SI140 Discussion 10

Li Zeng, Tao Huang, Xinyi Liu

ShanghaiTech University, China
{zengli,huangtao1,liuxy10}@shanghaitech.edu.cn

1 Joint Distribution

1.1 Random vectors and joint distributions

Recall that a random variable X is a real-valued function on the sample space $(S, \mathcal{E}, P)$, where P is a probability measure on S; and that it induces a probability measure P_X on $\boldsymbol{R}$, called the distribution of X, given by

$$P_X(B) = P(X \in B) = P(\{s \in S : X(s) \in B\})$$

for every subset of $\boldsymbol{R}$. The distribution is enough to calculate the expectation of any (Borel) function of X.

Now suppose I have more than one random variable on the same sample space. Then I can consider the random vector (X, Y), or $\boldsymbol{X} = (X_1, \ldots, X_n)$ - A random vector X defines a probability P_X on R^n, called the distribution of X via:

$$P_{\boldsymbol{X}}(B) = P(\boldsymbol{X} \in B) = P\left\{s \in S : (X_1(s), \ldots, X_n(s)) \in B\right\}$$

for every subset of R^n. This distribution is also called the joint distribution of $X_1, \ldots, X_n$ We can use this to define a joint cumulative distribution function, denoted F_X, by

$$F_X\left(x_1, \ldots, x_n\right) = P\left(X_i \leq x_i, \text{ for all } i = 1, \ldots, n\right)$$

1.2 Joint PMFs & PDFs

Joint PMFs A random vector X on a probability space $(S, \mathcal{E}, P)$ is discrete, if you can enumerate its range. When $X_1, \ldots, X_n$ are discrete, the joint probability mass function of the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$ is usually denoted $p_{\boldsymbol{X}}$, and is given by $p_{\boldsymbol{X}}\left(x_1, x_2, \ldots, x_n\right) = P\left(X_1 = x_1\right.$ and $X_2 = x_2$ and $\cdots$ and $\left.X_n = x_n\right)$ If X and Y are independent random variables, then $p_{X,Y}(x, y) = p_X(x) p_Y(y)$. For a function g of X and Y we have

$$\boldsymbol{E} g(X, Y) = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$

Joint PDFs Let X and Y be random variables on a probability space $(S, \mathcal{E}, P)$. The random vector (X, Y) has a joint density $f_{X,Y}(x, y)$ if for every (Borel) subset $B \subset R^2$

$$P((X, Y) \in B) = \iint_B f_{X,Y}(x, y) dx dy$$

If X and Y are independent, then $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. For example,

$$P(X \geq Y) = \int_{-\infty}^{\infty} \int_y^{\infty} f_{X,Y}(x, y) dx dy = \int_{-\infty}^{\infty} f_Y(y) \int_y^{\infty} f_X(x) dx dy$$

For a function g of X and Y we have

$$\boldsymbol{E} g(X, Y) = \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy$$

Exercise 1 (Independence Depending on Range). Suppose that you have a disk surface, modeled as a circle of radius R. Suppose that you know that there is a single point imperfection uniformly distributed on the disk. Therefore the coordinates (X, Y) of this imperfection is distributed according to the following joint PDF:

$$f(x,y) = \begin{cases} \frac{1}{\pi R^2} & x^2+y^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases}$$

What are the marginal distributions of X and Y? Are X and Y independent?

Solution 1. Solution: To compute $f_X(x)$, the marginal PDF of X, we note that we must integrate over y in the support, where $x^2+y^2 \leq R^2$ and therefore $y \in \left[-\sqrt{R^2-x^2}, \sqrt{R^2-x^2}\right]$ for $-R \leq x \leq R$.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \frac{1}{\pi R^2}\int_{x^2+y^2\leq R^2} dy \quad \text{where } -R \leq x \leq R \\ &= \frac{1}{\pi R^2}\int_{y=-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy = \frac{2\sqrt{R^2-x^2}}{\pi R^2} \end{aligned}$$

By symmetry, we observe that we could switch x and y above and obtain the marginal PDF of y :

$$f_Y(y) = \frac{2\sqrt{R^2-y^2}}{\pi R^2} \text{ where } -R \leq y \leq R$$

X and Y are not independent; they are dependent because $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$

1.3 Marginal distributions

The marginal distribution of X is just the distribution P_X of X alone. We can recover its probability mass function from the joint probability mass function $P_{X,Y}$ as follows.

- In the discrete case:

  $$p_X(x) = P(X=x) = \sum_y p_{X,Y}(x,y)$$

  Likewise

  $$p_Y(y) = P(Y=y) = \sum_y p_{X,Y}(x,y)$$

  If X and Y are independent random variables, then $p_{X,Y}(x,y) = p_X(x)p_Y(y)$
- For the continuous case, the marginal density of X, denoted f_X is given by

  $$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

  and the marginal density f_Y of Y is given by

  $$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

1.4 The distribution of a sum

- In the discrete case, let $Z = X+Y$. Discrete case:

  $$P(Z=z) = \sum_{(x,y):x+y=z} P(x,y) = \sum_{\text{all } x} p_{X,Y}(x, z-x)$$

– For the continuous case, If (X, Y) has joint density $f_{X,Y}(x, y)$, what is the density of $X + Y$. Recall that the density is the derivative of the CDF, so

$$
\begin{aligned}
f_{X+Y}(t) = \frac{d}{dt} P(X + Y \le t) &= \frac{d}{dt} \iint_{\{(x,y):x \le t-y\}} f_{X,Y}(x, y) dx dy \\
&= \frac{d}{dt} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{t-y} f_{X,Y}(x, y) dx \right) dy \\
&= \int_{-\infty}^{\infty} \frac{d}{dt} \left(\int_{-\infty}^{t-y} f_{X,Y}(x, y) dx \right) dy \\
&= \int_{-\infty}^{\infty} f_{X,Y}(t - y, y) dy
\end{aligned}
$$

So if X and Y are independent, we get the convolution

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(t - y) f_Y(y) dy$$

2 Covariance & Correlation

When X and Y are independent, we proved

$$\operatorname{Var}(X + Y) = \operatorname{Var} X + \operatorname{Var} Y$$

More generally however, since expectation is a positive linear operator,

$$
\begin{aligned}
\operatorname{Var}(X + Y) &= E((X + Y) - E(X + Y))^2 \\
&= E((X - EX) + (Y - EY))^2 \\
&= E\left((X - EX)^2 + 2(X - EX)(Y - EY) + (Y - EY)^2\right. \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2E(X - EX)(Y - EY) \\
&= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \operatorname{Cov}(X, Y)
\end{aligned}
$$

Definition 1 (Covariance). *The covariance of X and Y is defined to be*

$$
\begin{aligned}
\operatorname{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

Remark 1. It follows immediately that for any random variable X, $\operatorname{Cov}(X, X) = Var(X)$

Remark 2. The product $(X - EX)(Y - EY)$ is positive at outcomes s where $X(s)$ and $Y(s)$ are either both above or both below their means, and negative when one is above and the other below. So one very loose interpretation of positive covariance is that the random variables are probably both above average or below average rather than not. Of course this is just a tendency.

2.1 Properties of covariance

1. $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$
2. $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$
3. $\operatorname{Cov}(X, c) = 0$ for any constant c $\operatorname{Cov}(a \cdot X, Y) = a \cdot \operatorname{Cov}(X, Y)$ for any constant a
4. $\operatorname{Cov}(X + Y, Z) = \operatorname{Cov}(X, Z) + \operatorname{Cov}(Y, Z)$
5. $\operatorname{Cov}(X + Y, Z + W) = \operatorname{Cov}(X, Z) + \operatorname{Cov}(X, W) + \operatorname{Cov}(Y, Z) + \operatorname{Cov}(Y, W)$
6. $\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \operatorname{Cov}(X, Y)$
7. For n r.v.s $X_1, \ldots, X_n$

$$\operatorname{Var}(X_1 + \ldots + X_n) = \operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n) + 2 \sum_{i<j} \operatorname{Cov}(X_i, Y_j)$$

Exercise 2. A scientist makes two measurements, considered to be independent standard Normal r.v.s. Find the correlation between the larger and smaller of the values.
Hint: Note that $max(x, y) + min(x, y) = x + y$ and $max(x, y) - min(x, y) = |x - y|$.

Solution 2. Let X and Y be i.i.d $N(0, 1)$ and $M = max(X, Y), L = min(X, Y)$. By the hint,

$$E(M) + E(L) = E(M + L) = E(X + Y) = E(X) + E(Y) = 0,$$

$$E(M) - E(L) = E(M - L) = E|X - Y| = \frac{2}{\pi},$$

where the last equality was shown in class. So $E(M) = \frac{1}{\pi}$, and

$$Cov(M, L) = E(ML) - E(M)E(L) = E(XY) + (EM)^2 = (EM)^2 = \frac{1}{\pi},$$

since $ML = XY$ has mean $E(XY) = E(X)E(Y) = 0$. To obtain the correlation, we also need $Var(M)$ and $Var(L)$. By symmetry of the Normal, $(-X, -Y)$ has the same distribution as (X, Y), so $Var(M) = Var(L)$; call this v. Then

$$E(X - Y)^2 = Var(X - Y) = 2, \text{and also}$$

$$E(X - Y)^2 = E(M - L)^2 = EM^2 + EL^2 - 2E(ML) = 2v + \frac{2}{\pi}.$$

So $v = 1 - \frac{1}{\pi}$ (alternatively, we can get this by taking the variance of both sides of $max(X, Y) + min(X, Y) = X + Y$). Thus

$$Corr(M, L) = \frac{Cov(M, L)}{\sqrt{Var(M)Var(L)}} = \frac{\frac{1}{\pi}}{1 - \frac{1}{\pi}} = \frac{1}{\pi}$$

■

2.2 Correlation

Covariance is interesting because it is a quantitative measurement of the relationship between two variables. Correlation between two random variables, $\rho(X, Y)$ is the covariance of the two variables normalized by the variance of each variable. This normalization cancels the units out and normalizes the measure so that it is always in the range [0,1]:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}}$$

Correlation measures linearity between X and Y.

$$\begin{array}{ll} \rho(X, Y) = 1 & Y = aX + b \text{ where } a = \sigma_y/\sigma_x \\ \rho(X, Y) = -1 & Y = aX + b \text{ where } a = -\sigma_y/\sigma_x \\ \rho(X, Y) = 0 & \text{absence of linear relationship} \end{array}$$

If $\rho(X, Y) = 0$ we say that X and Y are "uncorrelated." If two varaibles are independent, then their correlation will be 0. However, it doesn't go the other way. A correlation of 0 does not imply independence.

When people use the term correlation, they are actually referring to a specific type of correlation called "Pearson" correlation. It measures the degree to which there is a linear relationship between the two variables. An alternative measure is "Spearman" correlation which has a formula almost identical to your regular correlation score, with the exception that the underlying random variables are first transformed into their rank.

Remark 3. If X, Y are independent, then $\rho(X, Y) = 0$, $(\text{Cov}(X, Y) = 0)$ however, **the inverse proposition is not true.**

Example 1 (Covariance = 0, but variables are not independent). Let U, V be independent and identically distributed random variables with $\boldsymbol{E}U = \boldsymbol{E}V = 0$. Define

$$X = U + V, \quad Y = U - V$$

since $\boldsymbol{E}X = \boldsymbol{E}Y = 0$

$$\operatorname{Cov}(X,Y) = \boldsymbol{E}(XY) = \boldsymbol{E}((U+V)(U-V)) = \boldsymbol{E}\left(U^2 - V^2\right) = \boldsymbol{E}U^2 - \boldsymbol{E}V^2 = 0$$

since U and V have the same distribution. But are X and Y independent?

If U and V are integer-valued, then X and Y are also integer-valued, but more importantly they have the same parity. That is, X is odd if and only if Y is odd. (This is a handy fact for KenKen solvers.

So let U and V be independent and assume the values ± 1 and ± 2 , each with probability $1/4$.($\boldsymbol{E}U = \boldsymbol{E}V = 0$.) Then

$$P(X \text{ is odd}) = P(X \text{ is even}) = P(Y \text{ is odd}) = P(Y \text{ is even}) = \frac{1}{2}$$

but

$$P(X \text{ is even and } Y \text{ is odd}) = 0 \neq \frac{1}{4} = P(X \text{ is even})P(Y \text{ is odd})$$

so X and Y are not independent.

3 Multinomial Distribution

The Multinomial distribution is a generalization of the Binomial. Whereas the Binomial distribution counts the successes in a fixed number of trials that can only be categorized as success or failure, the Multinomial distribution keeps track of trials whose outcomes can fall into multiple categories, such as excellent, adequate, poor; or red, yellow, green, blue.

Theorem 1. *(Multinomial joint PMF) If $\mathbf{X} \sim Mult_k(n, \mathbf{p})$, then the joint PMF of $\mathbf{X}$ is*

$$P(X_1 = n_1, ..., X_k = n_k) = \frac{n!}{n_1! n_2! \cdots n_k!} \cdot p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

for $n_1, \cdots, n_k$ satisfying $n_1 + \cdots + n_k = n$.

Theorem 2. *(Multinomial marginals) If $\mathbf{X} \sim Mult_k(n, \mathbf{p})$, then $X_j \sim Bin(n, p_j)$.*

Theorem 3. *(Multinomial lumping) If $\mathbf{X} \sim Mult_k(n, \mathbf{p})$, then for any distinct i and j, $X_i + X_j \sim Bin(n, p_i + p_j)$. The random vector of counts obtained from merging categories i and j is still Multinomial. For example, merging categories 1 and 2 gives*

$$(X_1 + X_2, X_3, ..., X_k) \sim Mult_{k-1}(n, (p_1 + p_2, p_3, ..., p_n))$$

Theorem 4. *(Multinomial conditioning) If $\mathbf{X} \sim Mult_k(n, \mathbf{p})$, then*

$$(X_2, ..., X_k) | X_1 = n_1 \sim Mult_{k-1}(n - n_1, (p_2', ..., p_k')),$$

where $p_j = p_j / (p_2 + ... + p_k)$.

Exercise 3. (Covariance in a Multinomial) Let $(X_1, ..., X_k) \sim \operatorname{Mult}_k(n, \mathbf{p})$ where $\mathbf{p} = (p_1, ..., p_k)$. Show that, for $i \neq j, Cov(X_i, X_j) = -np_i p_j$.

Solution 3. Let $i = 1, j = 2$, without loss of generality. Using the lumping property and the marginal distributions of a Multinomial, we know $X_1 + X_2 \sim Bin(n, p_1 + p_2), X_1 \sim Bin(n, p_1), X_2 \sim Bin(n, p_2)$. Therefore,

$$Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$$

becomes

$$n(p_1 + p_2)(1 - (p_1 + p_2)) = np_1(1 - p_1) + np_2(1 - p_2) + 2Cov(X_1, X_2).$$

Solving for $Cov(X_1, X_2)$ gives $Cov(X_1, X_2) = -np_1 p_2$. Through the same logic, for $i \neq j$, we have $Cov(X_i, X_j) = -np_i p_j$. The components are negatively correlated, as we would expect: if we know there are a lot of objects in category i, then there aren't as many objects left over that could possibly be in category j. ■

Exercise 4. Consider the birthdays of 100 people. Assume people's birthdays are independent, and the 365 days of the year are equally likely. Find the covariance and correlation between how many of the people were born on January 1 and how many were born on January 2.

Solution 4. Let X_j be the number of people born on January j. Then

$$Cov(X_1, X_2) = -\frac{100}{365^2},$$

using the result about covariance in a Multinomial. Since $X_j \sim Bin(100, \frac{1}{365})$, we then have

$$Corr(X_1, X_2) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = -\frac{\frac{100}{365^2}}{100 \times \frac{1}{365} \times \frac{364}{365}} = -\frac{1}{364}.$$

■

4 Multivariate Normal (MVN)/ Joint Gaussian (JG)

In class, we learn the following definition of *Multivariate Normal(MVN) distribution*:

Definition 2 (MVN). *A random vector is said have a Multivariate Normal(MVN) distribution if every linear combination of its k components has a univariate normal distribution.*

Example 2 (An Actual MVN). For $Z, W \overset{i.i.d}{\sim} \mathcal{N}(0, 1), (Z, W)$ is Bivariate Normal because the sum of independent Normals is Normal. Also, $(Z + 2W, 3Z + 5W)$ is Bivariate Normal, since an arbitrary linear combination

$$t_1(Z + 2W) + t_2(3Z + 5W)$$

can also be written as a linear combination of Z and W,

$$(t_1 + 3t_2)\, Z + (2t_1 + 5t_2)\, W$$

which is Normal.

This definition have a parametric version, which has a clearer geometric interpretation. In this section, we will have a brief discussion on the general parametric definition of MVN with its geometric meaning, and some properties of MVN that would be helpful in further study,

4.1 Parameterizations

The multivariate Gaussian distribution is commonly expressed in terms of the parameters μ and Σ For $X \sim \mathcal{N}(\mu, \Sigma)$, we have the following form for the density function:

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\}$$

where

- $\mu \in \mathbb{R}^n$ is the mean vector. x is also a vector in $\mathbb{R}^n$.
- $\Sigma \in \mathbb{R}^{n \times n}$ is the covariance matrix, $\Sigma = \Sigma^T > 0$.

4.2 Properties

- In the convairance matrix, the entry of ith row and jth column is $Cov(X_i, X_j)$.
- All the random vector components are independent if and only if Σ is diagonal, when PDF simplifies to the product of n scalar MVNs:

$$f_X(x) = \prod_{i=1,\ldots,D} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)$$

– For two MVN r.v.s, uncorrelated $\iff$ independent.
– Some facts that does not require your proof.
  - An (affine) linear transformation of a MVN r.v. is a MVN r.v.
  - A linear combination of two jointly MVN r.v.s is a MVN r.v.
  - Conditioning jointly MVN r.v.s gives a MVN r.v.

Exercise 5. Show that in bivariate normal case, uncorrelated $\iff$ independent.Given

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

where ρ is the correlation between X and Y and where $\sigma_X > 0$ and $\sigma_Y > 0$. In this case,

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

Solution 5.

4.3 Geometric Interpretation

The equidensity contours of a non-singular multivariate normal distribution are ellipsoids centered at the mean. The directions of the principal axes of the ellipsoids are given by the eigenvectors of the covariance matrix $\boldsymbol{\Sigma}$. The squared relative lengths of the principal axes are given by the corresponding eigenvalues.

If $\Sigma = \mathrm{U} \cap \mathrm{U}^{\mathrm{T}} = \mathrm{U}\Lambda^{1/2}\left(\mathrm{U}\Lambda^{1/2}\right)^{\mathrm{T}}$ is an eigen decomposition where the columns of U are unit eigen vectors and Λ is a diagonal matrix of the eigenvalues, then we have

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \iff \mathbf{X} \sim \boldsymbol{\mu} + \mathbf{U}\boldsymbol{\Lambda}^{1/2}\mathcal{N}(0, \mathbf{I}) \iff \mathbf{X} \sim \boldsymbol{\mu} + \mathbf{U}\mathcal{N}(0, \boldsymbol{\Lambda})$$

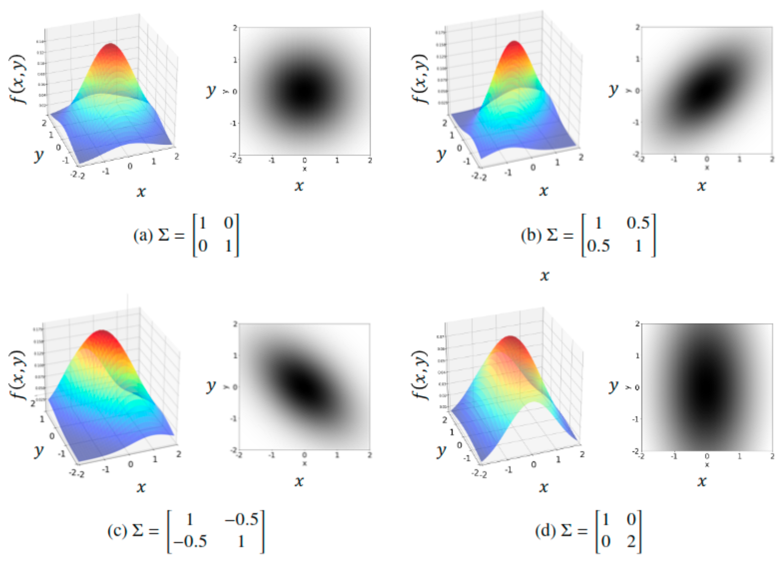

Fig. 1. Visualization: the 3-D view and top-down view for Bivariate Normal r.v.s with mean **0**.

Exercise 6 (Bivariate Normal generation). Suppose that we have access to i.i.d. *r.v.s* $X, Y \sim \mathcal{N}(0,1)$, but want to generate a Bivariate Normal (Z, W) with $\mathrm{Corr}(Z, W) = \rho$ and Z,W marginally $\mathcal{N}(0,1)$, for the purpose of running a simulation. How can we construct Z and W from linear combinations of X and Y?

Solution 6. By definition of Multivariate Normal, any (Z, W) of the form

$$Z = aX + bY$$
$$W = cX + dY$$

will be Bivariate Normal. So let's try to find suitable a, b, c, d.$\mathrm{E}(Z) = \mathrm{E}(W) = 0$ has been satisfied. Setting the variances equal to 1 gives

$$\mathrm{Var}(Z) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) = a^2 + b^2 = 1$$
$$\mathrm{Var}(W) = c^2\,\mathrm{Var}(X) + d^2\,\mathrm{Var}(Y) = c^2 + d^2 = 1$$

Setting the covariance of Z and W equal to ρ gives

$$\mathrm{Corr}(Z, W) = \rho$$

$$\Rightarrow \mathrm{Cov}(aX + bY, cX + dY) = \rho$$
$$\Rightarrow ac\,\mathrm{Var}(X) + bd\,\mathrm{Var}(Y) = \rho$$
$$\Rightarrow ac + bd = \rho$$

In conclusion, we need to find the solution of the equation set

$$\begin{cases} a^2 + b^2 = 1 \\ c^2 + d^2 = 1 \\ ac + bd = \rho \end{cases}$$

There are more unknowns than equations here, and we just need one solution. To simplify, let's look for a solution with $b = 0$. Then $a^2 = 1$, so let's take $a = 1$. Now $ac + bd = \rho$ reduces to $c = \rho$, and then we can use $c^2 + d^2 = 1$ to find $d = \sqrt{1 - \rho^2}$. Putting everything together, we can generate (Z, W) as

$$Z = X$$
$$W = \rho X + \sqrt{1 - \rho^2}Y$$

Exercise 7 (Extension: The Decomposition of a given JG). Provided that $\mathbf{X} \sim (\mu, \mathbf{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1/4 \\ 1/4 & 1 \end{pmatrix}$$

Show how to decompose X into linear combinations of 2 i.i.d $\mathcal{N}(0, 1)$

Solution 7.