Probability distributions EBP038A05
Lecture slides

Nicky van Foreest, Ruben van Beesten, Arpan Rijal

March 1, 2022

1

WEEK 1

1.1 LECTURE 1

Ex 1.1.1. Consider 12 football players on a football field. Eleven of them are players of F.C. Barcelona, the other one is an arbiter. We select a random player, uniform. This player must take a penalty. The probability that a player of Barcelona scores is 70%, for the arbiter it is 50%. Let $P \in \{A, B\}$ be r.v that corresponds to the selected player, and $S \in \{0, 1\}$ be the score.

1. What is the PMF? In other words, determine $\mathsf{P}\{P = B, S = 1\}$ and so on for all possibilities.
2. What is $\mathsf{P}\{S = 1\}$? What is $\mathsf{P}\{P = B\}$?
3. Show that S and P are dependent.

An insurance company receives on a certain day two claims $X, Y \geq 0$. We will find the PMF of the loss $Z = X + Y$ under different assumptions.

The joint CDF $F_{X,Y}$ and joint PMF $p_{X,Y}$ are assumed known.

Ex 1.1.2. Why is it not interesting to consider the case $\{X = 0, Y = 0\}$?

Ex 1.1.3. Find an expression for the PMF of $Z = X + Y$.

Suppose $p_{X,Y}(i, j) = c\, I_{i=j}\, I_{1 \leq i \leq 4}$.

Ex 1.1.4. What is c?

Ex 1.1.5. What is $F_X(i)$? What is $F_Y(j)$?

Ex 1.1.6. Are X and Y dependent? If so, why, because $1 = F_{X,Y}(4, 4) = F_X(4) F_Y(4)$?

Ex 1.1.7. What is $\mathsf{P}\{Z = k\}$?

Ex 1.1.8. What is $\mathsf{V}[Z]$?

Now take X, Y iid $\sim$ Unif($\{1, 2, 3, 4\}$) (so now no longer $p_{X,Y}(i, j) = I_{i=j}\, I_{1 \leq i \leq 4}$).

Ex 1.1.9. What is $\mathsf{P}\{Z = 4\}$?

Remark 1.1.10. We can make lots of variations on this theme.

1. Let $X \in \{1,2,3\}$ and $Y \in \{1,2,3,4\}$.
2. Take $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$. (Use the chicken-egg story)
3. We can make X and Y such that they are (both) continuous, i.e., have densities. The conceptual ideas[1] don't change much, except that the summations become integrals.
4. Why do people often/sometimes (?) model the claim sizes as iid $\sim \text{Norm}(\mu, \sigma^2)$? There is a slight problem with this model (can real claim sizes be negative?), but what is the way out?
5. The example is more versatile than you might think. Here is another interpretation.

   A supermarket has 5 packets of rice on the shelf. Two customers buy rice, with amounts X and Y. What is the probability of a lost sale, i.e., $\mathsf{P}\{X+Y>5\}$? What is the expected amount lost, i.e., $\mathsf{E}[\max\{X+Y-5,0\}]$?

   Here is yet another. Two patients arrive in to the first aid of a hospital. They need X and Y amounts of service, and there is one doctor. When both patients arrive at 2 pm, what is the probability that the doctor has work in overtime (after 5 pm), i.e., $\mathsf{P}\{X+Y>5-2\}$?

Ex 1.1.11. We have a continuous r.v. $X \geq 0$ with finite expectation. Use 2D integration and indicators to prove that

$$\mathsf{E}[X] = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0 G(x)\,\mathrm{d}x, \tag{1.1.1}$$

where $G(x)$ is the survival function.

Ex 1.1.12. A variation on BH.7.1. Alice is prepared to wait 20 minutes for Bob, while Bob doesn't want to wait longer than 10 minutes. What is the probability that they meet?

Use the fundamental bridge and indicator functions to write this probability as a 2D integral. Then use repeated integration to solve the 2D integral.

1.2 LECTURE 2

Ex 1.2.1. Let $L = \min\{X,Y\}$, where $X, Y \sim \text{Geo}(p)$ and independent. What is the domain of L? Then, use the fundamental bridge and 2D LOTUS to show that

$$\mathsf{P}\{L \geq i\} = q^{2i} \implies L \sim \text{Geo}(1-q^2).$$

Ex 1.2.2. Let $M = \max\{X,Y\}$, where $X, Y \sim \text{Geo}(p)$ and independent. Show that

$$\mathsf{P}\{M=k\} = 2pq^k(1-q^k) + p^2q^{2k}.$$

---

[1] Unless you start digging deeper. Then things change drastically, but we skip this technical stuff.

Ex 1.2.3. Explain that

$$\mathsf{P}\{L = i, M = k\} = 2p^2 q^{i+k} I_{k>i} + p^2 q^{2i} I_{i=k}).$$

Ex 1.2.4. With the previous exercise, use marginalization to compute the marginal PMF $\mathsf{P}\{M = k\}$.x

---

Ex 1.2.5. Now take X, Y iid and $\sim \text{Exp}(\lambda)$. Use the fundamental bridge to show that for $u \le v$, the joint CDF has the form

$$F_{L,M}(u, v) = \mathsf{P}\{L \le u, M \le v\} = 2\int_0^u (F_Y(v) - F_Y(x)) f_X(x)\, \mathrm{d}x.$$

Ex 1.2.6. Take partial derivatives to show that for the joint PDF,

$$f_{L,M}(u, v) = 2 f_X(u) f_Y(v)\, I_{u \le v}.$$

2

WEEK 2

2.1 LECTURE 3

Ex 2.1.1. We ask a married woman on the street her height X. What does this tell us about the height Y of her spouse? We suspect that taller/smaller people choose taller/smaller partners, so, given X, a simple estimator $\hat{Y}$ of Y is given by

$$\hat{Y} = aX + b.$$

(What is the sign of a if taller people tend to choose taller people as spouse?) But how to determine a and b? A common method is to find a and b such that the function

$$f(a,b) = \mathsf{E}\left[(Y-\hat{Y})^2\right]$$

is minimized. Show that the optimal values are such that

$$\hat{Y} = \mathsf{E}[Y] + \rho\frac{\sigma_Y}{\sigma_X}(X - \mathsf{E}[X]),$$

where ρ is the correlation between X and Y and where σ_X and σ_Y are the standard deviations of X and Y respectively.

Ex 2.1.2. Using scaling laws often can help to find errors. For instance, the prediction $\hat{Y}$ should not change whether we measure the height in meters or centimeters. In view of this, explain that

$$\hat{Y} = \mathsf{E}[Y] + \rho\frac{\mathsf{V}[Y]}{\sigma_X}(X - \mathsf{E}[X])$$

must be wrong.

Ex 2.1.3. n people throw their hat in a box. After shuffling, each of them takes out a hat at random. How many people do you expect to take out their own hat (i.e., the hat they put in the box); what is the variance?

$$\begin{aligned}
\mathsf{E}\left[I_{X_i=i}\right] &= 1/n, \quad \text{for all } i. \\
\mathsf{E}[S] &= \sum_{i=1}^n \mathsf{E}\left[I_{X_i=i}\right] = \sum_{i=1}^n 1/n = 1. \\
\mathsf{E}\left[S^2\right] &= \sum_{i=1}^n \mathsf{E}\left[I_{X_i=i}\right] + \sum_{i\neq j} \mathsf{E}\left[I_{X_i=i} I_{X_j=j}\right] = 1 + n(n-1)\cdot\frac{1}{n}\frac{1}{n-1} = 1+1 = 2. \\
\mathsf{V}[S] &= \mathsf{E}\left[S^2\right] - (\mathsf{E}[S])^2 = 2-1 = 1.
\end{aligned}$$

Ex 2.1.4. Continuation of the previous exercise. Write a simulator for compute the expectation and variance.

2.2 LECTURE 4

Ex 2.2.1. BH.7.65 Let $(X_1, \dots, X_k)$ be Multinomial with parameters n and $(p_1, \dots, p_k)$. Use indicator rvs to show that $\mathsf{Cov}\left[X_i, X_j\right] = -np_ip_j$ for $i \neq j$.

Ex 2.2.2. Suppose (X, Y) are bi-variate normal distributed with mean vector $\mu = (\mu_X, \mu_Y) = (0, 0)$, standard deviations $\sigma_X = \sigma_Y = 1$ and correlation ρ_{XY} between X and Y. Specify the joint pdf of X and $X + Y$.

The following exercises will show how probability theory can be used in finance. We will look at the trade off between risk and return in a financial portfolio.

John is an investor who has $\$10,000$ to invest. There are three stocks he can choose from. The returns on investment (A, B, C) of these three stocks over the following year (in terms of percentages) follow a Multivariate Normal distribution. The expected returns on investment are $\mu_A = 7.5\%$, $\mu_B = 10\%$, $\mu_C = 20\%$. The corresponding standard deviations are $\sigma_A = 7\%$, $\sigma_B = 12\%$ and $\sigma_C = 17\%$. Note that risk (measured in standard deviation) increases with expected return. The correlation coefficients between the different returns are $\rho_{AB} = 0.7$, $\rho_{AC} = -0.8$, $\rho_{BC} = -0.3$.

Ex 2.2.3. Suppose the investor decides to invest $\$2,000$ in stock A, $\$4,000$ in stock B, $\$2,000$ in stock C and to put the remaining $\$2,000$ in a savings account with a zero interest rate. What the expected value of his portfolio after a year?

Ex 2.2.4. What is the standard deviation of the value of the portfolio in a year?

Ex 2.2.5. John does not like losing money. What is his probability of having made a net loss after a year?

John has a friend named Mary, who is a first-year EOR student. She has never invested money herself, but she is paying close attention during the course Probability Distributions. She tells her friend: "John, your investment plan does not make a lot of sense. You can easily get a higher expected return at a lower level of risk!"

Ex 2.2.6. Show that Mary is right. That is, make a portfolio with a higher expected return, but with a lower standard deviation.
Hint: Make use of the ***negative correlation*** *between C and the other two stocks!*

---

3 WEEK 3

3.1 LECTURE 5

HERE IS A NICE geometrical explanation of how the normal distribution originates.

Ex 3.1.1. Suppose $z_0 = (x_0, y_0)$ is the target on a dart board at which Barney (our national darts hero) aims, but you can also interpret it as the true position of a star in the sky. Let z be the actual position at which the dart of Barney lands on the board, or the measured position of the star. For ease, take z_0 as the origin, i.e., $z_0 = (0,0)$. Then make the following assumptions:

1. The disturbance (x, y) has the same distribution in any direction.
2. The disturbance (x, y) along the x direction and the y direction are independent.
3. Large disturbances are less likely than small disturbances.

Show that the disturbance along the x-axis (hence y-axis) is normally distributed. You can use BH.8.17 as a source of inspiration. (This is perhaps a hard exercise, but the solution is easy to understand and very useful to memorize.)

We next find the normalizing constant of the normal distribution (and thereby offer an opportunity to practice with change of variables).

Ex 3.1.2. For this purpose consider two circles in the plane: $C(N)$ with radius N and $C(\sqrt{2}N)$ with radius $\sqrt{2}N$. It is obvious that the square $S(N) = [-N, N] \times [-N, N]$ contains the first circle, and is contained in the second. Therefore,

$$\iint_{C(N)} f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y \leq \iint_{S(N)} f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y \leq \iint_{C(\sqrt{2}N)} f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y. \tag{3.1.1}$$

Now substitute the normal distribution of [3.1.1]. Then use polar coordinates (See BH.8.1.9) to solve the integrals over the circles, and derive the normalization constant.

BENFORD'S LAW MAKES a statement on the first significant digit of numbers. Look it up on the web; it is a fascinating law. It's used to detect fraud by insurance companies and the tax department, but also to see whether the US elections in 2020 have been rigged, or whether authorities manipulate the statistics of the number of deceased by Covid. You can find the rest of the analysis in Section 5.5 of 'The art of probability for scientists and engineers' by R.W. Hamming. The next exercise is a first step in the analysis of Benford's law.

Ex 3.1.3. Let X, Y be iid with density f and support $[1, 10)$. Find an expression for the density of $Z = XY$. What is the support (domain) of Z? If $X, Y \sim \mathrm{Unif}([1, 10))$, what is f_Z?

3.2 LECTURE 6

Ex 3.2.1. Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. What is the distribution of $Z = X + Y$?

Ex 3.2.2. (BH.8.4.3.) Let $X_1, X_2, \ldots$ be i.i.d. $\text{Exp}(\lambda)$ distributed. Let $T_n = \sum_{k=1}^n X_k$. Show that T_n has the following pdf:

$$f_{T_n}(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}, \quad t > 0. \tag{3.2.1}$$

That is, show that T_n follows a *Gamma distribution* with parameters n and λ. (We will learn about the Gamma distribution in BH.8.4.)

Ex 3.2.3. Let X, Y be i.i.d. $\mathcal{N}(0,1)$ distributed and define $Z = X + Y$. Show that $Z \sim \mathcal{N}(0,2)$ using a convolution integral.

4

WEEK 4

WEEK 5

5

6

WEEK 6

WEEK 7

7.1 HINTS

h.1.1.11. Check the proof of BH.4.4.8

h.1.2.1. The fundamental bridge and 2D LOTUS have the general form

$$\mathsf{P}\left\{g(X,Y\right\} \in A\} = \mathsf{E}\left[I_{g(X,Y)\in A}\right] = \sum_i \sum_j I_{g(i,j)\in A}\, \mathsf{P}\left\{X=i, Y=j\right\}.$$

Take $g(i,j) = \min\{i,j\}$.

h.1.2.2. Use 2D LOTUS on $g(x,y) = I_{\max\{x,y\}=k}$.

h.2.1.3. Take $I_{X_i=i}$. When this is 1, person i picks its own hat, and if 0, the person picks somebody else's hat. What is the meaning of $S = \sum_{i=1}^n I_{X_i=i}$?

h.3.2.1. Use the Binomial theorem:

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}, \tag{7.1.1}$$

for any non negative integer n.

h.3.2.2. Use mathematical induction.

7.2 SOLUTIONS

s.1.1.1. Here is the joint PMF:

$$\mathsf{P}\{P=A, S=1\} = \frac{1}{12}0.5 \qquad \mathsf{P}\{P=A, S=0\} = \frac{1}{12}0.5 \tag{7.2.1}$$

$$\mathsf{P}\{P=B, S=1\} = \frac{11}{12}0.7 \qquad \mathsf{P}\{P=B, S=0\} = \frac{11}{12}0.3. \tag{7.2.2}$$

Now the marginal PMFs

$$\mathsf{P}\{S=1\} = \mathsf{P}\{P=A, S=1\} + \mathsf{P}\{P=B, S=1\} = 0.042 + 0.64 = 0.683 = 1 - \mathsf{P}\{S=0\}$$

$$\mathsf{P}\{P=B\} = \frac{11}{12} = 1 - \mathsf{P}\{P=A\}.$$

For independence we take the definition. In general, for all outcomes x, y we must have that $\mathsf{P}\left\{X=x, Y=y\right\} = \mathsf{P}\{X=x\}\,\mathsf{P}\left\{Y=y\right\}$. For our present example, let's check for a particular outcome:

$$\mathsf{P}\{P=B, S=1\} = \frac{11}{12}\cdot 0.7 \neq \mathsf{P}\{P=B\}\,\mathsf{P}\{S=1\} = \frac{11}{12}\cdot 0.683$$

The joint PMF is obviously not the same as the product of the marginals, which implies that P and S are not independent.

s.1.1.2. When the claim sizes are 0, then the insurance company does not receive a claim.

s.1.1.3. By the fundamental bridge,

$$\mathsf{P}\{Z=k\} = \sum_{i,j} I_{i+j=k} p_{X,Y}(i,j) \tag{7.2.3}$$

$$= \sum_{i,j} I_{i,j\geq 0} I_{j=k-i} p_{X,Y}(i,j) \tag{7.2.4}$$

$$= \sum_{i=0}^{k} p_{X,Y}(i,k-i). \tag{7.2.5}$$

s.1.1.4. $c = 1/4$ because there are just four possible values for i and j.

s.1.1.5. Use marginalization:

$$F_X(k) = F_{X,Y}(k,\infty) = \sum_{i\leq k}\sum_{j} p_{X,Y}(i,j) \tag{7.2.6}$$

$$= \frac{1}{4}\sum_{i\leq k}\sum_{j} I_{i=j}\, I_{1\leq i\leq 4} \tag{7.2.7}$$

$$= \frac{1}{4}\sum_{i\leq k} I_{1\leq i\leq 4} \tag{7.2.8}$$

$$= k/4, \tag{7.2.9}$$

$$F_Y(j) = j/4. \tag{7.2.10}$$

s.1.1.6. The equality in the question must hold for all i, j, not only for $i = j = 4$. If you take $i = j = 1$, you'll see immediately that $F_{X,Y}(1,1) \neq F_X(1)F_Y(1)$:

$$\frac{1}{4} = F_{X,Y}(1,1) \neq F_X(1)F_Y(1) = \frac{1}{4}\frac{1}{4}. \tag{7.2.11}$$

s.1.1.7. $\mathsf{P}\{Z=2\} = \mathsf{P}\{X=1, Y=1\} = 1/4 = \mathsf{P}\{Z=4\}$, etc. $\mathsf{P}\{Z=k\} = 0$ for $k \not\in \{2,4,6,8\}$.

s.1.1.8. Here is one approach

$$\mathsf{V}[Z] = \mathsf{E}\left[Z^2\right] - (\mathsf{E}[Z])^2 \tag{7.2.12}$$

$$\mathsf{E}\left[Z^2\right] = \mathsf{E}\left[(X+Y)^2\right] = \mathsf{E}\left[X^2\right] + 2\,\mathsf{E}[XY] + \mathsf{E}\left[Y^2\right] \tag{7.2.13}$$

$$(EZ)^2 = (\mathsf{E}[X] + \mathsf{E}[Y])^2 \tag{7.2.14}$$

$$= (\mathsf{E}[X])^2 + 2\,\mathsf{E}[X]\,\mathsf{E}[Y] + (\mathsf{E}[Y])^2 \tag{7.2.15}$$

$$\Longrightarrow \tag{7.2.16}$$

$$\mathsf{V}[Z] = \mathsf{E}\left[Z^2\right] - (\mathsf{E}[Z])^2 \tag{7.2.17}$$

$$= \mathsf{V}[X] + \mathsf{V}[Y] + 2(\mathsf{E}[XY] - (\mathsf{E}[X]\,\mathsf{E}[Y])) \tag{7.2.18}$$

$$\mathsf{E}[XY] = \sum_{ij} i\,j\,p_{X,Y}(i,j) = \frac{1}{4}(1+4+9+16) = \dots \tag{7.2.19}$$

$$\mathsf{E}\left[X^2\right] = \dots \tag{7.2.20}$$

The numbers are for you to compute.

s.1.1.9.

$$\mathsf{P}\{Z=4\} = \sum_{i,j} I_{i+j=4}\, p_{X,Y}(i,j) \tag{7.2.21}$$
$$= \sum_{i=1}^{4}\sum_{j=1}^{4} I_{j=4-i}\frac{1}{16} \tag{7.2.22}$$
$$= \sum_{i=1}^{3}\frac{1}{16} \tag{7.2.23}$$
$$= \frac{3}{16}. \tag{7.2.24}$$

s.1.1.11. The trick is to realize that $x = \int_0^\infty I_{y\leq x}\,\mathrm{d}y$. Using this,

$$\mathsf{E}[X] = \int_0^\infty x f(x)\,\mathrm{d}x \tag{7.2.25}$$
$$= \int_0^\infty\int_0^\infty I_{y\leq x} f(x)\,\mathrm{d}y\,\mathrm{d}x \tag{7.2.26}$$
$$= \int_0^\infty\int_0^\infty I_{y\leq x} f(x)\,\mathrm{d}x\,\mathrm{d}y \tag{7.2.27}$$
$$= \int_0^\infty\int_0^\infty I_{x\geq y} f(x)\,\mathrm{d}x\,\mathrm{d}y \tag{7.2.28}$$
$$= \int_0^\infty\int_y^\infty f(x)\,\mathrm{d}x\,\mathrm{d}y \tag{7.2.29}$$
$$= \int_0^\infty G(y)\,\mathrm{d}y. \tag{7.2.30}$$

s.1.1.12. Let A, B be the arrival times of Alice and Bob. They meet if $I_{A<B+1/3}\, I_{B<A+1/6}$ is true, i.e., is equal to 1. Therefore, by letting M be the event that they meet:

$$\mathsf{P}\{M\} = \mathsf{E}[I_{A<B+1/3}\, I_{B<A+1/6}] = \int_0^1\int_0^1 I_{x<y+1/3}\, I_{y<x+1/6}\,\mathrm{d}y\,\mathrm{d}x.$$

We can solve this integral by first integrating along y, and then along x. Let's focus on the integral over y first.

$$\begin{aligned}\int_0^1 I_{x<y+1/3}\, I_{y<x+1/6}\,\mathrm{d}y &= \int_0^1 I_{x-1/3<y<x+1/6}\,\mathrm{d}y\\ &= \int_0^1 I_{\max\{0,x-1/3\}<y\min\{1,x+1/6\}}\,\mathrm{d}y\\ &= \min\{1,x+1/6\} - \max\{0,x-1/3\}\end{aligned}$$

Now the integral over x:

$$\begin{aligned}\int_0^1 (\min\{1,x+1/6\} - \max\{0,x-1/3\})\,\mathrm{d}x &= \int_0^1 \min\{1,x+1/6\}\,\mathrm{d}x - \int_0^1 \max\{0,x-1/3\}\,\mathrm{d}x\\ &= \int_0^{5/6}(x+1/6)\,\mathrm{d}x + \int_{5/6}^1 1\,\mathrm{d}x - \int_{1/3}^1 (x-1/3)\,\mathrm{d}x\\ &= 0.5x^2\Big|_0^{5/6} + 1/6\cdot 5/6 - 0.5x^2\Big|_{1/3}^1 + 1/3\cdot 2/3\end{aligned}$$

Of course, we can find the probability with some simple geometric arguments (compute the area of two triangles). However, this does not work any longer if the density is not uniform. Then we have to do the integration, and that is the reason why I show above how to handle the general case.

s.1.2.1. With the hint,

$$
\begin{aligned}
\mathsf{P}\{L \geq k\} &= \sum_i \sum_j I_{\min\{i,j\}\geq k} \mathsf{P}\{X = i, Y = j\} \\
&= \sum_{i\geq k} \sum_{j \geq k} \mathsf{P}\{X = i\} \mathsf{P}\{Y = j\} \\
&= \mathsf{P}\{X \geq k\} \mathsf{P}\{Y \geq k\} = q^k q^k = q^{2k}.
\end{aligned}
$$

$\mathsf{P}\{L > i\}$ has the same form as $\mathsf{P}\{X > i\}$, but now with q^{2i} rather than q^i.

s.1.2.2.

$$
\begin{aligned}
\mathsf{P}\{M = k\} &= \mathsf{P}\{\max\{X, Y\} = k\} \\
&= p^2 \sum_{ij} I_{\max\{i,j\}=k} q^i q^j \\
&= 2p^2 \sum_{ij} I_{i=k} I_{j<k} q^i q^j + p^2 \sum_{ij} I_{i=j=k} q^i q^j \\
&= 2p^2 q^k \sum_{j<k} q^j + p^2 q^{2k} \\
&= 2p^2 q^k \frac{1-q^k}{1-q} + p^2 q^{2k}
\end{aligned}
$$

s.1.2.3.

$$
\begin{aligned}
\mathsf{P}\{L = i, M = k\} &= 2\mathsf{P}\{X = i, Y = k\} I_{k>i} + \mathsf{P}\{X = Y = i\} I_{i=k} \\
&= 2p^2 q^{i+k} I_{k>i} + p^2 q^{2i} I_{i=k}.
\end{aligned}
$$

s.1.2.4.

$$
\begin{aligned}
\mathsf{P}\{M = k\} &= \sum_i \mathsf{P}\{L = i, M = k\} \\
&= \sum_i (2p^2 q^{i+k} I_{k>i} + p^2 q^{2i} I_{i=k}) \\
&= 2p^2 q^k \sum_{i=0}^{k-1} q^i + p^2 q^{2k} \\
&= 2pq^k (1-q^k) + p^2 q^{2k} \\
&= 2pq^k + (p^2 - 2p) q^{2k},
\end{aligned}
$$

s.1.2.5. First the joint distribution. With $u \leq v$,

$$
\begin{aligned}
F_{L,M}(u, v) &= \mathsf{P}\{L \leq u, M \leq v\} \\
&= 2 \iint I_{x\leq u, y\leq v, x\leq y} f_{X,Y}(x, y) \,\mathrm{d}x \,\mathrm{d}y \\
&= 2 \int_0^u \int_x^v f_Y(y) \,\mathrm{d}y f_X(x) \,\mathrm{d}x && \text{independence} \\
&= 2 \int_0^u (F_Y(v) - F_Y(x)) f_X(x) \,\mathrm{d}x.
\end{aligned}
$$

s.1.2.6. Taking partial derivatives,

$$\begin{aligned}
f_{L,M}(u,v) &= \partial_v \partial_u F_{L,M}(u,v) \\
&= 2\partial_v\partial_u \int_0^u (F_Y(v) - F_Y(x)) f_X(x)\,\mathrm{d}x \\
&= 2\partial_v \left\{(F_Y(v) - F_Y(u)) f_X(u)\right\} \\
&= 2 f_X(u) \partial_v F_Y(v) \\
&= 2 f_X(u) f_Y(v).
\end{aligned}$$

s.2.1.1. We take the partial derivatives of f with respect to a and b, and solve for a and b. In the derivation, we use that

$$\rho = \frac{\mathsf{Cov}[X,Y]}{\sqrt{\mathsf{V}[X]\,\mathsf{V}[Y]}} = \frac{\mathsf{Cov}[X,Y]}{\sigma_X\sigma_Y} \implies \rho\frac{\sigma_Y}{\sigma_X} = \frac{\mathsf{Cov}[X,Y]}{\mathsf{V}[X]}. \tag{7.2.31}$$

Hence,

$$\begin{aligned}
f(a,b) &= \mathsf{E}\left[(Y-\hat{Y})^2\right] \\
&= \mathsf{E}\left[(Y - aX - b)^2\right] \\
&= \mathsf{E}\left[Y^2\right] - 2a\mathsf{E}[YX] - 2b\mathsf{E}[Y] + a^2\mathsf{E}\left[X^2\right] + 2ab\mathsf{E}[X] + b^2 \\
\partial_a f &= -2\mathsf{E}[YX] + 2a\mathsf{E}\left[X^2\right] + 2b\mathsf{E}[X] = 0 \\
&\implies a\mathsf{E}\left[X^2\right] = \mathsf{E}[YX] - b\mathsf{E}[X] \\
\partial_b f &= -2\mathsf{E}[Y] + 2a\mathsf{E}[X] + 2b = 0 \\
&\implies b = \mathsf{E}[Y] - a\mathsf{E}[X] \\
a\mathsf{E}\left[X^2\right] &= \mathsf{E}[YX] - \mathsf{E}[X](\mathsf{E}[Y] - a\mathsf{E}[X]) \\
&\implies a(\mathsf{E}\left[X^2\right] - \mathsf{E}[X]\,\mathsf{E}[X]) = \mathsf{E}[YX] - \mathsf{E}[X]\,\mathsf{E}[Y] \\
&\implies a = \frac{\mathsf{Cov}[X,Y]}{\mathsf{V}[X]} = \rho\frac{\sigma_Y}{\sigma_X} \\
b &= \mathsf{E}[Y] - \rho\frac{\sigma_Y}{\sigma_X}\mathsf{E}[X] \\
\hat{Y} &= aX + b \\
&= \rho\frac{\sigma_Y}{\sigma_X}X + \mathsf{E}[Y] - \rho\frac{\sigma_Y}{\sigma_X}\mathsf{E}[X] \\
&= \mathsf{E}[Y] + \rho\frac{\sigma_Y}{\sigma_X}(X - \mathsf{E}[X]).
\end{aligned}$$

What a neat formula! Memorize the derivation, at least the structure. You'll come across many more optimization problems.

What if $\rho = 0$?

s.2.1.2. If we measure X in centimeters instead of meters, then X, $\mathsf{E}[X]$ and σ_X are all multiplied by 100, and the prediction $\hat{Y}$ should also be expressed in centimeters But $\mathsf{V}[Y]$ scales as length squared. This messes up the units.

s.2.1.3. Use the hint.

s.2.1.4. Let us first do one run.

Python Code

```
import numpy as np

np.random.seed(3)

n = 4
X = np.arange(n)
np.random.shuffle(X)
print(X)
print(np.arange(n))
print((X == np.arange(n)))
print((X == np.arange(n)).sum())
```

Here are the results of the print statements: $X = [3102]$. The matches are [False True False False]; we see that $X[1] = 1$ (recall, python arrays start at index 0, not at 1, so $X[1]$ is the second element of X, not the first), so that the second person picks his own hat. The number of matches is therefore 1 for this simulation.

Now put the people to work, and let them pick hats for 50 times.

Python Code

```
import numpy as np

np.random.seed(3)

num_samples = 50
n = 5

res = np.zeros(num_samples)
for i in range(num_samples):
    X = np.arange(n)
    np.random.shuffle(X)
    res[i] = (X == np.arange(n)).sum()

print(res.mean(), res.var())
```

Here is the number of matches for each round: [0. 1. 1. 0. 1. 0. 1. 0. 1. 1. 1. 2. 2. 1. 0. 1. 1. 1. 0. 2. 0. 1. 2. 2. 0. 0. 0. 1. 0. 1. 3. 1. 1. 2. 3. 0. 1. 0. 3. 1. 2. 0. 2. 0. 1. 0. 3. 0. 1. 0.] The mean and variance are as follows: $\mathsf{E}[X] = 0.96$ and $\mathsf{V}[X] = 0.8384$.

For your convenience, here's the R code

R Code

```
# set seed such that results can be recreated
set.seed(42)

# number simulations and people
numSamples <- 50
n <- 5

# initialize empty result vector
res <- c()

# for loop to simulate repeatedly
for (i in 1:numSamples) {

  # shuffle the n hats
  x <- sample(1:n)

  # number of people picking own hat (element by element the vectors x and
  # 1:n are compared, which yields a vector of TRUE and FALSE, TRUE = 1 and
  # FALSE = 0)
  correctPicks <- sum(x == 1:n)

  # append the result vector by the result of the current simulation
  res <- append(res, correctPicks)
}

# printing of observed mean and variance
print(mean(res))
print(var(res))
```

s.2.2.1. See solution manual.

s.2.2.2. Define $V := X$ and $W := X + Y$. Observe that for any t_V, t_W, we have

$$t_V V + t_W W = t_V X + t_W (X + Y) \tag{7.2.32}$$

$$= (t_V + t_W) X + t_W Y. \tag{7.2.33}$$

Hence, any linear combination of V and W is a linear combination of X and Y. Since (X, Y) is bi-variate normal, every linear combination of X and Y is normally distributed. Hence, every linear combination of V and W is normally distributed. Hence, by definition, (V, W) is bi-variate normally distributed.

We need to compute the mean vector and covariance matrix of (V, W). We have

$$\mu_V = \mathsf{E}[V] = \mathsf{E}[X] = \mu_X = 0, \tag{7.2.34}$$

and

$$\mu_W = \mathsf{E}[W] = \mathsf{E}[X + Y] = \mu_X + \mu_Y = 0. \tag{7.2.35}$$

Next, we have

$$\mathsf{V}[V] = \mathsf{V}[X] = \sigma_X^2 = 1, \tag{7.2.36}$$

and

$$\mathsf{V}[W] = \mathsf{V}[X+Y] = \mathsf{V}[X] + \mathsf{V}[Y] + 2\,\mathsf{Cov}[X,Y] \tag{7.2.37}$$
$$= 1 + 1 + 2\rho_{XY}\sigma_X\sigma_Y = 2(1+\rho_{XY}). \tag{7.2.38}$$

Finally,

$$\mathsf{Cov}[V,W] = \mathsf{Cov}[X, X+Y] = \mathsf{Cov}[X,X] + \mathsf{Cov}[X,Y] \tag{7.2.39}$$
$$= \sigma_X^2 + \rho_{XY}\sigma_X\sigma_Y = 1 + \rho_{XY}, \tag{7.2.40}$$

and hence,

$$\rho_{VW} := \mathsf{Cor}[()\,V, W) = \frac{\mathsf{Cov}[V,W]}{\sqrt{\mathsf{V}[V]\,\mathsf{V}[W]}} \tag{7.2.41}$$
$$= \frac{1+\rho_{XY}}{\sqrt{1\cdot 2(1+\rho_{XY})}} \tag{7.2.42}$$
$$= \sqrt{\frac{1+\rho_{XY}}{2}}. \tag{7.2.43}$$

We have now specified all parameters of the bi-variate normal distribution. This yields the following joint pdf:

$$f_{V,W}(v,w) = \frac{1}{2\pi\sigma_V\sigma_W\tau_{VW}}\exp\left(-\frac{1}{2\tau_{VW}^2}\left(\left(\frac{v}{\sigma_V}\right)^2 + \left(\frac{w}{\sigma_W}\right)^2 - 2\frac{\rho_{VW}}{\sigma_V\sigma_W}vw\right)\right), \tag{7.2.44}$$

where $\tau_{VW} := \sqrt{1-\rho_{VW}^2} = \sqrt{1-\frac{1+\rho_{XY}}{2}} = \sqrt{\frac{1-\rho_{XY}}{2}}$ and $\sigma_V = \sqrt{\mathsf{V}[V]} = 1$ and $\sigma_W = \sqrt{\mathsf{V}[W]} = \sqrt{2(1+\rho_{XY})}$. Hence,

$$f_{V,W}(v,w) = \frac{1}{2\pi\sqrt{1-(\rho_{XY})^2}}\exp\left(-\frac{1}{1-\rho_{XY}}\left(v^2 + \frac{w^2}{2(1+\rho_{XY})} - vw\right)\right). \tag{7.2.45}$$

s.2.2.3. Let X denote the value of the portfolio after a year in thousands of dollars. Then,

$$X := 2(1+A) + 4(1+B) + 2(1+C) + 2 \tag{7.2.46}$$
$$= 10 + 2A + 4B + 2C. \tag{7.2.47}$$

Then,

$$\mathsf{E}[X] = \mathsf{E}[10 + 2A + 4B + 2C] \tag{7.2.48}$$
$$= 10 + 2\,\mathsf{E}[A] + 4\,\mathsf{E}[B] + 2\,\mathsf{E}[C] \tag{7.2.49}$$
$$= 10 + 2\cdot 0.075 + 4\cdot 0.1 + 2\cdot 0.2 \tag{7.2.50}$$
$$= 10 + 0.15 + 0.4 + 0.4 \tag{7.2.51}$$
$$= 10.95 \tag{7.2.52}$$

s.2.2.4. We have

$$
\begin{aligned}
\mathsf{V}[X] &= \mathsf{V}[10+2A+4B+2C] && (7.2.53)\\
&= \mathsf{V}[2A]+\mathsf{V}[4B]+\mathsf{V}[2C] && (7.2.54)\\
&\quad +2\Big(\mathsf{Cov}[2A,4B]+\mathsf{Cov}[2A,2C]+\mathsf{Cov}[4B,2C]\Big) && (7.2.55)\\
&= 4\mathsf{V}[A]+16\mathsf{V}[B]+4\mathsf{V}[C] && (7.2.56)\\
&\quad +2\Big(8\,\mathsf{Cov}[A,B]+4\,\mathsf{Cov}[A,C]+8\,\mathsf{Cov}[B,C]\Big) && (7.2.57)\\
&= 4\sigma_A^2+16\sigma_B^2+4\sigma_C^2 && (7.2.58)\\
&\quad +2\Big(8\rho_{AB}\sigma_A\sigma_B+4\rho_{AC}\sigma_A\sigma_C+8\rho_{BC}\sigma_B\sigma_C\Big) && (7.2.59)\\
&= 4(0.07)^2+16(0.12)^2+4(0.17)^2 && (7.2.60)\\
&\quad +2\Big(8(0.7)(0.07)(0.12)+4(-0.8)(0.07)(0.17)+8(-0.3)(0.12)(0.17)\Big) && (7.2.61)\\
&= 0.2856. && (7.2.62)
\end{aligned}
$$

So

$$\sigma_X = \sqrt{0.2856} = 0.5344. \qquad (7.2.63)$$

So X has a standard deviation of \$534.

s.2.2.5. We need to compute the probability $\mathsf{P}\{X \le 10\}$. We have

$$
\begin{aligned}
\mathsf{P}\{X \le 10\} &= \mathsf{P}\left\{X-\mu_X \le 10-10.95\right\} && (7.2.64)\\
&= \mathsf{P}\left\{\frac{X-\mu_X}{\sigma_X} \le \frac{10-10.95}{0.5344}\right\} && (7.2.65)\\
&= \mathsf{P}\left\{Z \le \frac{10-10.95}{0.5344}\right\} && (7.2.66)\\
&= \mathsf{P}\{Z \le -1.7777\} && (7.2.67)\\
&= 0.0377. && (7.2.68)
\end{aligned}
$$

So John has a probability of 3.77% of losing money with his investment.

s.2.2.6. Observe that C has the highest expected return *and* it is negatively correlated with the other two stocks. We will use these facts to our advantage.

Starting out with portfolio X, we construct a portfolio Y by splitting the investment in stock B in two halves, which we add to our investments in stock A and C. Since the average expected return of A and C is higher than that of B, we must have that $\mathsf{E}[Y] > \mathsf{E}[X]$. Moreover, the fact that A and C are negatively correlated will mitigate the level of risk. If one stock goes up, we expect the other to go down, so the stocks cancel out each others variability. This is the idea behind the investment principle of *diversification*.

Mathematically, we define

$$
\begin{aligned}
Y &:= 4(1+A)+4(1+C)+2 && (7.2.69)\\
&= 10+4A+4C. && (7.2.70)
\end{aligned}
$$

Then,

$$\begin{aligned}
\mathsf{E}[Y] &= \mathsf{E}[10+4A+4C] && (7.2.71)\\
&= 10+4\,\mathsf{E}[A]+4\,\mathsf{E}[C] && (7.2.72)\\
&= 10+4(0.075)+4(0.20) && (7.2.73)\\
&= 11.1 && (7.2.74)
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathsf{V}[Y] &= \mathsf{V}[10+4A+4C] && (7.2.75)\\
&= \mathsf{V}[4A]+\mathsf{V}[4C]+2\,\mathsf{Cov}[4A,4C] && (7.2.76)\\
&= 4^2\,\mathsf{V}[A]+4^2\,\mathsf{V}[C]+2\cdot 4\cdot 4\cdot \mathsf{Cov}[A,C] && (7.2.77)\\
&= 16(.07)^2+16(.17)^2+32(-.8)(.07)(.17) && (7.2.78)\\
&= 0.23616, && (7.2.79)
\end{aligned}$$

which corresponds to a standard deviation of

$$\sigma_Y = \sqrt{\mathsf{V}[Y]} = \sqrt{0.23616} = 0.4860 \qquad (7.2.80)$$

So indeed, $\mathsf{E}[Y] > \mathsf{E}[X]$, while $\sigma_Y < \sigma_X$. Clearly, portfolio Y is more desirable.

s.3.1.1. Since the disturbance (x, y) has the same distribution in any direction, it has in particular the same distribution in the x and y direction. From this and property 2 we conclude that the joint PDF of the disturbance (x, y) must satisfy

$$f_{X,Y}(x,y) = f_X(x)f_Y(x) =: f(x)f(y), \qquad (7.2.81)$$

where we use property 2 first and then property 1, and we write $f(x)$ for ease. Since the disturbance has the same distribution in *any* direction, the density f can only depend on the distance r from the origin but not on the angle. Therefore, the probability that the dart lands on some square $\mathrm{d}x\,\mathrm{d}y$ must be such that

$$f(x)f(y)\,\mathrm{d}x\,\mathrm{d}y = g(r)\,\mathrm{d}x\,\mathrm{d}y, \qquad (7.2.82)$$

for some function g, hence $g(r) = f(x)f(y)$. But since g does not depend on the angle ϕ,

$$\partial_\phi g(r) = 0 = f(x)\partial_\phi f(y) + f(y)\partial_\phi f(x). \qquad (7.2.83)$$

What can we about $\partial_\phi f(x)$ and $\partial_\phi f(y)$? The relation between x and y and r and ϕ is given by the relations:

$$x = r\cos\phi, \qquad y = r\sin\phi. \qquad (7.2.84)$$

Using the chain rule,

$$\begin{aligned}
\partial_\phi f(x) &= \partial_x f(x)\frac{\mathrm{d}x}{\mathrm{d}\phi} = f'(x)r(-\sin\phi) = -f'(x)y, && (7.2.85)\\
\partial_\phi f(y) &= \partial_y f(y)\frac{\mathrm{d}y}{\mathrm{d}\phi} = f'(y)r\cos\phi = f'(y)x. && (7.2.86)
\end{aligned}$$

All this gives for (7.2.83)

$$0 = xf(x)f'(y) - yf(y)f'(x). \qquad (7.2.87)$$

Simplifying,

$$\frac{f'(x)}{xf(x)} = \frac{f'(y)}{yf(y)}. \tag{7.2.88}$$

But now notice that must hold for all x and y at the same time. The only possibility is that there is some constant α such that

$$\frac{f'(x)}{xf(x)} = \frac{f'(y)}{yf(y)} = \alpha. \tag{7.2.89}$$

Hence, our f must satisfy for all x

$$f'(x) = \alpha x f(x). \tag{7.2.90}$$

Differentiating the guess $f(x) = ae^{x^2/2\alpha}$, for some constant a, shows that this f satisfies this differential equation.

Finally, by the third property, we want that f decays as x increases, so that necessarily $\alpha < 0$. We set $\alpha = -1/\sigma^2$ to get the final answer:

$$f(x) = ae^{-x^2/2\sigma^2}. \tag{7.2.91}$$

It remains to find the normalization constant a; recall, f must be a PDF. This is the topic of the next exercise.

s.3.1.2.

$$\iint_{C(N)} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = a^2 \iint_{C(N)} e^{-(x^2+y^2)/2\sigma^2}\,\mathrm{d}x\,\mathrm{d}y. \tag{7.2.92}$$

Since $x = r\cos\phi$ and $y = r\sin\phi$, we get that $x^2 + y^2 = r^2$. For the Jacobian,

$$\frac{\partial(x,y)}{\partial(r,\phi)} = \begin{vmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{vmatrix} = r(\cos^2\phi + \sin^2\phi) = r. \tag{7.2.93}$$

Therefore

$$\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\phi, \tag{7.2.94}$$

from which

$$\begin{aligned}
a^2 \iint_{C(N)} e^{-(x^2+y^2)/2\sigma}\,\mathrm{d}x\,\mathrm{d}y &= a^2 \iint_{C(N)} e^{-r^2/2\sigma^2}\, r\,\mathrm{d}r\,\mathrm{d}\phi &(7.2.95)\\
&= a^2 \int_0^N \int_0^{2\pi} e^{-r^2/2\sigma^2}\, r\,\mathrm{d}r\,\mathrm{d}\phi &(7.2.96)\\
&= a^2 2\pi \int_0^N e^{-r^2/2\sigma^2}\, r\,\mathrm{d}r &(7.2.97)\\
&= -a^2 2\pi\sigma^2 e^{-r^2/2\sigma^2}\Big|_0^N &(7.2.98)\\
&= a^2 2\pi\sigma^2 (1 - e^{-N^2/2\sigma^2}), &(7.2.99)
\end{aligned}$$

where we use (7.2.90).

Therefore, for the square,

$$a^2 2\pi\sigma^2 (1 - e^{-N^2/2\sigma^2}) \le \iint_{S(N)} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \le a^2 2\pi\sigma^2 (1 - e^{-2N^2/2\sigma^2}). \tag{7.2.100}$$

Taking $N \to \infty$ we conclude that

$$a^2 2\pi\sigma^2 = \iint f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = a^2 \iint e^{-x^2/2\sigma^2} e^{-y^2/2\sigma}\,\mathrm{d}x\,\mathrm{d}y \tag{7.2.101}$$

$$= a^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} e^{-y^2/2\sigma}\,\mathrm{d}x\,\mathrm{d}y = a^2 \left(\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\,\mathrm{d}x\right)^2, \tag{7.2.102}$$

and therefore

$$\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\,\mathrm{d}x = \sqrt{2\pi}\sigma. \tag{7.2.103}$$

s.3.1.3. Let's first find the density $f_{X,Z}$. Let $g(x,y) = (x,z) = (x,xy)$, i.e., we take $z = xy$. It is simple to see that $y = z/x$.

We use the mnemonic

$$f_{X,Z}(x,z)\,\mathrm{d}x\,\mathrm{d}z = f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{7.2.104}$$

to see that the density $f_{X,Z}$ must be given by

$$f_{X,Z}(x,z) = f_{X,Y}(x,y)\frac{\partial(x,y)}{\partial(x,z)}. \tag{7.2.105}$$

Now (I take this form because I find it easier to differentiate in this sequence),

$$\left(\frac{\partial(x,y)}{\partial(x,z)}\right)^{-1} = \frac{\partial(x,z)}{\partial(x,y)} = \begin{vmatrix} 1 & 0 \\ y & x \end{vmatrix} = x. \tag{7.2.106}$$

and therefore, using that X and Y are iid with density f,

$$f_{X,Z}(x,z) = f_{X,Y}(x,y)\frac{1}{x} = f(x)f(y)/x = f(x)f(z/x)/x. \tag{7.2.107}$$

(Don't forget to take $1/x$ instead of x.)

It remains to tackle the domain of $f_{X,Z}$. In particular, we have to account for the fact that $X, Y \in [1,10)$. Surely, $Z \in [1,100)$, but if $Z = 80$, say, than necessarily $X > 8$. Hence, the domain is not $(x,z) \in [1,10] \times [1,100]$, but more complicated.

We already have that $1 \leq x < 10$. We also have the condition $z/x = y \in [1,10)$, which we can simplify to a condition on x,

$$1 \leq z/x < 10 \iff 1 \geq x/z > 1/10 \iff z \geq x > z/10. \tag{7.2.108}$$

Combining both constraints gives

$$\max\{1, z/10\} < x \leq \min\{10, z\}. \tag{7.2.109}$$

All in all,

$$f_{X,Z}(x,z) = f(x)f(z/x)/x\, I_{\max\{1,z/10\}<x\leq\min\{10,z\}}. \tag{7.2.110}$$

As a test, $z = 110 \implies x > 110/10 > 10$, but the indicator says that $x \leq 10$, hence we get 0 for the indicator, which is what we want in this case. (You should test $Z = 0$, $Z = 1$, $Z = 5$.)

I advice you to make a sketch of the support of X and Z.

With marginalization

$$f_Z(z) = \int_1^{10} f_{X,Z}(x,z)\,\mathrm{d}x. \tag{7.2.111}$$

We can plug in the above expression, but that just results in a longer expression that we cannot solve unless we make a specific choice for f.

Finally, if X, Y uniform on $[1, 10)$, then $f(x) = 1/9$, hence,

$$f_Z(z) = \frac{1}{9^2}\int_1^{10} I_{\max\{1,z/10\}<x\leq\min\{10,z\}} \frac{\mathrm{d}x}{x} \tag{7.2.112}$$

$$= \frac{\log(\min\{10,z\}) - \log(\max\{1,z/10\})}{81} I_{1\leq z<100}. \tag{7.2.113}$$

Do we need an indicator to ensure that $f_{X,Z}(x,z) \geq 0$ for all $z \in [1,100)$, or is this already satisfied by the expression above?

It's easy to make some interesting variations for the exam:

1. Change the domains or the distributions of X and Y.

2. Take $Z = X/Y$, or $Z = X + Y$.

s.3.2.1. We use a convolution sum. First note that the domain of X and Y is $0, 1, 2, \ldots$. For any $n = 0, 1, 2, \ldots$ we get

$$\mathsf{P}\{Z = n\} = \sum_{k=0}^{\infty} \mathsf{P}\{X = k\}\,\mathsf{P}\{Z = n \mid X = k\} \tag{7.2.114}$$

$$= \sum_{k=0}^{\infty} \mathsf{P}\{X = k\}\,\mathsf{P}\{Y = n - X \mid X = k\} \tag{7.2.115}$$

$$= \sum_{k=0}^{n} \mathsf{P}\{X = k\}\,\mathsf{P}\{Y = n - k\} \tag{7.2.116}$$

$$= \sum_{k=0}^{n} \frac{e^{-\lambda}}{k!}\lambda^k \cdot \frac{e^{-\mu}}{(n-k)!}\mu^{n-k} \tag{7.2.117}$$

$$= e^{-(\lambda+\mu)} \sum_{k=0}^{n} \frac{1}{k!(n-k)!} \cdot \lambda^k \mu^{n-k} \tag{7.2.118}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \cdot \lambda^k \mu^{n-k} \tag{7.2.119}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \cdot \lambda^k \mu^{n-k} \tag{7.2.120}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} (\lambda + \mu)^k. \tag{7.2.121}$$

We recognize this as the PMF of a Poisson distribution with parameter $\lambda + \mu$. Hence, $Z \sim \mathrm{Pois}(\lambda + \mu)$.

s.3.2.2. We use mathematical induction. For $n = 1$ we have $T_1 = X_1$, which follows an exponential distribution with rate λ. We get

$$\frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} = \frac{\lambda^1}{0!} t^0 e^{-\lambda t} \tag{7.2.122}$$

$$= \lambda e^{-\lambda t}, \tag{7.2.123}$$

for all $t > 0$. Hence, the statement is true for $n = 1$. Now suppose the statement is true for $n - 1 \geq 1$. That is, we assume that

$$f_{T_{n-1}}(t) = \frac{\lambda^{n-1}}{(n-2)!} t^{n-2} e^{-\lambda t}, \quad t > 0. \tag{7.2.124}$$

We need to prove that it follows that the statement holds for n. Note that $T_n = T_{n-1} + X_n$. Moreover, the domain of both X_n and T_{n-1} is $(0,\infty)$. This yields the convolution integral for all $t > 0$:

$$
\begin{aligned}
f_{T_n}(t) &= \int_{-\infty}^{\infty} f_{T_{n-1}}(t-x) f_{X_n}(x)\,dx && (7.2.125)\\
&= \int_0^t f_{T_{n-1}}(t-x) f_{X_n}(x)\,dx && (7.2.126)\\
&= \int_0^t \frac{\lambda^{n-1}}{(n-2)!}(t-x)^{n-2} e^{-\lambda(t-x)} \lambda e^{-\lambda x}\,dx && (7.2.127)\\
&= \frac{\lambda^n e^{-\lambda t}}{(n-2)!} \int_0^t (t-x)^{n-2}\,dx && (7.2.128)\\
&= \frac{\lambda^n e^{-\lambda t}}{(n-2)!} \left[-\frac{(t-x)^{n-1}}{n-1} \right]_{x=0}^{t} && (7.2.129)\\
&= \frac{\lambda^n e^{-\lambda t}}{(n-2)!} \left(0 - \frac{-t^{n-1}}{n-1} \right) && (7.2.130)\\
&= \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}. && (7.2.131)
\end{aligned}
$$

Hence, the statement holds for n. By mathematical induction, the statement holds for any $n = 1, 2, \ldots$.

s.3.2.3. Recall the pdf of a $\mathcal{N}(\mu,\sigma^2)$ random variable T:

$$
f_T(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(t-\mu)^2}{2\sigma^2} \right\}, \quad t \in \mathbb{R}. \tag{7.2.132}
$$

We use a convolution integral. We have for every $z \in \mathbb{R}$:

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_Y(z-x) f_X(x)\,dx && (7.2.133)\\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(z-x)^2}{2} \right\} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{x^2}{2} \right\} dx && (7.2.134)\\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ -\frac{z^2 - 2zx + 2x^2}{2} \right\} dx && (7.2.135)\\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ -\left[\frac{1}{2} z^2 - zx + x^2 \right] \right\} dx && (7.2.136)\\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ -\left[\left(x - \frac{1}{2} z \right)^2 + \frac{1}{4} z^2 \right] \right\} dx && (7.2.137)\\
&= \frac{1}{2\pi} \exp\left\{ -\frac{1}{4} z^2 \right\} \int_{-\infty}^{\infty} \exp\left\{ -\left(x - \frac{1}{2} z \right)^2 \right\} dx && (7.2.138)\\
&= \frac{1}{\sqrt{2\pi}\cdot\sqrt{2}} \exp\left\{ -\frac{z^2}{2\cdot\sqrt{2}^2} \right\} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\cdot(1/\sqrt{2})} \exp\left\{ -\frac{\left(x - \frac{1}{2}z\right)^2}{2\cdot(1/\sqrt{2})^2} \right\} dx && (7.2.139)\\
&= \frac{1}{\sqrt{2\pi}\cdot\sqrt{2}} \exp\left\{ -\frac{z^2}{2\cdot\sqrt{2}^2} \right\}, && (7.2.140)
\end{aligned}
$$

where in the last step we recognize that the integral on the right is the integral of the pdf of a $\mathcal{N}(\frac{1}{2}z, 1/2)$ random variable, which integrates to one (since any pdf integrates to one). We are left with the pdf of a $\mathcal{N}(0,2)$ random variable. Hence, $Z \sim \mathcal{N}(0,2)$.