Probability distributions EBP038A05
Lecture slides

Nicky van Foreest, Ruben van Beesten, Arpan Rijal

March 17, 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

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

---

[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 $\sim$ Exp(λ). Use the fundamental bridge to show that for $u \leq v$, the joint CDF has the form

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

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

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

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 $(X_1, \ldots, X_k)$ be Multinomial with parameters n and $(p_1, \ldots, p_k)$. Use indicator rvs to show that $\mathrm{Cov}\left[X_i, X_j\right] = -np_ip_j$ 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!*

WEEK 3

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.

WEEK 4

4

4.1 LECTURE 7

Ex 4.1.1. At the end of Story 2 of Bayes' billiards (BH.8.3.2) there is the expression

$$\beta(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}. \tag{4.1.1}$$

Derive this equation.

Ex 4.1.2. In the Beta-Binomial conjugacy story, BH take as prior $f(p) = I_{p\in[0,1]}$, and then they remark that when $f(p) \sim \beta(a,b)$ for general $a, b \in \mathbb{N}$, we must obtain the negative hypergeometric distribution. I found this pretty intriguing, so my question is: Relate the Bayes' billiards story to the story of the Negative Hypergeometric distribution, and, in passing, provide an interpretation of a and b in terms of white and black balls. Before trying to answer this question, look up the details of the negative hypergeometric distribution. (In other words, this exercise is meant to help you sort out the details of the remark of BH about the negative hypergeometric distribution.)

The next real exercise is about recursion applied to the negative hypergeometric distribution. But to get in the mood, here is short fun question on how to use recursion.

Ex 4.1.3. We have a chocolate bar consisting of n small squares. The bar is in any shape you like, square, rectangular, whatever. What is the number of times you have to break the bar such that you end up with the n single pieces?

Ex 4.1.4. Use recursion to find the expected number X of black balls drawn without replacement at random from an urn containing $w \geq 1$ white balls and b black balls before we draw 1 white ball. In other words, I ask to use recursion to compute $\mathsf{E}[X]$ for X a negative hypergeometric distribution with parameters $w, b, r = 1$ and show that

$$\mathsf{E}[X] = \frac{b}{w+1} \tag{4.1.2}$$

Ex 4.1.5. Extend the previous exercise to cope with the case $r \geq 2$. For this, write $N_r(w,b)$ for an urn with w white balls and b black balls, and r white balls to go.

4.2 LECTURE 8

Ex 4.2.1. Let X be a continuous random variable with a pdf

$$f_X(x) = \begin{cases} c, & \text{if } 0 \leq x \leq 4, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2.1}$$

1. What is the value of c?
2. What is the distribution of X?
3. Do we need to know the value of c to determine the distribution of X?

Ex 4.2.2. Let X be a continuous random variable with a pdf

$$f_X(x) = c \cdot e^{-\frac{(x-4)^2}{8}}, \quad x \in \mathbb{R}. \tag{4.2.2}$$

1. What is the value of c?
2. What is the distribution of X?
3. Do we need to know the value of c to determine the distribution of X?

(BH.8.4.5) Fred is waiting at a bus stop. He knows that buses arrive according to a Poisson process with rate λ buses per hour. Fred does not know the value of λ, though. Fred quantifies his uncertainty of λ by the *prior distribution* $\lambda \sim \text{Gamma}(r_0, b_0)$, where r_0 and b_0 are known, positive constants with r_0 an integer.

Fred has waited for t hours at the bus stop. Let Y denote the number of buses that arrive during this time interval. Suppose that Fred has observed that $Y = y$.

Ex 4.2.3. Find Fred's (hybrid) joint distribution for Y and λ.

Ex 4.2.4. Find Fred's marginal distribution for Y. Use this to compute $\mathsf{E}[Y]$. Interpret the result.

Ex 4.2.5. Find Fred's posterior distribution for λ, i.e., his conditional distribution of λ given the data y.

Ex 4.2.6. Find Fred's posterior mean $\mathsf{E}[\lambda | Y = y]$ and variance $\mathsf{V}[\lambda | Y = y]$.

---

5 WEEK 5

5.1 LECTURE 9

Ex 5.1.1. We draw, with replacement, balls, numbered 1 to N, from an urn. Find a recursion to compute the expected number $\mathsf{E}[T]$ of draws necessary to see all balls.

Ex 5.1.2. Write code to compute $\mathsf{E}[T]$ for $N = 45$.

Ex 5.1.3. We draw, with replacement, balls, numbered 1 to N, from an urn, but 6 at a time (not just one as in the previous exercise). Find a recursion to compute the expected number $\mathsf{E}[T]$ of draws necessary to see all balls.

Ex 5.1.4. For the previous exercise, compute $\mathsf{E}[T]$ for $N = 45$.

Ex 5.1.5. The lifetime X of a machine is $\text{Exp}(\lambda)$.

a. Compute $\mathsf{E}[X \mid X \leq \tau]$ where τ is some positive constant.

b. As a *check* on the result of a., use Adam's law and LOTP to show that $\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X \mid Y]] = 1/\lambda$, where $Y = I_{X \leq \tau}$.

Ex 5.1.6. We have a station with two machines, one is working, the other is off. If the first fails, the other machine takes over. The repair time of the first machine is a constant τ. If the second machine fails before the first is repaired, the station stops working, i.e., is 'down'. If the second machine does not fail before the first is repaired, the first machine takes over, and a new cycle starts (i.e., wait for the first to fail again, and so on).

Use a conditioning argument to find the expected time $\mathsf{E}[T]$ until the station is down when the lifetimes of both machines is iid $\sim \text{Exp}(\lambda)$.

5.2 LECTURE 10

We have a wooden stick of length 100 cm that we break twice. First, we break the stick at a random point that is uniformly distributed over the entire stick. We keep the left end of the stick. Then, we break the remaining stick again at a random point, uniformly distributed again, and we keep the left end again.

Ex 5.2.1. What is the expected length of the stick we end up with?

Ex 5.2.2. Now we change the story slightly. Every time we break a stick, we keep the *longest* part. What is the expected length of the remaining stick?

(Same story as last week; BH.8.4.5) Fred is waiting at a bus stop. He knows that buses arrive according to a Poisson process with rate λ buses per hour. Fred does not know the value of λ, though. Fred quantifies his uncertainty of λ by the *prior distribution* $\lambda \sim \text{Gamma}(r_0, b_0)$, where r_0 and b_0 are known, positive constants with r_0 an integer.

Fred has waited for t hours at the bus stop. Let Y denote the number of buses that arrive during this time interval. Suppose that Fred has observed that $Y = y$.

Ex 5.2.3. How many buses does Fred expect to observe? I.e., compute $\mathsf{E}[Y]$

(BH.9.3.10). An extremely widely used method for data analysis in statistics is *linear regression.* In its most basic form, the linear regression model uses a single explanatory variable X to predict a response variable Y. For instance, let X be the number of hours studied for an exam and let Y be the grade on the exam. The linear regression model assumes that the conditional expectation of Y is *linear* in X:

$$\mathsf{E}[Y|X] = a + bX. \tag{5.2.1}$$

Ex 5.2.4. Show that an equivalent way to express this is to write

$$Y = a + bX + \varepsilon, \tag{5.2.2}$$

where ε is a random variable (called the *error*) with $\mathsf{E}[\varepsilon|X] = 0$.

Ex 5.2.5. Solve for the constants a and b in terms of $\mathsf{E}[X]$, $\mathsf{E}[Y]$, $\mathsf{Cov}[X,Y]$, and $\mathsf{V}[Y]$.

6

WEEK 6

6.1 LECTURE 11

Ex 6.1.1. As a continuation of BH.9.2.4, we now ask you to use Eve's law to compute $\mathsf{V}[Y]$, the variance of the second break point.

Ex 6.1.2. We have a standard fair die and we paint red the sides with $1,2,3$ and blue the other sides. Let X be the color after a throw and Y the number. Use the definitions of conditional expectation, conditional variance and Adam's and Eve's law to compute $\mathsf{E}[Y]$ and $\mathsf{V}[Y]$. Compare that to the results you would obtain from using the standard methods; you should get the same result.

This exercises is handy to check before you solve BH.9.1.

BTW, it's easy to make some variations on this exercise by painting other combinations of sides, e.g., only 1 and 2 are red.

Ex 6.1.3. We have a continuous r.v. X with PDF

$$f(x) = \begin{cases} 1/2, & 0 \leq x < 1, \\ 1/4, & 1 \leq x \leq 3, \end{cases} \tag{6.1.1}$$

and 0 elsewhere. Compute $\mathsf{E}[X]$ and $\mathsf{V}[X]$.

Then define the r.v. Y such that $Y = I_{X<1} + 2\, I_{X\geq 1}$. Explain what information you obtain about X when somebody tells that $Y = 1$, or that $Y = 2$. Use conditioning on Y and Adam's and Eve's laws to recompute $\mathsf{E}[X]$ and $\mathsf{V}[X]$.

Again, with this exercise we can check whether we are applying all concepts in the correct way.

Ex 6.1.4. A server (e.g., a machine, a mechanic, a doctor) spends a random amount T on a job. While the server works on the job, it sometimes gets interrupted to do other tasks $\{R_i\}$. Such tasks can be failures of the machine, and they need to be repaired before the machine can continue working again. (In the case of a mechanic, interruptions occur when the mechanic has to check some other machine or help other mechanics.) It is clear that such interruptions can only occur when the server is working. Assume that the interruptions arrive according to a Poisson process with rate λ, i.e., the number of failures $N|T \sim \text{Pois}(\lambda T)$. The interruptions $\{R_i\}$ are independent of T and form an iid sequence with common mean $\mathsf{E}[R]$ and variance $\mathsf{V}[R]$. (Typically, we get estimates for $\mathsf{E}[T]$ and $\mathsf{V}[T]$ from measurements.)

Use Adam and Eve's laws to express the expectation and variance of the total time S to complete a job in terms of $\mathsf{E}[T]$ and $\mathsf{V}[R]$.

6.2 LECTURE 12

Some inequalities. Fill in either "≤" or "≥" at the location of the question mark.

Ex 6.2.1. $\mathsf{E}[XY]^2 \quad ? \quad \mathsf{E}[X^2]\,\mathsf{E}[Y^2]$.

Ex 6.2.2. $\mathsf{E}[\log(X)] \quad ? \quad \log(\mathsf{E}[X])$

Ex 6.2.3. $\mathsf{V}[Y] \quad ? \quad \mathsf{E}[\mathsf{V}[Y|X]]$

Ex 6.2.4. $\mathsf{E}[|X|] \quad ? \quad \sqrt{\mathsf{E}[X^2]}$

Ex 6.2.5. $\mathsf{P}\{X^2 \geq 4\} \quad ? \quad \mathsf{E}[|X|]/2$

Ex 6.2.6. Let $Z \sim N(0,1)$. Then, $\mathsf{P}\{Z > \sqrt{2}\} \quad ? \quad 1/e$.

We consider the height of a certain population of people (e.g., all students at the University of Groningen). For some reason, we don't know the value of the mean μ of the population, but we do know that the standard deviation σ is 10 cm. We use the sample mean $\bar{X}_n$ of an i.i.d. sample $X_1, \ldots, X_n$, from the population (measured in cm) to estimate the true mean μ. We want to choose the sample size n in such a way that our estimate $\bar{X}_n$ is sufficiently reliable.

Ex 6.2.7. One measure of reliability of our estimator $\bar{X}_n$ is its standard deviation. Let's say we find our estimator reliable if its standard deviation is at most 1 cm. Give a lower bound for n for which we can guarantee that our esitmator is reliable in this sense.

Ex 6.2.8. Another measure of reliability of our estimator is given by the probability that our estimate is very bad. Specifically, we say our estimate is reliable if we can be 99% sure that our estimate is off by less than 5 cm. Give a lower bound for n for which we can guarantee that our estimator is reliable in this sense.

Little Mike needs to get ready for school. He must leave within 15 minutes, but there are two more things he needs to do: eat his breakfast and get dressed. The time it takes Mike to eat his breakfast has a mean value of 6 minutes with a standard deviation of 3 minutes. The time it takes Mike to get dressed has a mean value of 4 minutes with a standard deviation of 1 minute.

Ex 6.2.9. Give a lower bound for the probability that Mike will be ready for school in time.

7

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.

h.4.1.4. For ease, write $N(w,b) = \mathsf{E}[X]$ for an urn with $w \geq r = 1$ white balls and b black balls. Then explain that

$$N(w,0) = 0 \quad \text{for all } w, \tag{7.1.2}$$

$$N(w,b) = \frac{b}{w+b}(1 + N(w,b-1)). \tag{7.1.3}$$

Then show that this implies that $N(w,b) = b/(w+1)$.

h.4.1.5. Explain that $N_0(w,b) = 0$ and

$$N_r(w,b) = \frac{w}{w+b} N_{r-1}(w-1,b) + \frac{b}{w+b}(1 + N_r(w,b-1)). \tag{7.1.4}$$

Then show that this implies that $N_r(w,b) = rb/(w+1)$.

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 $\mathsf{P}\{X = x, Y = y\} = \mathsf{P}\{X = x\}\,\mathsf{P}\{Y = y\}$. 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

$$\mathsf{V}[Z] = \mathsf{E}[Z^2] - (\mathsf{E}[Z])^2 \tag{7.2.12}$$
$$\mathsf{E}[Z^2] = \mathsf{E}[(X+Y)^2] = \mathsf{E}[X^2] + 2\,\mathsf{E}[XY] + \mathsf{E}[Y^2] \tag{7.2.13}$$
$$(EZ)^2 = (\mathsf{E}[X] + \mathsf{E}[Y])^2 \tag{7.2.14}$$
$$= (\mathsf{E}[X])^2 + 2\,\mathsf{E}[X]\,\mathsf{E}[Y] + (\mathsf{E}[Y])^2 \tag{7.2.15}$$
$$\implies \tag{7.2.16}$$
$$\mathsf{V}[Z] = \mathsf{E}[Z^2] - (\mathsf{E}[Z])^2 \tag{7.2.17}$$
$$= \mathsf{V}[X] + \mathsf{V}[Y] + 2(\mathsf{E}[XY] - (\mathsf{E}[X]\,\mathsf{E}[Y])) \tag{7.2.18}$$
$$\mathsf{E}[XY] = \sum_{ij} i\,j\,p_{X,Y}(i, j) = \frac{1}{4}(1 + 4 + 9 + 16) = \ldots \tag{7.2.19}$$
$$\mathsf{E}[X^2] = \ldots \tag{7.2.20}$$

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:

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

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,

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

$\mathsf{P}\{L > i\}$ 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

$$\begin{aligned} t_V V + t_W W &= t_V X + t_W(X+Y) && (7.2.32)\\ &= (t_V + t_W)X + t_W Y. && (7.2.33)\end{aligned}$$

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)

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

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

$$a^2 \iint_{C(N)} e^{-(x^2+y^2)/2\sigma}\,\mathrm{d}x\,\mathrm{d}y = a^2 \iint_{C(N)} e^{-r^2/2\sigma^2}\,r\,\mathrm{d}r\,\mathrm{d}\phi \tag{7.2.95}$$

$$= a^2 \int_0^N \int_0^{2\pi} e^{-r^2/2\sigma^2}\,r\,\mathrm{d}r\,\mathrm{d}\phi \tag{7.2.96}$$

$$= a^2 2\pi \int_0^N e^{-r^2/2\sigma^2}\,r\,\mathrm{d}r \tag{7.2.97}$$

$$= -a^2 2\pi\sigma^2 e^{-r^2/2\sigma^2}\Big|_0^N \tag{7.2.98}$$

$$= a^2 2\pi\sigma^2(1 - e^{-N^2/2\sigma^2}), \tag{7.2.99}$$

where we use (7.2.90).

Therefore, for the square,

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

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, $Z \sim \text{Pois}(\lambda + \mu)$.

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, $Z \sim \mathcal{N}(0,2)$.

s.4.1.1. We have by the Bayes' billiard stories that

$$\binom{n}{k}\int_0^1 p^k(1-p)^{n-k}\,\mathrm{d}p = \frac{1}{n+1}, \tag{7.2.141}$$

but the integral is equal to $\beta(k+1, n-k+1)$ by the definition of the β function. Hence,

$$\beta(k+1, n-k+1) = \frac{1}{(n+1)\binom{n}{k}} = \frac{(n-k)!k!}{(n+1)!}.$$

Taking $a-1=k$, $b-1=n-k$, then $n=b-1+k=a+b-2$, so that

$$\beta(a,b) = \frac{(b-1)!(a-1)!}{(a+b-1)!}.$$

s.4.1.2. Let's follow the Bayes' billiards story no 1, but in the context of the negative hypergeometric distribution. We have w white balls, and b black balls. We place, arbitrarily all balls on a line, and mark the rth, $r \leq w$, white ball. Call this ball the pivot. Then we ask: what is the probability that the number of black balls X lying at the left of the pivot is equal to k, i.e, $\mathsf{P}\{X=k\}$?

Conditional on the position of the pivot being p, we must have that

$$\mathsf{P}\{X=k\} = w\int_0^1 \binom{b}{k} p^k(1-p)^{b-k} \times \binom{w-1}{r-1} p^{r-1}(1-p)^{w-r}\,\mathrm{d}p. \tag{7.2.142}$$

To see this, reason as follows. There are w ways to choose one of the white balls as the pivot. (I initially forgot this factor.) Then, given p, k black balls lie to the left of the pivot; the rest lies to the right. Then, of the $w-1$ remaining white balls, $r-1$ also have to lie the left of the pivot, and the others to the right (observe that $w-1-(r-1) = w-r$, hence $(1-p)^{w-r}$.)

With the definition of the β function, we can simplify to this:

$$\begin{aligned}
\mathsf{P}\{X=k\} &= w\binom{b}{k}\binom{w-1}{r-1}\int_0^1 p^{k+r-1}(1-p)^{b-k+w-r}\,\mathrm{d}p && (7.2.143)\\
&= w\binom{b}{k}\binom{w-1}{r-1}\beta(k+r, b-k+w-r+1) && (7.2.144)\\
&= w\binom{b}{k}\binom{w-1}{r-1}\frac{(r+k-1)!(b-k+w-r)!}{(w+b)!}. && (7.2.145)
\end{aligned}$$

Now, from the story of the negative hypergeometric distribution, we have that

$$\mathsf{P}\{X=k\} = \frac{\binom{w}{r-1}\binom{b}{k}}{\binom{w+b}{r+k-1}} \frac{w-(r-1)}{w-(r-1)+b-k} \tag{7.2.146}$$

$$\tag{7.2.147}$$

To see this, note that when $X=k$, the $r+k$th ball is white (hence we stop). The probability to select a white ball then $w-(r-1)$ out of the remaining $w-(r-1)+b-k$. (For X to be equal to k we have removed k black balls and $r-1$ white balls). And we have to multiply by the number of combinations to select $r-1$ of the w balls and k of the black balls.

Next, with a some algebra—here we do the work, but the books just say 'after a bit of algebra we find', and then it is left to you to do the algebra—,

$$\mathsf{P}\{X=k\} = \frac{\binom{w}{r-1}\binom{b}{k}}{\binom{w+b}{r+k-1}} \frac{w-r+1}{w+b-r-k+1} \tag{7.2.148}$$

$$= \binom{w}{r-1}\binom{b}{k}\frac{(r+k-1)!(w+b-r-k+1)!}{(w+b)!}\frac{w-r+1}{w+b-r-k+1} \tag{7.2.149}$$

$$= \binom{w}{r-1}\binom{b}{k}\frac{(r+k-1)!(w+b-r-k)!}{(w+b)!}(w-r+1) \tag{7.2.150}$$

$$= w\binom{w-1}{r-1}\binom{b}{k}\frac{(r+k-1)!(w+b-r-k)!}{(w+b)!}, \tag{7.2.151}$$

where we use that

$$w\binom{w-1}{r-1} = w\frac{(w-1)!}{(w-r)!(r-1)!} = \frac{w!}{(w-r)!(r-1)!} \tag{7.2.152}$$

$$= \frac{w!}{(w-r+1)!(r-1)!}(w-r+1) = \binom{w}{r-1}(w-r+1). \tag{7.2.153}$$

The results for the Bayes' billiards match with those of the negative hypergeometric distribution!

So, inspecting (7.2.142) we see that the a and b parameters in the β distribution for the prior $f(p)$ are such that $a=r$ and $b=w-r+1$. (Don't confuse the b here with the number of black balls. I took the b to stick to the notation of the book.) In other words, the a corresponds to the white ball we mark as the pivot, and b to the remaining number of white balls (plus 1).

s.4.1.3. Write $T(n)$ for the number of times you have to break the bar when there are n squares. Clearly, $T(1)=0$. In general, suppose $n=m+k$, $1\le k<n$, then $T(n)=T(m)+T(k)+1$, because we need $T(k)$ to break the part with k pieces, and $T(m)$ for the part with m pieces, and we need 1 to break the big bar into the two parts. But then, with the *boundary condition* $T(1)=0$,

$$T(2) = T(1)+T(1)+1 = 1,$$
$$T(3) = T(1)+T(2)+1 = 1+1 = 2,$$

and so on. So we guess that $T(n)=n-1$ for any n. Let's check. It certainly holds for $n=1$. Generally,

$$n-1 = T(n) = T(k)+T(m)+1 = k-1+m-1+1 = k+m-1 = n-1.$$

s.4.1.4. If there are no black balls left, then we cannot pick a black ball which implies that $N(w,0)=0$. As soon as we pick a white ball, we have to stop.

When $w,b\ge 1$ two things can happen. Suppose we pick a white ball, then we stop right away and we cannot increase the number of black balls picked. Suppose we pick a black ball, which happens with probability $b/(b+w)$, we obtain one black ball, and we continue with one black ball less. Hence, with LOTE,

$$N(w,b) = \frac{b}{w+b}(1+N(w,b-1)). \tag{7.2.154}$$

You should realize that here we use conditioning; we condition on the color of the ball picked.

Now we need a real tiny bit of inspiration. Let's *guess* that $N(w,b) = \alpha b$ for some $\alpha > 0$. Guessing is always allowed; if it works, we are done (since the recursion, together with the boundary condition, has a unique solution: just repeatedly applying it allows us to calculate all values). In general, you can make any guess whatsoever, but mind that only the good guesses work. So, trying the form $N(w,b) = \alpha b$, we find that

$$\alpha b = \frac{b}{w+b}(1+\alpha(b-1)) \overset{\text{algebra}}{\Longrightarrow} \alpha = \frac{1}{w+1} \implies N(w,b) = \frac{b}{w+1}.$$

This is consistent with the boundary condition $N(w,0) = 0$. Hence, we can move the case $b = 0$ to the case with $b = 1$, and so on, thereby proving the validity of the formula in general.

s.4.1.5. The recursion follows right away by noticing that we pick a white ball with probability $w/(w+b)$, but then we remove one white ball *and* we have one white ball less to go. With probability $b/(w+b)$ we pick a black ball, in which case the number of black balls drawn increases by one, but there is one black ball less in the urn.

The boundary condition is also clear: $N_0(w,b) = 0$ because we stop. Note that this is consistent with the previous exercise.

If $r = 2$, we use from the previous exercise that

$$N_{r-1}(w-1,b) = N_1(w-1,b) = \frac{b}{w-1+1} = \frac{b}{w}.$$

Using this in the recursion,

$$\begin{aligned}
N_2(w,b) &= \frac{w}{w+b}N_1(w-1,b) + \frac{b}{w+b}(1+N_2(w,b-1)) \\
&= \frac{w}{w+b}\frac{b}{w} + \frac{b}{w+b}(1+N_2(w,b-1)) \\
&= \frac{b}{w+b}(2+N_2(w,b-1)).
\end{aligned}$$

But this has the same form as (4.1.2), except that the 1 has been replaced by a 2. But, now we are dealing with the case $r = 2$ instead of $r = 1$. Hence, by the same token,

$$N_2(w,b) = 2\frac{b}{w+1}$$

Knowing the result for $r = 2$, we fill this for $r = 3$, and so on, to get the general result.

What an elegant procedure; we pull ourselves out of the swamp, just like Baron Munchhausen.

s.4.2.1. We need that the pdf f_X integrates to one. Hence, we need

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1 \iff \tag{7.2.155}$$

$$\int_0^4 c\,dx = 1 \iff \tag{7.2.156}$$

$$4c = 1 \iff \tag{7.2.157}$$

$$c = 1/4. \tag{7.2.158}$$

Clearly, X is uniformly distributed on $[0,4]$. In fact, we do not need to know the value of c to determine this. It is sufficient to know that the pdf of X is constant on the interval $[0,4]$.

s.4.2.2. We need that the pdf f_X integrates to one. Hence, we need

$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1 \iff \tag{7.2.159}$$

$$\int_{-\infty}^{\infty} c\cdot e^{-\frac{(x-4)^2}{8}}\,dx = 1 \iff \tag{7.2.160}$$

$$c\cdot\sqrt{2\pi}\cdot 2\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\cdot 2}e^{-\frac{1}{2}\frac{(x-4)^2}{2^2}}\,dx = 1 \iff \tag{7.2.161}$$

$$c\cdot\sqrt{2\pi}\cdot 2\cdot 1 = 1 \iff \tag{7.2.162}$$

$$c = \frac{1}{\sqrt{2\pi}\cdot 2}. \tag{7.2.163}$$

Clearly, X is $N(4,2)$ distributed. In fact, we do not need to know the value of c to determine this. It is sufficient to observe the structure of the pdf as a function of x.

s.4.2.3. See the book.

s.4.2.4. See the book for the solution to the first part. From the result, it follows that $\mathsf{E}[Y] = \frac{r_0}{b_0}t$ (just plug in the parameters of the Negative Binomial distribution into the expression for the expected value on page 161).
Note that the expected number of buses is linear in time t and has a rate of r_0/b_0 per hour, which is the expected value of the rate λ of our Poisson process.

s.4.2.5. See the book.

s.4.2.6. See the book.

s.5.1.1. Write T_n for expected time to finish given that we have seen n different balls. Thus, $\mathsf{E}[T] = T_0$. Then for $n < N$ and using conditioning and LOTE

$$T_n = 1 + T_{n+1}\,\mathsf{P}\{\text{draw a new ball}\} + T_n\,\mathsf{P}\{\text{draw an old ball}\} \tag{7.2.164}$$

$$= 1 + T_{n+1}\frac{N-n}{N} + T_n\frac{n}{N} \tag{7.2.165}$$

$$\implies \tag{7.2.166}$$

$$T_n\frac{N-n}{N} = 1 + T_{n+1}\frac{N-n}{N} \tag{7.2.167}$$

$$\implies \tag{7.2.168}$$

$$T_n = \frac{N}{N-n} + T_{n+1}. \tag{7.2.169}$$

How about the boundary conditions, i.e., what is T_n for $n = N$? That's obvious: $T_N = 0$. Hence,

$$T_{N-1} = \frac{N}{N-(N-1)} + T_N = N, \tag{7.2.170}$$

and so on.

If you are interested, push the result a bit further: approximate T_0 for $N \gg 0$.

s.5.1.2. Here is the python code.

Python Code

```python
N = 45


def T(n):
    if n >= N:
        return 0
    return N / (N - n) + T(n + 1)

# print(T(44))
# print(T(0))
```

R Code

```r
N = 45

bigT = function(n) {
  if(n >= N) {
    return(0)
  }
  return(N / (N - n) + bigT(n + 1))
}

#print(bigT(44))
#print(bigT(0))
```

Here are the results $T(44) = 45.0$ (check!) and $T(0) = 197.77266519980952$.

s.5.1.3. Write T_n for expected time to finish given that we have seen n different balls. Take 6 balls. If we would know the number k of new balls drawn, then $T_n = 1 + T_{n+k}$. What is the probability to draw k new balls out of the 6 we pick? This must be

$$\binom{n}{6-k}\binom{N-n}{k}\Big/\binom{N}{6}. \tag{7.2.171}$$

Therefore, by LOTE, and when $N - n \geq 6$

$$T_n = 1 + \sum_{k=0}^{6} \frac{\binom{n}{6-k}\binom{N-n}{k}}{\binom{N}{6}} T_{n+k}. \tag{7.2.172}$$

This formula is not ok when are just 2 new balls as in that case we cannot pick $k=6$ new balls. In general, we can pick $k=\min\{6,N-n\}$ new balls. Hence,

$$T_n = 1 + \sum_{k=0}^{\min\{6,N-n\}} \frac{\binom{n}{6-k}\binom{N-n}{k}}{\binom{N}{6}} T_{n+k} \tag{7.2.173}$$

$$\implies \tag{7.2.174}$$

$$T_n - \frac{\binom{n}{6}\binom{N-n}{0}}{\binom{N}{6}} T_n = 1 + \sum_{k=1}^{\min\{6,N-n\}} \frac{\binom{n}{6-k}\binom{N-n}{k}}{\binom{N}{6}} T_{n+k} \tag{7.2.175}$$

$$\implies \tag{7.2.176}$$

$$T_n\left(\binom{N}{6} - \binom{n}{6}\right) = \binom{N}{6} + \sum_{k=1}^{\min\{6,N-n\}} \binom{n}{6-k}\binom{N-n}{k} T_{n+k} \tag{7.2.177}$$

$$\tag{7.2.178}$$

And this is the final result.

s.5.1.4. When we draw $m=6$ balls at a time, we should take care how we compute the recursion.

To see this, consider for ease the case in which we draw two balls at at time. Now I want to know how often each function gets called in the computation. For this, I write a simple helper function $S(n,N)$ that returns 1 for being called plus $S(n+1,N)+S(n+2,N)$. Like this, I can find out how often the functions are called in the recursion.

Python Code

```python
def S(n, N):
    if n >= N:
        return 0
    return 1 + S(n + 1, N) + S(n + 2, N)


for N in range(1, 25):
    print(S(0, N))
```

R Code

```r
S = function(n, N) {
  if(n >= N) {
    return(0)
  }
  return(1 + S(n + 1, N) + S(n + 2, N))
}

for (N in 1:24) {
  print(S(0, N))
}
```

The result is this: 1 2 4 7 12 20 33 54 88 143 232 376 609 986 1596 2583 4180 6764 10945 17710 28656 46367 75024 121392 . Apparently, the number of functions called grows very rapidly. When we want to compute for 6 balls and $N=45$ the situation must be much, much worse.

The way out is to *store* intermediate results rather than computing them time and again. For this we use the concept *memoization*. You should memorize this concept when dealing with the computation of recursions. In short, this means that we check whether we computed the value of a function earlier. If so, get the value from memory, otherwise, do the computation, and store it for later purposes.

Let's check what happens if we print out how often the function gets called when using memoization.

Python Code

```python
from functools import lru_cache

@lru_cache   # This realizes the memoization.
def S(n, N):
    if n >= N:
        return 0
    print(f"Called for n = {n}\n")
    return 1 + S(n + 1, N) + S(n + 2, N)


S(0, 5)
```

R Code

```r
library("memoise")

S = function(n, N) {
  if(n >= N) {
    return(0)
  }
  cat("Called for n =", n, "\n")
  return(1 + S(n + 1, N) + S(n + 2, N))
}
S = memoise(S)

S(0, 5)
```

Now the result is this:

Called for n = 0
Called for n = 1
Called for n = 2
Called for n = 3
Called for n = 4

This is much better, for each n the function in the recursion gets called just once.

To convince you that memoization really works, let's remove the memoization.

Python Code

```python
def S(n, N):
    if n >= N:
```

```
        return 0
    print(f"Called for n = {n}\n")
    return 1 + S(n + 1, N) + S(n + 2, N)


S(0, 5)
```

R Code

```
S = function(n, N) {
  if(n >= N) {
    return(0)
  }
  cat("Called for n =", n, "\n")
  return(1 + S(n + 1, N) + S(n + 2, N))
}

S(0, 5)
```

This is the result:

Called for n = 0
Called for n = 1
Called for n = 2
Called for n = 3
Called for n = 4
Called for n = 4
Called for n = 3
Called for n = 4
Called for n = 2
Called for n = 3
Called for n = 4
Called for n = 4

You see how often the function is called with $n = 4$?

For the course you don't have to look up on the web how memoization is implemented, but I advice you to read about. Understanding such ideas makes you much better at programming, which is important in view of the fact that many of you will have to use computers a lot in your profession careers. Besides this, memoization lies at the heart of how spreadsheet programs such as excel work.

Now we know that we have to use memoization, we can turn to the balls and urn problem. I build a general function $T(n, m)$ for the situation in which we have seen n balls and we pick m balls at a time. Like this, I can check with the earlier case in which we picked just one ball by computing $T(n, 1)$. (Hopefully you learn from this that you should also test your code.)

Python Code

```
from math import comb
from functools import lru_cache

```

```python
N = 45


@lru_cache
def T(n, m):
    if n >= N:
        return 0
    res = comb(N, m)
    for k in range(1, min(m, N - n) + 1):
        P = comb(n, m - k)
        P *= comb(N - n, k)
        res += P * T(n + k, m)
    return res / (comb(N, m) - comb(n, m))
```

R Code

```r
library("memoise")

N = 45

bigT = function(n, m) {
  if(n >= N) {
    return(0)
  }
  res = choose(N, m)
  for (k in 1:min(m, N - n)) {
    P = choose(n, m - k)
    P = P * choose(N - n, k)
    res = res + P * bigT(n + k, m)
  }
  return(res / (choose(N, m) - choose(n, m)))
}
bigT = memoise(bigT)

bigT(0, 1)
bigT(0, 6)
```

Here is the check: $T(0,1) = 197.7726651998094$, which is the same as our earlier computation.

Next, $T(0,6) = 31.497085595869386$. This is more than 6 as fast, i.e., $6 \times T(0,6) = 188.98251357521633 < T(0,1)$. Why is that? Well, the 6 balls we pick from the urn are guaranteed to be drawn without replacement.

Can you solve the following extension? Suppose for each time you draw a set of balls, you get the value (in Euro's say) of the numbers on the balls, and then you put the balls back in the urn. The game stops when you have seen all balls at least once. What is your expected gain?

s.5.1.5. Since $X \sim \text{Exp}(\lambda)$,

$$\mathsf{P}\{X \leq t | X \leq \tau\} = \frac{\mathsf{P}\{X \leq \min\{t,\tau\}\}}{\mathsf{P}\{X \leq \tau\}} = \frac{1 - e^{-\lambda \min\{t,\tau\}}}{1 - e^{-\lambda\tau}}.$$

Defining

$$f(t) = \partial_t \mathsf{P}\{X \leq t | X \leq \tau\} = \lambda \frac{e^{-\lambda t}}{1 - e^{-\lambda\tau}} I_{0 \leq t \leq \tau},$$

we get

$$\mathsf{E}[X|X \leq \tau] = \int t f(t)\,\mathrm{d}t \tag{7.2.179}$$
$$= \frac{1}{1-e^{-\lambda\tau}} \int_0^\tau \lambda t e^{-\lambda t}\,\mathrm{d}t. \tag{7.2.180}$$

Simplifying,

$$\int_0^\tau \lambda t e^{-\lambda t}\,\mathrm{d}t = -t e^{-\lambda t}\Big|_0^\tau + \int_0^\tau e^{-\lambda t}\,\mathrm{d}t \tag{7.2.181}$$
$$= -\tau e^{-\lambda\tau} + \frac{e^{-\lambda\tau}-1}{-\lambda} \tag{7.2.182}$$
$$= \frac{1-e^{-\lambda\tau}-\tau\lambda e^{-\lambda\tau}}{\lambda}, \tag{7.2.183}$$

$$\mathsf{E}[X|X \leq \tau] = \frac{1-e^{-\lambda\tau}-\tau\lambda e^{-\lambda\tau}}{\lambda(1-e^{-\lambda\tau})}. \tag{7.2.184}$$

For part 2: by memorylessness, $\mathsf{E}[X \mid X > \tau] = \tau + 1/\lambda$. Hence,

$$\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X \mid Y]], \quad \text{Adam's law} \tag{7.2.185}$$
$$= \mathsf{E}[X \mid Y=1]\,\mathsf{P}\{Y=1\} + \mathsf{E}[X \mid Y=0]\,\mathsf{P}\{Y=0\}, \quad \text{LOTE} \tag{7.2.186}$$
$$= \mathsf{E}[X \mid X \leq \tau]\,\mathsf{P}\{X \leq \tau\} + \mathsf{E}[X \mid X > \tau]\,\mathsf{P}\{X > \tau\} \tag{7.2.187}$$
$$= \frac{1-e^{-\lambda\tau}-\tau\lambda e^{-\lambda\tau}}{\lambda(1-e^{-\lambda\tau})}(1-e^{-\lambda\tau}) + (\tau+1/\lambda)e^{-\lambda\tau} \tag{7.2.188}$$
$$= 1/\lambda - e^{-\lambda\tau}/\lambda - \tau e^{-\lambda\tau} + \tau e^{-\lambda T} + e^{-\lambda\tau}/\lambda \tag{7.2.189}$$
$$= 1/\lambda. \tag{7.2.190}$$

Here is a final point for you to think hard about: $X \leq \tau$ is an event, $Y = I_{X\leq\tau}$ is a r.v. Therefore the following is sloppy notation: $\mathsf{E}[\mathsf{E}[X \mid X \leq \tau]]$, and when using, it is easy to get confused, because where is the random variable on which we should condition according to Adam's law?

s.5.1.6. Let X_i be the lifetime of machine i. Suppose that the second machine does not break down before τ, then

$$\mathsf{E}[T|X_2 > \tau] = \mathsf{E}[X_1 + \tau + C|X_2 > \tau] = 1/\lambda + \tau + \mathsf{E}[T], \tag{7.2.191}$$

because the expected time until machine 1 fails is $1/\lambda$, then we wait until it is repaired, and then, by memoryless, a new cycle C starts, which has an expected duration of $\mathsf{E}[T]$.

Next condition on the second machine breaking down before τ. Then, since X_1 and X_2 are independent and by the previous exercise,

$$\mathsf{E}[T|X_2 \leq \tau] = \mathsf{E}[X_1 + X_2|X_2 \leq \tau] = \frac{1}{\lambda} + \frac{1-e^{-\lambda\tau}-\tau\lambda e^{-\lambda\tau}}{\lambda(1-e^{-\lambda\tau})}. \tag{7.2.192}$$

By LOTE and using $\mathsf{P}\{X_2 \geq \tau\} = e^{-\lambda\tau}$,

$$\mathsf{E}[T] = \mathsf{E}[T|X_2 \leq \tau]\,\mathsf{P}\{X_2 \leq \tau\} + \mathsf{E}[T|X_2 \geq \tau]\,\mathsf{P}\{X_2 \geq \tau\} \tag{7.2.193}$$

$$= (1-e^{-\lambda\tau})\left(\frac{1}{\lambda} + \frac{1-e^{-\lambda\tau} - \tau\lambda e^{-\lambda\tau}}{\lambda(1-e^{-\lambda\tau})}\right) + e^{-\lambda\tau}\left(\frac{1}{\lambda} + \tau + \mathsf{E}[T]\right) \tag{7.2.194}$$

$$= \frac{1}{\lambda} + \frac{1-e^{-\lambda\tau} - \tau\lambda e^{-\lambda\tau}}{\lambda} + e^{-\lambda\tau}\tau + e^{-\lambda\tau}\,\mathsf{E}[T] \tag{7.2.195}$$

$$\implies \tag{7.2.196}$$

$$\left(1-e^{-\lambda\tau}\right)\mathsf{E}[T] = \frac{2-e^{-\lambda\tau}}{\lambda} \tag{7.2.197}$$

$$\implies \tag{7.2.198}$$

$$\mathsf{E}[T] = \frac{2-e^{-\lambda\tau}}{\lambda(1-e^{-\lambda\tau})}. \tag{7.2.199}$$

For you to do: Check the limits $\tau = 0$ and $\tau = \infty$. Is the result in accordance with your intuition?

Two challenges: Does the problem become much harder when X_1 and X_2 are still exponential, but have different failure rates λ_1 and λ_2? Can you extend the analysis to 3 machines, or n machines?

s.5.2.1. Let $X \sim \text{Unif}(0, 100)$ denote the point where the first stick is broken (in cm). Let S denote the point where the second stick is broken (i.e., the length of the remaining stick). Then, $S \sim \text{Unif}(0, X)$. We want to compute $\mathsf{E}[Y]$. Using Adam's law we obtain

$$\mathsf{E}[S] = \mathsf{E}[\mathsf{E}[S|X]] \tag{7.2.200}$$

$$= \mathsf{E}[X/2] \tag{7.2.201}$$

$$= \mathsf{E}[X]/2 \tag{7.2.202}$$

$$= 100/4 = 25. \tag{7.2.203}$$

Hence, we end up with a stick of expected length 25 cm.

s.5.2.2. By symmetry we can assume that the point of the first break is $Y \sim \text{Unif}(1/2, 1)$, in meters. By the same reasoning, the point of the second break, conditional on $Y = y$, is $Z|Y = y \sim \text{Unif}(y/2, y)$. Then, $\mathsf{E}\left[Z|Y = y\right] = 3y/4$. Replacing y by the rv Y, we have that $\mathsf{E}[Z|Y] = 3Y/4$. With this, and Adam's law,

$$\mathsf{E}[Z] = \mathsf{E}[\mathsf{E}[Z|Y]] = \mathsf{E}[3Y/4] = 3/4 \cdot \mathsf{E}[Y] = 3/4 \cdot (1/2+1)/2 = (3/4)^2, \tag{7.2.204}$$

in meters.

s.5.2.3. Last week, we solved this question by first deriving the marginal distribution of Y (which was negative binomial) and then computing the mean of this distribution. We obtained $\mathsf{E}[Y] = \frac{r_0}{b_0}t$. Now, let's use conditional expectations instead. By Adam's law,

$$\mathsf{E}[Y] = \mathsf{E}[\mathsf{E}[Y|\lambda]]. \tag{7.2.205}$$

Since $Y|\lambda \sim \text{Pois}(\lambda t)$, we have

$$\mathsf{E}[\mathsf{E}[Y|\lambda]] = \lambda t. \tag{7.2.206}$$

Hence,

$$\begin{aligned}
\mathsf{E}[Y] &= \mathsf{E}[\mathsf{E}[Y|\lambda]] && (7.2.207)\\
&= \mathsf{E}[\lambda t] && (7.2.208)\\
&= \mathsf{E}[\lambda]\, t && (7.2.209)\\
&= \frac{r_0}{b_0} t. && (7.2.210)
\end{aligned}$$

s.5.2.4. See the book.

s.5.2.5. See the book.

s.6.1.1. Recall that $\mathsf{V}[X] = 1/12$. More generally, if the stick has length l, then $\mathsf{V}[X] = l^2/12$. With this idea and using the solution of BH.9.2.4,

$$\begin{aligned}
\mathsf{E}[Y\,|\,X=x] &= x/2, && (7.2.211)\\
\mathsf{V}[Y\,|\,X=x] &= \mathsf{E}\left[Y^2\,\middle|\,X=x\right] - (\mathsf{E}[Y\,|\,X=x])^2 && (7.2.212)\\
&= \frac{1}{x}\int_0^x y^2\,\mathrm{d}y - \left(\frac{1}{x}\int_0^x y\,\mathrm{d}y\right)^2 = x^3/3x - (x^2/2x)^2 = x^2/12 && (7.2.213)
\end{aligned}$$

Replacing x by X:

$$\begin{aligned}
\mathsf{E}[Y\,|\,X] &= X/2, && (7.2.214)\\
\mathsf{V}[Y\,|\,X] &= X^2/12, && (7.2.215)
\end{aligned}$$

Adam's law:

$$\mathsf{E}[Y\,|\,X] = X/2 \implies \mathsf{E}[Y] = \mathsf{E}[\mathsf{E}[Y\,|\,X]] = \mathsf{E}[X/2] = 1/4. \quad (7.2.216)$$

Eve's law:

$$\begin{aligned}
&\mathsf{V}[Y] = \mathsf{E}[\mathsf{V}[Y\,|\,X]] + \mathsf{V}[\mathsf{E}[Y\,|\,X]] && (7.2.217)\\
&\mathsf{V}[Y\,|\,X] = X^2/12 \implies \mathsf{E}[\mathsf{V}[Y\,|\,X]] = \mathsf{E}\left[X^2\right]/12 = \frac{1}{12}\int_0^1 x^2\,\mathrm{d}x = \frac{1}{12\cdot 3} && (7.2.218)\\
&\mathsf{E}[Y\,|\,X] = X/2 \implies \mathsf{V}[\mathsf{E}[Y\,|\,X]] = \mathsf{V}[X/2] = \frac{1}{4}\frac{1}{12}, && (7.2.219)\\
&\mathsf{V}[Y] = 1/12\cdot(1/3+1/4) = 7/144. && (7.2.220)
\end{aligned}$$

s.6.1.2. The first step is easy:

$$\begin{aligned}
\mathsf{E}[Y\,|\,X=r] &= \sum_y y\,\mathsf{P}\left\{Y=y\,\middle|\,X=r\right\} = \sum_y y\frac{\mathsf{P}\{Y=y, X=r\}}{\mathsf{P}\{X=r\}} = \frac{(1+2+3)/6}{1/2} = 2, && (7.2.221)\\
\mathsf{E}[Y\,|\,X=b] &= \frac{(4+5+6)/6}{1/2} = 5. && (7.2.222)
\end{aligned}$$

Recall that in the stick breaking exercise, we first computed $\mathsf{E}[Y\,|\,X=x] = x/2$, and then we replaced x by X to get $\mathsf{E}[Y\,|\,X] = X/2$. In the present case (with the die), where is the x at the RHS? In other words, how to turn the above into a function $g(x)$ so that we can replace x by X to get a rv?

Here is the solution. Define g as

$$g(x) = \begin{cases} \mathsf{E}[Y\,|\,X=r], & \text{if } x=r\\ \mathsf{E}[Y\,|\,X=b], & \text{if } x=b. \end{cases} \quad (7.2.223)$$

But this is equal to

$$g(x) = \mathsf{E}[Y \mid X = r]\ I_{x=r} + \mathsf{E}[Y \mid X = b]\ I_{x=b}. \tag{7.2.224}$$

With this insight:

$$\begin{aligned}
\mathsf{E}[Y \mid x] = g(x) = \mathsf{E}[Y \mid X = r]\ I_{x=r} + \mathsf{E}[Y \mid X = b]\ I_{x=b}, \\
\mathsf{E}[Y \mid X] = g(X) = \mathsf{E}[Y \mid X = r]\ I_{X=r} + \mathsf{E}[Y \mid X = b]\ I_{X=b}.
\end{aligned}$$

Here, the indicators are the rvs!

And now,

$$\begin{aligned}
\mathsf{E}[Y] &= \mathsf{E}[\mathsf{E}[Y \mid X]] = \mathsf{E}\left[g(X)\right] && (7.2.225)\\
&= \mathsf{E}[\mathsf{E}[Y \mid X = r]\ I_{X=r} + \mathsf{E}[Y \mid X = b]\ I_{X=b}] && (7.2.226)\\
&= \mathsf{E}[\mathsf{E}[Y \mid X = r]\ I_{X=r}] + \mathsf{E}[\mathsf{E}[Y \mid X = b]\ I_{X=b}] && (7.2.227)\\
&= \mathsf{E}[Y \mid X = r]\ \mathsf{E}[I_{X=r}] + \mathsf{E}[Y \mid X = b]\ \mathsf{E}[I_{X=b}] && (7.2.228)\\
&= \mathsf{E}[Y \mid X = r]\ \mathsf{P}\{X = r\} + \mathsf{E}[Y \mid X = b]\ \mathsf{P}\{X = b\} && (7.2.229)\\
&= 2\frac{1}{2} + 5\frac{1}{2}. && (7.2.230)
\end{aligned}$$

In view of this, let's consider LOTE again. Observing that

$$\begin{aligned}
A_1 &= \{s \in S : \text{paint of } s \text{ is red}\} = \{1,2,3\} && (7.2.231)\\
A_2 &= \{s \in S : \text{paint of } s \text{ is blue}\} = \{4,5,6\}, && (7.2.232)
\end{aligned}$$

we have

$$\begin{aligned}
\mathsf{E}[Y \mid X = r] &= \mathsf{E}[Y \mid A_1], && (7.2.233)\\
\mathsf{E}[Y \mid X = b] &= \mathsf{E}[Y \mid A_2] && (7.2.234)\\
g(x) &= \mathsf{E}[Y \mid X = r]\ I_{x=r} + \mathsf{E}[Y \mid X = b]\ I_{x=b} && (7.2.235)\\
&= \mathsf{E}[Y \mid A_1]\ I_{x\in A_1} + \mathsf{E}[Y \mid A_2]\ I_{x=A_2} && (7.2.236)\\
&= \sum_{i=1,2} \mathsf{E}[Y \mid A_i]\ I_{x\in A_i} && (7.2.237)
\end{aligned}$$

The last is simple to generalize!

$$\begin{aligned}
g(x) &= \sum_i \mathsf{E}[Y \mid A_i]\ I_{x\in A_i} && (7.2.238)\\
g(X) &= \sum_i \mathsf{E}[Y \mid A_i]\ I_{X\in A_i}. && (7.2.239)
\end{aligned}$$

And now we define $\mathsf{E}[Y \mid X] = g(X)$. Remember these general equations!

Now observe that Adam's law and LOTE are the same for discrete rvs:

$$\begin{aligned}
\mathsf{E}[Y] &= \mathsf{E}[\mathsf{E}[Y \mid X]]\\
&= \mathsf{E}\left[g(X)\right]\\
&= \mathsf{E}\left[\sum_i \mathsf{E}[Y \mid A_i]\ I_{X\in A_i}\right]\\
&= \sum_i \mathsf{E}[Y \mid A_i]\ \mathsf{E}\left[I_{X\in A_i}\right]\\
&= \sum_i \mathsf{E}[Y \mid A_i]\ \mathsf{P}\{A_i\}.
\end{aligned}$$

s.6.1.3. Using the properties of the uniform distribution,

$$\mathsf{E}[X] = \frac{1}{2}\int_0^1 x\,\mathrm{d}x + \frac{1}{4}\int_1^3 x\,\mathrm{d}x = \frac{1}{4} + \frac{9-1}{8} = \frac{5}{4} \tag{7.2.240}$$
$$\mathsf{E}\left[X^2\right] = \frac{1}{2}\int_0^1 x^2\,\mathrm{d}x + \frac{1}{4}\int_1^3 x^2\,\mathrm{d}x = \frac{1}{6} + \frac{27-1}{12} = \frac{7}{3} \tag{7.2.241}$$
$$\mathsf{V}[X] = \frac{7}{3} - \frac{25}{16} = \frac{37}{48}. \tag{7.2.242}$$

$Y = 1 \implies 0 \le X < 1$, $Y = 2 \implies 2 \le X \le 4$.

$$\mathsf{E}[X\,|\,Y=1] = 1/2, \tag{7.2.243}$$
$$\mathsf{E}[X\,|\,Y=2] = 2, \tag{7.2.244}$$
$$g(y) = \mathsf{E}[X\,|\,Y=1]\,I_{y=1} + \mathsf{E}[X\,|\,Y=2]\,I_{y=2} \tag{7.2.245}$$
$$g(Y) = \mathsf{E}[X\,|\,Y] = \frac{1}{2}I_{Y=1} + 2I_{Y=2}, \tag{7.2.246}$$
$$\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X\,|\,Y]] = \mathsf{E}\left[g(Y)\right] \tag{7.2.247}$$
$$= \mathsf{E}\left[\frac{1}{2}I_{Y=1} + 2I_{Y=2}\right] \tag{7.2.248}$$
$$= \frac{1}{2}\mathsf{E}[I_{Y=1}] + 2\mathsf{E}[I_{Y=2}] \tag{7.2.249}$$
$$= \frac{1}{2}\mathsf{P}\{Y=1\} + 2\mathsf{P}\{Y=2\} \tag{7.2.250}$$
$$\mathsf{P}\{Y=1\} = \mathsf{E}[I_{Y=1}] = \mathsf{E}[I_{X<1}] = \frac{1}{2} \tag{7.2.251}$$
$$\mathsf{P}\{Y=2\} = \mathsf{E}[I_{Y=2}] = \mathsf{E}[I_{X\ge 1}] = \frac{1}{2} \tag{7.2.252}$$
$$\mathsf{E}[X] = \frac{1}{2}\frac{1}{2} + 2\frac{1}{2} = \frac{5}{4}. \tag{7.2.253}$$

$$\mathsf{V}[X\,|\,Y] = \mathsf{V}[X\,|\,Y=1]\,I_{Y=1} + \mathsf{V}[X\,|\,Y=2]\,I_{Y=2} \tag{7.2.254}$$
$$= \frac{1}{12}I_{Y=1} + \frac{4}{12}I_{Y=2}, \tag{7.2.255}$$
$$\mathsf{E}[\mathsf{V}[X\,|\,Y]] = \frac{1}{12}\frac{1}{2} + \frac{4}{12}\frac{1}{2} = \frac{5}{24}, \tag{7.2.256}$$
$$\mathsf{V}[\mathsf{E}[X\,|\,Y]] = \frac{1}{4}\mathsf{V}[I_{Y=1}] + 4\mathsf{V}[I_{Y=2}] + 2\,\mathsf{Cov}\left[\frac{1}{2}I_{Y=1}, 2I_{Y=2}\right] \tag{7.2.257}$$
$$\mathsf{V}[I_{Y=1}] = \mathsf{V}[I_{Y=2}] = \frac{1}{4}, \tag{7.2.258}$$
$$2\,\mathsf{Cov}\left[\frac{1}{2}I_{Y=1}, 2I_{Y=2}\right] = 2\,\mathsf{Cov}[I_{Y=1}, I_{Y=2}] = -2\mathsf{E}[I_{Y=1}]\,\mathsf{E}[I_{Y=2}] = -\frac{1}{2} \tag{7.2.259}$$
$$\mathsf{V}[\mathsf{E}[X\,|\,Y]] = \frac{1}{16} + 1 - \frac{1}{2} = \frac{9}{16} \tag{7.2.260}$$
$$\mathsf{V}[X] = \mathsf{E}[\mathsf{V}[X\,|\,Y]] + \mathsf{V}[\mathsf{E}[X\,|\,Y]] = \frac{5}{24} + \frac{9}{16}. \tag{7.2.261}$$

I forgot the covariance at first :-(

s.6.1.4. Below we need some properties of the Poisson distribution.

$$\mathsf{E}[N(T)] = \mathsf{E}[\mathsf{E}[N(T) \,|\, T]] \tag{7.2.262}$$

$$\mathsf{E}[\mathsf{E}[N(T) \,|\, T = t]] = \mathsf{E}[N(t)] = \lambda t \tag{7.2.263}$$

$$\mathsf{E}[\mathsf{E}[N(T) \,|\, T]] = \lambda T \tag{7.2.264}$$

$$\mathsf{E}[N(T)] = \mathsf{E}[\mathsf{E}[N(T) \,|\, T]] = \mathsf{E}[\lambda T] = \lambda \mathsf{E}[T] \tag{7.2.265}$$

$$\mathsf{V}[N(t)] = \lambda t \tag{7.2.266}$$

The duration S of a job is its own service time T plus all interruptions, hence $S = T + \sum_{i=1}^{N(T)} R_i$, hence

$$\mathsf{E}[S] = \mathsf{E}[\mathsf{E}[S \,|\, T]] = \mathsf{E}\left[\mathsf{E}\left[T + \sum_{i=1}^{N(T)} R_i \,\middle|\, T\right]\right] \tag{7.2.267}$$

$$\mathsf{E}[S \,|\, T = t] = \mathsf{E}\left[T + \sum_{i=1}^{N(T)} R_i \,\middle|\, T = t\right] = \mathsf{E}\left[t + \sum_{i=1}^{N(t)} R_i\right] = t + \mathsf{E}\left[\sum_{i=1}^{N(t)} R_i\right] \tag{7.2.268}$$

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i\right] = \mathsf{E}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] \tag{7.2.269}$$

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t) = n\right] = \mathsf{E}\left[\sum_{i=1}^{n} R_i\right] = n\mathsf{E}[R] \implies \mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right] = N(t)\mathsf{E}[R] \tag{7.2.270}$$

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i\right] = \mathsf{E}[N(t)\mathsf{E}[R]] = \mathsf{E}[R]\,\mathsf{E}[N(t)] = \lambda t \mathsf{E}[R] \tag{7.2.271}$$

$$\mathsf{E}[S \,|\, T = t] = t + \lambda \mathsf{E}[R]\, t = (1 + \lambda \mathsf{E}[R])\, t \tag{7.2.272}$$

$$\mathsf{E}[S \,|\, T] = (1 + \lambda \mathsf{E}[R]) T \tag{7.2.273}$$

$$\mathsf{E}[S] = \mathsf{E}[(1 + \lambda \mathsf{E}[R]) T] = (1 + \lambda \mathsf{E}[R])\, \mathsf{E}[T]. \tag{7.2.274}$$

(When rereading this exercise, explain per step which assumption we use to get to the next step.)

For the variance,

$$\mathsf{V}[S] = \mathsf{E}[\mathsf{V}[S \,|\, T]] + \mathsf{V}[\mathsf{E}[S \,|\, T]]. \tag{7.2.275}$$

From the above we immediately have

$$\mathsf{V}[\mathsf{E}[S \,|\, T]] = \mathsf{V}[(1 + \lambda \mathsf{E}[R]) T] = (1 + \lambda \mathsf{E}[R])^2 \mathsf{V}[T]. \tag{7.2.276}$$

For $\mathsf{E}[\mathsf{V}[S \,|\, T]]$ we go step by step.

$$\mathsf{V}[S \,|\, T = t] = \mathsf{V}\left[T + \sum_{i=1}^{N(T)} R_i \,\middle|\, T = t\right] = \mathsf{V}\left[\sum_{i=1}^{N(t)} R_i\right], \tag{7.2.277}$$

as $\mathsf{V}[t] = 0$. But $N(t)$ is a Poisson rv. So we need to use Eve's law again, but now condition on $N(t)$.

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i\right] = \mathsf{E}\left[\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] + \mathsf{V}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right]. \tag{7.2.278}$$

For the middle part:

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)=n\right] = \mathsf{V}\left[\sum_{i=1}^{n} R_i\right] = n\,\mathsf{V}[R] \tag{7.2.279}$$

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right] = N(t)\,\mathsf{V}[R] \tag{7.2.280}$$

$$\mathsf{E}\left[\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] = \mathsf{E}[N(t)\,\mathsf{V}[R]] = \lambda t\,\mathsf{V}[R]. \tag{7.2.281}$$

For the right part:

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)=n\right] = n\,\mathsf{E}[R] \tag{7.2.282}$$

$$\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right] = N(t)\,\mathsf{E}[R] \tag{7.2.283}$$

$$\mathsf{V}\left[\mathsf{E}\left[\sum_{i=1}^{N(t)} R_i \,\middle|\, N(t)\right]\right] = \mathsf{V}[N(t)\,\mathsf{E}[R]] = (\mathsf{E}[R])^2\,\mathsf{V}[N(t)] = (\mathsf{E}[R])^2\lambda t. \tag{7.2.284}$$

And so,

$$\mathsf{V}\left[\sum_{i=1}^{N(t)} R_i\right] = \lambda t\,\mathsf{V}[R] + \lambda t(\mathsf{E}[R])^2 = \lambda t\,\mathsf{E}\left[R^2\right]. \tag{7.2.285}$$

Putting the things together:

$$\mathsf{V}[S\,|\,T=t] = \mathsf{V}\left[\sum_{i=1}^{N(t)} R_i\right] = \lambda t\,\mathsf{E}\left[R^2\right] \tag{7.2.286}$$

$$\mathsf{V}[S\,|\,T] = \lambda T\,\mathsf{E}\left[R^2\right] \tag{7.2.287}$$

$$\mathsf{E}[\mathsf{V}[S\,|\,T]] = \lambda\,\mathsf{E}\left[R^2\right]\mathsf{E}[T]. \tag{7.2.288}$$

So, finally,

$$\mathsf{V}[S] = \lambda\,\mathsf{E}\left[R^2\right]\mathsf{E}[T] + (1+\lambda\,\mathsf{E}[R])^2\,\mathsf{V}[T]. \tag{7.2.289}$$

Special cases for you to sort out:

1. R and T constant. Why is still $\mathsf{V}[S] > 0$?
2. Take R and T exponential; then the expressions for $\mathsf{V}[T]$ etc are still easy to manage.

s.6.2.1. By the Cauchy-Schwartz inequality,

$$\mathsf{E}[XY] \leq \sqrt{\mathsf{E}\left[X^2\right]\mathsf{E}\left[Y^2\right]}. \tag{7.2.290}$$

Squaring both sides yields the solution

$$\mathsf{E}[XY]^2 \leq \mathsf{E}\left[X^2\right]\mathsf{E}\left[Y^2\right]. \tag{7.2.291}$$

s.6.2.2. By Jensen's inequality, $\mathsf{E}\left[\log X\right] \leq \log \mathsf{E}[X]$, since $\log(\cdot)$ is a concave function.

s.6.2.3. By Eve's law we have

$$\mathsf{V}[Y] = \mathsf{E}[\mathsf{V}[Y|X]] + \mathsf{V}[\mathsf{E}[Y|X]]. \tag{7.2.292}$$

Since $\mathsf{V}[\mathsf{E}[Y|X]] \geq 0$, it follows that

$$\mathsf{V}[Y] \geq \mathsf{E}[\mathsf{V}[Y|X]]. \tag{7.2.293}$$

s.6.2.4. Note that $|X| = \sqrt{X^2}$. Define $Y = X^2$ and note that $g(y) = \sqrt{y}$ is a concave function. Hence, by Jensen's inequality,

$$\mathsf{E}[|X|] = \mathsf{E}\left[g(Y)\right] \leq g(\mathsf{E}[Y]) = \sqrt{\mathsf{E}\left[X^2\right]}. \tag{7.2.294}$$

Hence, the solution is $\mathsf{E}[|X|] \leq \sqrt{\mathsf{E}\left[X^2\right]}$

s.6.2.5. We have

$$\mathsf{P}\left\{X^2 \geq 4\right\} = \mathsf{P}\{|X| \geq 2\} \leq \mathsf{E}[|X|]/2, \tag{7.2.295}$$

by Markov's inequality.

s.6.2.6. By Chernoff's inequality,

$$\begin{aligned}
\mathsf{P}\left\{Z > \sqrt{2}\right\} &\leq \frac{\mathsf{E}\left[e^{tZ}\right]}{e^{\sqrt{2}t}} && (7.2.296)\\
&= \frac{e^{\frac{1}{2}t^2}}{e^{\sqrt{2}t}} && (7.2.297)\\
&= e^{\frac{1}{2}t^2 - \sqrt{2}t}, && (7.2.298)
\end{aligned}$$

for every $t > 0$. This inequality is tightest for $t = \sqrt{2}$, as this minimizes $\frac{1}{2}t^2 - \sqrt{2}t$. Plugging in this value yields

$$\begin{aligned}
\mathsf{P}\left\{Z > \sqrt{2}\right\} &\leq e^{\frac{1}{2}\sqrt{2}^2 - \sqrt{2}\cdot\sqrt{2}} && (7.2.299)\\
&\leq e^{1-2} = 1/e. && (7.2.300)
\end{aligned}$$

s.6.2.7. We want to guarante that $sd(\bar{X}_n) \leq 1$, which is equivalent to $\mathsf{V}\left[\bar{X}_n\right] \leq 1$. We have

$$\begin{aligned}
\mathsf{V}\left[\bar{X}_n\right] &= \mathsf{V}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] && (7.2.301)\\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{V}[X_i] && (7.2.302)\\
&= \frac{1}{n^2}\cdot n \cdot \sigma^2 && (7.2.303)\\
&= \frac{100}{n}. && (7.2.304)
\end{aligned}$$

So to guarantee we have a reliable estimator, we need $\frac{100}{n} \leq 1$, which is equivalent to $n \geq 100$.

s.6.2.8. We need to guarantee that

$$\mathsf{P}\left\{|\bar{X}_n - \mu| \leq 5\right\} \geq 0.99, \tag{7.2.305}$$

which is equivalent to

$$\mathsf{P}\left\{|\bar{X}_n - \mu| > 5\right\} < 0.01, \tag{7.2.306}$$

Note that $\bar{X}_n$ is a random variable with mean μ and variance $\sigma^2/n = 100/n$. Using Chebyshev's inequality, we obtain

$$\mathsf{P}\left\{|\bar{X}_n - \mu| > 5\right\} < \frac{100/n}{5^2} = 4/n. \tag{7.2.307}$$

Equating the right-hand side to 0.01 yields $n = 400$. So we need $n \geq 400$.

s.6.2.9. Let X and Y denote the time it takes Mike to eat his breakfast and get dressed, respectively. Let $Z = X + Y$ denote the total time Mike needs to finish his morning routine. Let μ denote the mean of Z, i.e., $\mu = \mathsf{E}[Z] = \mathsf{E}[X + Y] = 6 + 4 = 10$. Then,

$$\mathsf{P}\{Z < 15\} = 1 - \mathsf{P}\{Z \geq 15\}, \tag{7.2.308}$$

where, using Chebyshev's inequality,

$$\begin{aligned}
\mathsf{P}\{Z \geq 15\} &= \mathsf{P}\left\{Z - \mu \geq 5\right\} && (7.2.309)\\
&\leq \mathsf{P}\left\{|Z - \mu| \geq 5\right\} && (7.2.310)\\
&\leq \frac{\mathsf{V}[Z]}{5^2} && (7.2.311)\\
&= \frac{\mathsf{V}[X] + \mathsf{V}[Y]}{25} && (7.2.312)\\
&= \frac{9+1}{25} && (7.2.313)\\
&= 10/25 = 0.4. && (7.2.314)
\end{aligned}$$

Hence,

$$\mathsf{P}\{Z < 15\} \geq 1 - 0.4 = 0.6. \tag{7.2.315}$$

So the probabilty that Mike will be ready in time is at least 60%.