bmatrix} A_{1j}B_{j1} & A_{2j}B_{j2} & \cdots & A_{nj}B_{jn} \\ A_{1j}B_{j2} & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ A_{1j}B_{jn} & \cdots & \cdots & A_{nj}B_{jn} \end{bmatrix} \\
&= \begin{bmatrix} B_{j1}A_{1j} & B_{j2}A_{2j} & \cdots & B_{jn}A_{nj} \\ B_{j2}A_{1j} & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ B_{jn}A_{1j} & \cdots & \cdots & B_{jn}A_{nj} \end{bmatrix} \\
&= b_j a_j'.
\end{aligned}$$

Using this, we have

$$(AB)' = \left(\sum_{j=1}^{n} a_j b_j'\right)' = \sum_{j=1}^{n} \left(a_j b_j'\right)' = \sum_{j=1}^{n} b_j a_j' = B'A'.$$

7

Let's prove it by contraposition. Denote A as

$$A = \begin{bmatrix} a_1' \\ a_2' \\ \vdots \\ a_m' \end{bmatrix},$$

where a_i' is the ith row with length n. Then

$$A' = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \end{bmatrix},$$

and

$$A'A = \sum_{i=1}^{m} a_i a_i'.$$

Notice that $a_i a_i'$ has nonnegative diagonal for any i. If $A \neq 0$, which implies there is at least one entry $A_{i,j} \neq 0$, then the jth diagonal element of $a_i a_i'$ is positive, and therefore the jth diagonal element is positive, $A'A \neq 0$.

8

The row rank of an $m \times n$ matrix A is the number of linear independent row vectors of A, or the dimension of the space spanned by the row vectors. The column rank of an $m \times n$ matrix A is the number of linear independent column vectors of A, or the dimension of the space spanned by the column vectors.

If the row rank of A is 1, it means there is only one linearly independent row vector. Denote one non-zero row vector of A as C_1, then $C_i = a_i C_1 \ \forall i = 1, ..., m$ where $a_i \in \mathbb{R}$.

$$A = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mn} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ a_2 C_{11} & a_2 C_{12} & \cdots & a_2 C_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_m C_{11} & a_m C_{12} & \cdots & a_m C_{1n} \end{bmatrix} = \begin{bmatrix} C_{11} & \frac{C_{12}}{C_{11}} C_{11} & \cdots & \frac{C_{1n}}{C_{11}} C_{11} \\ a_2 C_{11} & \frac{C_{12}}{C_{11}} a_2 C_{11} & \cdots & \frac{C_{1n}}{C_{11}} a_2 C_{11} \\ \vdots & \vdots & \vdots & \vdots \\ a_m C_{m1} & \frac{C_{12}}{C_{11}} a_m C_{11} & \cdots & \frac{C_{1n}}{C_{11}} a_m C_{11} \end{bmatrix}.$$

Denote the column vector of A as $D_j, j = 1, ..., n$. We can see $D_j = \frac{C_{1j}}{C_{11}} D_1 \ \forall j \geq 2$. Therefore, the column rank is also 1.

See a more general proof here: https://ocw.mit.edu/courses/mathematics/18-701-algebra-i-fall-2010/study-materials/MIT18_701F10_rrk_crk.pdf

9

a $AB = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix} B = \begin{bmatrix} B_{11}a_1 + B_{21}a_2 + \cdots B_{n1}a_n & B_{12}a_1 + B_{22}a_2 + \cdots B_{n2}a_n & \cdots & B_{1m}a_1 + B_{2m}a_2 + \cdots B_{nm}a_n \end{bmatrix}$
Notice that each column of AB is a linear combination of columns of A. The dimension of space spanned by linear combinations of the column vectors of A is no bigger than the dimension of the space spanned by the column vectors of A. By definition, $rank(AB) \leq rank(A)$.

b $AB = A \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} A_{11}b_1 + A_{12}b_2 + \cdots A_{1n}b_n \\ \vdots \\ A_{m1}b_1 + A_{m2}b_2 + \cdots A_{mn}b_n \end{bmatrix}$. Similar argument in (a) applies here.

c The coclusion in Q8 and 9(a)(b) implies $rank(AB) \leq rank(A)$ and $rank(AB) \leq rank(B)$.

10

$$AB = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mn} \end{bmatrix} \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1m} \\ D_{21} & D_{22} & \cdots & D_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ D_{n1} & D_{n2} & \cdots & D_{nm} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} C_{1i}D_{i1} & \cdots & \cdots & \cdots \\ \vdots & \sum_{i=1}^{n} C_{2i}D_{i2} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \sum_{i=1}^{n} C_{mi}D_{im} \end{bmatrix}$$

$$BA = \begin{bmatrix} D_{11} & D_{12} & \cdots & D_{1m} \\ D_{21} & D_{22} & \cdots & D_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ D_{n1} & D_{n2} & \cdots & D_{nm} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mn} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} D_{1j}C_{j1} & \cdots & \cdots & \cdots \\ \vdots & \sum_{j=1}^{m} D_{2j}C_{j2} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \sum_{j=1}^{m} D_{nj}C_{jn} \end{bmatrix}$$

$$trace(AB) = \sum_{j=1}^{m} \sum_{i=1}^{n} C_{ji}D_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{m} D_{ij}C_{ji} = trace(BA).$$