Probability distributions
Lectures, questions and solutions

Nicky van Foreest and Ruben van Beesten

March 3, 2021

CONTENTS

1 LECTURE 1

Ex 1.1. Consider 12 football players on a football field. Eleven of them are players of FC Barcelone, the other one is an arbiter. We select a random player, uniform. This player must take a penalty. The probability that a player of Barcelone 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 $P\{P = B, S = 1\}$ and so on for all possibilities.

2. What is $P\{S = 1\}$? What is $P\{P = B\}$?

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

Ex 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 1.4. What is c?

Ex 1.5. What is $F_X(i)$? What is $F_Y(j)$?

Ex 1.6. Are X and Y dependent? If so, why, because $1 = F_{X,Y}(4,4) = F_X(4)F_Y(4)$?

Ex 1.7. What is $P\{Z = k\}$?

Ex 1.8. What is $V[Z]$?

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

Ex 1.9. What is $P\{Z = 4\}$?

Remark 1.1. 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?

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

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., $\mathsf{P}\{X+Y>5-2\}$?

HERE IS SOME extra material for you practice on 2D integration.

Ex 1.10. We have a continuous r.v. $X \geq 0$ with finite expectation. Use 2D integration and indicators to prove that

$$\mathsf{E}[X] = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0 G(x)\,\mathrm{d}x, \tag{1.1}$$

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

Ex 1.11. Explain that for a continuous r.v. X with CDF F and a and b (so it might be that $a > b$),

$$\mathsf{P}\{a < X < b\} = [F(b) - F(a)]^+. \tag{1.2}$$

INDICATORS ARE GREAT functions, and I suspect you underestimated the importance of these functions. They help to keep your formulas clean. You can use them in computer code as logical conditions, or to help counting relevant events, something you need when numerically estimating multi-D integrals for machine learning for instance. And, even though I(=NvF) often prefer to use figures over algebra to understand something, when it comes to integration (and reversing the sequence of integration in multiple integrals) I find indicators easier to use.

Moreover, you should know that in fact, *expectation* is the fundamental concept in probability theory, and the *probability of an event is defined* as

$$\mathsf{P}\{A\} := \mathsf{E}[I_A]. \tag{1.3}$$

Thus, the fundamental bridge is actually an application of LOTUS to indicator functions. REREAD BH.4.4!

Ex 1.12. What is $\int_{-\infty}^\infty I_{0 \leq x \leq 3}\,\mathrm{d}x$?

Ex 1.13. What is

$$\int x I_{0 \leq x \leq 4}\,\mathrm{d}x? \tag{1.4}$$

When we do an integral over a 2D surface we can first integrate over the x and then over the y, or the other way around, whatever is the most convenient. (There are conditions about how to handle multi-D integral, but for this course these are irrelevant.)

Ex 1.14. What is

$$\iint xy I_{0 \leq x \leq 3} I_{0 \leq y \leq 4}\,\mathrm{d}x\,\mathrm{d}y? \tag{1.5}$$

Ex 1.15. What is

$$\iint I_{0 \leq x \leq 3} I_{0 \leq y \leq 4} I_{x \leq y}\,\mathrm{d}x\,\mathrm{d}y? \tag{1.6}$$

Ex 1.16. Take $X \sim \text{Unif}([1,3]), Y \sim \text{Unif}([2,4])$ and independent. Compute

$$\mathsf{P}\{Y \leq 2X\}. \tag{1.7}$$

2 LECTURE 2

Read the problems of `memoryless\_excursions.pdf`. All the problems in that document relate to topics discussed in Sections BH.7.1 and BH.7.2, and quite a lot of topics you have seen in the previous course on probability theory.

3   LECTURE 3

Ex 3.1. 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 3.2. 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 centimetres. 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 3.3. 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? In BH.7.46 you have to solve this analytically. In the exercise here you have to write a simulator for compute the expectation and variance.

4   LECTURE 4

Ex 4.1. BH.7.65 Let $(X_1,\ldots,X_k)$ be Multinomial with parameters n and $(p_1,\ldots,p_k)$. Use indicator r.v.s to show that $\mathrm{Cov}(X_i,X_j)=-np_ip_j$ for $i\neq j$.

Ex 4.2. Suppose (X,Y) are bivariate 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 tradeoff 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 multinomial 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 4.3. 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 4.4. What is the standard deviation of the value of the portfolio in a year?

Ex 4.5. 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 4.6. 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!*

5   LECTURE 5

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

Ex 5.1. 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 5.2. 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,

$$\iint_{C(N)} f_{X,Y}(x,y) \, dx \, dy \leq \iint_{S(N)} f_{X,Y}(x,y) \, dx \, dy \leq \iint_{C(\sqrt{2}N)} f_{X,Y}(x,y) \, dx \, dy. \tag{5.1}$$

Now substitute the normal distribution of [5.1]. 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 5.3. 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 $X, Y \sim \text{Unif}([1, 10))$, what is f_Z?

6   LECTURE 6

Ex 6.1. Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. What is the distribution of $Z = X + Y$?

Ex 6.2. (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:

$$f_{T_n}(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}, \quad t > 0. \tag{6.1}$$

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 6.3. 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.

7   LECTURE 7

Ex 7.1. At the end of Story 2 of Bayes' billiards (BH.8.3.2) there is the expression

$$\beta(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}. \tag{7.1}$$

Derive this equation.

Ex 7.2. 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 7.3. 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 7.4. 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

$$\mathsf{E}[X] = \frac{b}{w+1} \tag{7.2}$$

Ex 7.5. 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.

8   LECTURE 8

Ex 8.1. Let X be a continuous random variable with a pdf

$$f_X(x) = \begin{cases} c, & \text{if } 0 \le x \le 4, \\ 0, & \text{otherwise.} \end{cases} \tag{8.1}$$

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 8.2. Let X be a continuous random variable with a pdf

$$f_X(x) = c \cdot e^{-\frac{(x-4)^2}{8}}, \quad x \in \mathbb{R}. \tag{8.2}$$

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

Ex 8.4. Find Fred's marginal distribution for Y. Use this to compute $\mathsf{E}[Y]$. Interpret the result.

Ex 8.5. Find Fred's posterior distribution for λ, i.e., his conditional distribution of λ given the data y.

Ex 8.6. Find Fred's posterior mean $\mathsf{E}[\lambda|Y = y]$ and variance $\mathsf{V}[\lambda|Y = y]$.

9   LECTURE 9

Ex 9.1. The lifetime X of a machine is $\text{Exp}(\lambda)$. Compute $\mathsf{E}[X\,|\,X \leq \tau]$ where τ is some positive constant.

   Define $Y = I_{X \leq \tau}$. Use Adam's law and LOTP to show that $\mathsf{E}[X] = \mathsf{E}[\,\mathsf{E}[X\,|\,Y]\,] = 1/\lambda$. Observe that this is just a check on our results.

Ex 9.2. 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'. 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 $\sim \text{Exp}(\lambda)$.

Ex 9.3. 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 9.4. Write code to compute $\mathsf{E}[T]$ for $N = 45$.

Ex 9.5. 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 9.6. For the previous exercise, compute $\mathsf{E}[T]$ for $N = 45$.

10   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 10.1. What is the expected length of the stick we end up with?

Ex 10.2. 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 10.3. 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:

$$\mathsf{E}[Y|X] = a + bX. \tag{10.1}$$

Ex 10.4. Show that an equivalent way to express this is to write

$$Y = a + bX + \varepsilon, \tag{10.2}$$

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

Ex 10.5. Solve for the constants a and b in terms of $\mathsf{E}[X]$, $\mathsf{E}[Y]$, $\mathsf{Cov}[X,Y]$, and $\mathsf{V}[Y]$.

11   HINTS

h.1.10. Check the proof of BH.4.4.8

h.1.11. Recall that $F \in [0,1]$.

h.6.1. Use the Binomial theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \tag{11.1}$$

for any nonnegative integer n.

h.6.2. Use mathematical induction.

h.7.4. 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

$$N(w,0) = 0 \quad \text{for all } w, \tag{11.2}$$

$$N(w,b) = \frac{b}{w+b}(1 + N(w, b-1)). \tag{11.3}$$

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

h.7.5. Explain that $N_0(w,b) = 0$ and

$$N_r(w,b) = \frac{w}{w+b} N_{r-1}(w-1,b) + \frac{b}{w+b}(1 + N_r(w, b-1)). \tag{11.4}$$

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

12   SOLUTIONS

s.1.1. Here is the joint PMF:

$$P\{P = A, S = 1\} = \frac{1}{12}0.5 \qquad\qquad P\{P = A, S = 0\} = \frac{1}{12}0.5 \qquad (12.1)$$

$$P\{P = B, S = 1\} = \frac{11}{12}0.7 \qquad\qquad P\{P = B, S = 0\} = \frac{11}{12}0.3. \qquad (12.2)$$

Now the marginal PMFs

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

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

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

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

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

s.1.2. When the claim sizes are 0, then the insurance company does not receive a claim.

s.1.3. By the fundamental bridge,

$$P\{Z = k\} = \sum_{i,j} I_{i+j=k}\, p_{X,Y}(i,j) \qquad (12.3)$$

$$= \sum_{i,j} I_{i,j \geq 0}\, I_{j=k-i}\, p_{X,Y}(i,j) \qquad (12.4)$$

$$= \sum_{i=0}^{k} p_{X,Y}(i, k - i). \qquad (12.5)$$

s.1.4. $c = 1/4$ because there are just four possible values for i and j.

s.1.5. Use marginalization:

$$F_X(k) = F_{X,Y}(k, \infty) = \sum_{i \leq k} \sum_{j} p_{X,Y}(i,j) \qquad (12.6)$$

$$= \frac{1}{4} \sum_{i \leq k} \sum_{j} I_{i=j}\, I_{1 \leq i \leq 4} \qquad (12.7)$$

$$= \frac{1}{4} \sum_{i \leq k} I_{1 \leq i \leq 4} \qquad (12.8)$$

$$= k/4, \qquad (12.9)$$

$$F_Y(j) = j/4. \qquad (12.10)$$

s.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 $F_{X,Y}(1,1) \neq F_X(1)F_Y(1)$:

$$\frac{1}{4} = F_{X,Y}(1,1) \neq F_X(1)F_Y(1) = \frac{1}{4}\frac{1}{4}. \qquad (12.11)$$

s.1.7. $P\{Z = 2\} = P\{X = 1, Y = 1\} = 1/4 = P\{Z = 4\}$, etc. $P\{Z = k\} = 0$ for $k \notin \{2, 4, 6, 8\}$.

s.1.8. Here is one approach

$$V[Z] = E\left[Z^2\right] - (E[Z])^2 \tag{12.12}$$

$$E\left[Z^2\right] = E\left[(X + Y)^2\right] = E\left[X^2\right] + 2E[XY] + E\left[Y^2\right] \tag{12.13}$$

$$(EZ)^2 = (E[X] + E[Y])^2 \tag{12.14}$$

$$= (E[X])^2 + 2E[X]E[Y] + (E[Y])^2 \tag{12.15}$$

$$\Longrightarrow \tag{12.16}$$

$$V[Z] = E\left[Z^2\right] - (E[Z])^2 \tag{12.17}$$

$$= V[X] + V[Y] + 2(E[XY] - (E[X]E[Y])) \tag{12.18}$$

$$E[XY] = \sum_{ij} ij p_{X,Y}(i, j) = \frac{1}{4}(1 + 4 + 9 + 16) = \dots \tag{12.19}$$

$$E\left[X^2\right] = \dots \tag{12.20}$$

The numbers are for you to compute.

s.1.9.

$$P\{Z = 4\} = \sum_{i,j} I_{i+j=4} p_{X,Y}(i, j) \tag{12.21}$$

$$= \sum_{i=1}^{4} \sum_{j=1}^{4} I_{j=4-i} \frac{1}{16} \tag{12.22}$$

$$= \sum_{i=1}^{3} \frac{1}{16} \tag{12.23}$$

$$= \frac{3}{16}. \tag{12.24}$$

s.1.10. The trick is to realize that $x = \int_0^\infty I_{y \le x}\, dy$. Using this,

$$E[X] = \int_0^\infty x f(x)\, dx \tag{12.25}$$

$$= \int_0^\infty \int_0^\infty I_{y \le x} f(x)\, dy\, dx \tag{12.26}$$

$$= \int_0^\infty \int_0^\infty I_{y \le x} f(x)\, dx\, dy \tag{12.27}$$

$$= \int_0^\infty \int_0^\infty I_{x \ge y} f(x)\, dx\, dy \tag{12.28}$$

$$= \int_0^\infty \int_y^\infty f(x)\, dx\, dy \tag{12.29}$$

$$= \int_0^\infty G(y)\, dy. \tag{12.30}$$

s.1.11.

$$a < b \implies P\{a < X < b\} = F(b) - F(a) = [F(b) - F(a)]^+ \tag{12.31}$$

$$a \ge b \implies P\{a < X < b\} = 0 = [F(b) - F(a)]^+, \tag{12.32}$$

where the last equality follows from the fact that F is increasing.

s.1.12.

$$\int_{-\infty}^{\infty} I_{0\leq x\leq 3}\,\mathrm{d}x = \int_0^3 \mathrm{d}x = 3.$$

s.1.13.

$$\int x\, I_{0\leq x\leq 4}\,\mathrm{d}x = \int_0^4 x\,\mathrm{d}x = 16/2 = 8.$$

s.1.14.

$$\iint xy\, I_{0\leq x\leq 3}\, I_{0\leq y\leq 4}\,\mathrm{d}x\,\mathrm{d}y = \int_0^3 x \int_0^4 y\,\mathrm{d}y\,\mathrm{d}x$$
$$= \int_0^3 x \frac{y^2}{2}\Big|_0^4\,\mathrm{d}x$$
$$= \int_0^3 x\cdot 8\,\mathrm{d}x = 8\cdot 9/2 = 4\cdot 9.$$

s.1.15. Two solutions. First we integrate over y.

$$\iint I_{0\leq x\leq 3}\, I_{0\leq y\leq 4}\, I_{x\leq y}\,\mathrm{d}x\,\mathrm{d}y = \int I_{0\leq x\leq 3}\int I_{0\leq y\leq 4}\, I_{x\leq y}\,\mathrm{d}y\,\mathrm{d}x \tag{12.33}$$
$$= \int I_{0\leq x\leq 3}\int I_{\max\{x,0\}\leq y\leq 4}\,\mathrm{d}y\,\mathrm{d}x \tag{12.34}$$
$$= \int_0^3 \int_{\max\{x,0\}}^4 \mathrm{d}y\,\mathrm{d}x \tag{12.35}$$
$$= \int_0^3 y\Big|_{\max\{x,0\}}^4\,\mathrm{d}x \tag{12.36}$$
$$= \int_0^3 (4-\max\{x,0\})\,\mathrm{d}x \tag{12.37}$$
$$= 12 - \int_0^3 \max\{x,0\}\,\mathrm{d}x \tag{12.38}$$
$$= 12 - \int_0^3 x\,\mathrm{d}x \tag{12.39}$$
$$= 12 - 9/2. \tag{12.40}$$

Let's now instead first integrate over x.

$$\iint I_{0\leq x\leq 3}\, I_{0\leq y\leq 4}\, I_{x\leq y}\,\mathrm{d}x\,\mathrm{d}y = \int I_{0\leq y\leq 4}\int I_{0\leq x\leq 3}\, I_{x\leq y}\,\mathrm{d}x\,\mathrm{d}y \tag{12.41}$$
$$= \int_0^4 \int I_{0\leq x\leq \min\{3,y\}}\,\mathrm{d}x\,\mathrm{d}y \tag{12.42}$$
$$= \int_0^4 \int_0^{\min\{3,y\}} \mathrm{d}x\,\mathrm{d}y \tag{12.43}$$
$$= \int_0^4 \min\{3,y\}\,\mathrm{d}y \tag{12.44}$$
$$= \int_0^3 \min\{3,y\}\,\mathrm{d}y + \int_3^4 \min\{3,y\}\,\mathrm{d}y \tag{12.45}$$
$$= \int_0^3 y\,\mathrm{d}y + \int_3^4 3\,\mathrm{d}y \tag{12.46}$$
$$= 9/2 + 3. \tag{12.47}$$

s.1.16. Take c the normalization constant (why is $c = 1/4$), then using the previous exercise

$$P\{Y \leq 2X\} = E[I_{Y \leq 2X}] \tag{12.48}$$

$$= c \int_1^3 \int_2^4 I_{y \leq 2x} \, dy \, dx \tag{12.49}$$

$$= c \int_1^3 \int I_{2 \leq y \leq \min\{4, 2x\}} \, dy \, dx \tag{12.50}$$

$$= c \int_1^3 [\min\{4, 2x\} - 2]^+ \, dx \tag{12.51}$$

Now make a drawing of the function $[\min\{4, 2x\} - 2]^+$ on the interval $[1, 3]$ to see that

$$\int_1^3 [\min\{4, 2x\} - 2]^+ \, dx = \int_1^2 (2x - 2) \, dx + \int_2^3 (4 - 2) \, dx. \tag{12.52}$$

I leave the rest of the computation to you.

s.3.1. 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

$$\rho = \frac{\mathrm{Cov}[X, Y]}{\sqrt{V[X] V[Y]}} = \frac{\mathrm{Cov}[X, Y]}{\sigma_X \sigma_Y} \implies \rho \frac{\sigma_Y}{\sigma_X} = \frac{\mathrm{Cov}[X, Y]}{V[X]}. \tag{12.53}$$

Hence,

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

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

What if $\rho = 0$?

s.3.2. If we measure X in centimetres instead of metres, then X, $E[X]$ and σ_X are all multiplied by 100, and the prediction $\hat{Y}$ should also be expressed in centimetres But $V[Y]$ scales as length squared. This messes up the units.

s.3.3. Let us first do one run.

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

# number simulations and people
numSamples <- 50
N <- 5
```

```
7
8   # initialize empty result vector
9   res <- c()
10
11  # for loop to simulate repeatedly
12  for (i in 1:numSamples) {
13
14    # shuffle the N hats
15    x <- sample(1:N)
16
17    # number of people picking own hat (element by element the vectors x and
18    # 1:N are compared, which yields a vector of TRUE and FALSE, TRUE = 1 and
19    # FALSE = 0)
20    correctPicks <- sum(x == 1:N)
21
22    # append the result vector by the result of the current simulation
23    res <- append(res, correctPicks)
24  }
25
26  # printing of observed mean and variance
27  print(mean(res))
28  print(var(res))
```

s.4.1. See solution manual.

s.4.2. Define $V := X$ and $W := X + Y$. Observe that for any t_V, t_W, we have

$$t_V V + t_W W = t_V X + t_W (X + Y) \tag{12.54}$$

$$= (t_V + t_W) X + t_W Y. \tag{12.55}$$

Hence, any linear combination of V and W is a linear combination of X and Y. Since (X, Y) is bivariate 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 bivariate normally distributed.

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

$$\mu_V = \mathsf{E}[V] = \mathsf{E}[X] = \mu_X = 0, \tag{12.56}$$

and

$$\mu_W = \mathsf{E}[W] = \mathsf{E}[X + Y] = \mu_X + \mu_Y = 0. \tag{12.57}$$

Next, we have

$$\mathsf{V}[V] = \mathsf{V}[X] = \sigma_X^2 = 1, \tag{12.58}$$

and

$$\mathsf{V}[W] = \mathsf{V}[X + Y] = \mathsf{V}[X] + \mathsf{V}[Y] + 2\mathrm{Cov}(X, Y) \tag{12.59}$$

$$= 1 + 1 + 2\rho_{XY} \sigma_X \sigma_Y = 2(1 + \rho_{XY}). \tag{12.60}$$

Finally,

$$\mathrm{Cov}(V,W) = \mathrm{Cov}(X, X+Y) = \mathrm{Cov}(X,X) + \mathrm{Cov}(X,Y) \tag{12.61}$$

$$= \sigma_X^2 + \rho_{XY}\sigma_X\sigma_Y = 1 + \rho_{XY}, \tag{12.62}$$

and hence,

$$\rho_{VW} := \mathrm{Corr}(V,W) = \frac{\mathrm{Cov}(V,W)}{\sqrt{\mathsf{V}[V]\,\mathsf{V}[W]}} \tag{12.63}$$

$$= \frac{1 + \rho_{XY}}{\sqrt{1 \cdot 2(1 + \rho_{XY})}} \tag{12.64}$$

$$= \sqrt{\frac{1 + \rho_{XY}}{2}}. \tag{12.65}$$

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

$$f_{V,W}(v,w) = \frac{1}{2\pi\sigma_V\sigma_W\tau_{VW}}\exp\left(-\frac{1}{2\tau_{VW}^2}\left(\left(\frac{v}{\sigma_V}\right)^2 + \left(\frac{w}{\sigma_W}\right)^2 - 2\frac{\rho_{VW}}{\sigma_V\sigma_W}vw\right)\right), \tag{12.66}$$

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,

$$f_{V,W}(v,w) = \frac{1}{2\pi\sqrt{1 - (\rho_{XY})^2}}\exp\left(-\frac{1}{1 - \rho_{XY}}\left(v^2 + \frac{w^2}{2(1 + \rho_{XY})} - vw\right)\right). \tag{12.67}$$

s.4.3. Let X denote the value of the portfolio after a year in thousands of dollars. Then,

$$X := 2(1 + A) + 4(1 + B) + 2(1 + C) + 2 \tag{12.68}$$

$$= 10 + 2A + 4B + 2C. \tag{12.69}$$

Then,

$$\mathsf{E}[X] = \mathsf{E}[10 + 2A + 4B + 2C] \tag{12.70}$$

$$= 10 + 2\,\mathsf{E}[A] + 4\,\mathsf{E}[B] + 2\,\mathsf{E}[C] \tag{12.71}$$

$$= 10 + 2 \cdot 0.075 + 4 \cdot 0.1 + 2 \cdot 0.2 \tag{12.72}$$

$$= 10 + 0.15 + 0.4 + 0.4 \tag{12.73}$$

$$= 10.95 \tag{12.74}$$

s.4.4. We have

$$\mathsf{V}[X] = \mathsf{V}[10 + 2A + 4B + 2C] \tag{12.75}$$

$$= \mathsf{V}[2A] + \mathsf{V}[4B] + \mathsf{V}[2C] \tag{12.76}$$

$$+ 2\Big(\mathrm{Cov}(2A, 4B) + \mathrm{Cov}(2A, 2C) + \mathrm{Cov}(4B, 2C)\Big) \tag{12.77}$$

$$= 4\mathsf{V}[A] + 16\mathsf{V}[B] + 4\mathsf{V}[C] \tag{12.78}$$

$$+ 2\Big(8\mathrm{Cov}(A, B) + 4\mathrm{Cov}(A, C) + 8\mathrm{Cov}(B, C)\Big) \tag{12.79}$$

$$= 4\sigma_A^2 + 16\sigma_B^2 + 4\sigma_C^2 \tag{12.80}$$

$$+ 2\Big(8\rho_{AB}\sigma_A\sigma_B + 4\rho_{AC}\sigma_A\sigma_C + 8\rho_{BC}\sigma_B\sigma_C\Big) \tag{12.81}$$

$$= 4(0.07)^2 + 16(0.12)^2 + 4(0.17)^2 \tag{12.82}$$

$$+ 2\Big(8(0.7)(0.07)(0.12) + 4(-0.8)(0.07)(0.17) + 8(-0.3)(0.12)(0.17)\Big) \tag{12.83}$$

$$= 0.2856. \tag{12.84}$$

So

$$\sigma_X = \sqrt{0.2856} = 0.5344. \tag{12.85}$$

So X has a standard deviation of \$534.

s.4.5. We need to compute the probability $\mathsf{P}\{X \leq 10\}$. We have

$$\mathsf{P}\{X \leq 10\} = \mathsf{P}\left\{X - \mu_X \leq 10 - 10.95\right\} \tag{12.86}$$

$$= \mathsf{P}\left\{\frac{X - \mu_X}{\sigma_X} \leq \frac{10 - 10.95}{0.5344}\right\} \tag{12.87}$$

$$= \mathsf{P}\left\{Z \leq \frac{10 - 10.95}{0.5344}\right\} \tag{12.88}$$

$$= \mathsf{P}\{Z \leq -1.9625\} \tag{12.89}$$

$$= 0.0377. \tag{12.90}$$

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

s.4.6. 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

$$Y := 4(1 + A) + 4(1 + C) + 2 \tag{12.91}$$

$$= 10 + 4A + 4C. \tag{12.92}$$

Then,

$$E[Y] = E[10 + 4A + 4C] \tag{12.93}$$

$$= 10 + 4E[A] + 4E[C] \tag{12.94}$$

$$= 10 + 4(0.075) + 4(0.20) \tag{12.95}$$

$$= 11.1 \tag{12.96}$$

Moreover,

$$V[Y] = V[10 + 4A + 4C] \tag{12.97}$$

$$= V[4A] + V[4C] + 2\text{Cov}(4A, 4C) \tag{12.98}$$

$$= 4^2 V[A] + 4^2 V[C] + 2 \cdot 4 \cdot 4 \cdot \text{Cov}(A, C) \tag{12.99}$$

$$= 16(.07)^2 + 16(.17)^2 + 32(-.8)(.07)(.17) \tag{12.100}$$

$$= 0.23616, \tag{12.101}$$

which corresponds to a standard deviation of

$$\sigma_Y = \sqrt{V[Y]} = \sqrt{0.23616} = 0.4860 \tag{12.102}$$

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

s.5.1. 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

$$f_{X,Y}(x, y) = f_X(x) f_Y(x) =: f(x) f(y), \tag{12.103}$$

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 $dx\,dy$ must be such that

$$f(x)f(y)\,dx\,dy = g(r)\,dx\,dy, \tag{12.104}$$

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

$$\partial_\phi g(r) = 0 = f(x)\partial_\phi f(y) + f(y)\partial_\phi f(x). \tag{12.105}$$

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:

$$x = r\cos\phi, \qquad\qquad y = r\sin\phi. \tag{12.106}$$

Using the chain rule,

$$\partial_\phi f(x) = \partial_x f(x)\frac{dx}{d\phi} = f'(x)r(-\sin\phi) = -f'(x)y, \tag{12.107}$$

$$\partial_\phi f(y) = \partial_y f(y)\frac{dy}{d\phi} = f'(y)r\cos\phi = f'(y)x. \tag{12.108}$$

All this gives for (12.105)

$$0 = xf(x)f'(y) - yf(y)f'(x). \tag{12.109}$$

Simplifying,

$$\frac{f'(x)}{xf(x)} = \frac{f'(y)}{yf(y)}. \tag{12.110}$$

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

$$\frac{f'(x)}{xf(x)} = \frac{f'(y)}{yf(y)} = \alpha. \tag{12.111}$$

Hence, our f must satisfy for all x

$$f'(x) = \alpha x f(x). \tag{12.112}$$

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 $\alpha = -1/2\sigma^2$ to get the final answer:

$$f(x) = ae^{-x^2/2\sigma^2}. \tag{12.113}$$

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

s.5.2.

$$\iint_{C(N)} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = a^2 \iint_{C(N)} e^{-(x^2+y^2)/2\sigma}\,\mathrm{d}x\,\mathrm{d}y. \tag{12.114}$$

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

$$\frac{\partial(x,y)}{\partial(r,\phi)} = \begin{vmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{vmatrix} = r(\cos^2\phi + \sin^2\phi) = r. \tag{12.115}$$

Therefore

$$\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\phi, \tag{12.116}$$

from which

$$a^2 \iint_{C(N)} e^{-(x^2+y^2)/2\sigma}\,\mathrm{d}x\,\mathrm{d}y = a^2 \iint_{C(N)} e^{-r^2/2\sigma^2} r\,\mathrm{d}r\,\mathrm{d}\phi \tag{12.117}$$

$$= a^2 \int_0^N \int_0^{2\pi} e^{-r^2/2\sigma^2} r\,\mathrm{d}r\,\mathrm{d}\phi \tag{12.118}$$

$$= a^2 2\pi \int_0^N e^{-r^2/2\sigma^2} r\,\mathrm{d}r \tag{12.119}$$

$$= -a^2 2\pi\sigma^2 e^{-r^2/2\sigma^2}\Big|_0^N \tag{12.120}$$

$$= a^2 2\pi\sigma^2 (1 - e^{-N^2/2\sigma^2}), \tag{12.121}$$

where we use (12.112).

Therefore, for the square,

$$a^2 2\pi\sigma^2 (1 - e^{-N^2/2\sigma^2}) \le \iint_{S(N)} f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \le a^2 2\pi\sigma^2 (1 - e^{-2N^2/2\sigma^2}). \tag{12.122}$$

Taking $N \to \infty$ we conclude that

$$a^2 2\pi\sigma^2 = \iint f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y = a^2 \iint e^{-x^2/2\sigma^2} e^{-y^2/2\sigma}\,\mathrm{d}x\,\mathrm{d}y \tag{12.123}$$

$$= a^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2/2\sigma^2} e^{-y^2/2\sigma}\,\mathrm{d}x\,\mathrm{d}y = a^2 \left(\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\,\mathrm{d}x \right)^2, \tag{12.124}$$

and therefore

$$\int_{-\infty}^{\infty} e^{-x^2/2\sigma^2}\,\mathrm{d}x = \sqrt{2\pi}\sigma. \tag{12.125}$$

s.5.3. 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

$$f_{X,Z}(x,z)\,\mathrm{d}x\,\mathrm{d}z = f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \tag{12.126}$$

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

$$f_{X,Z}(x,z) = f_{X,Y}(x,y)\frac{\partial(x,y)}{\partial(x,z)}. \tag{12.127}$$

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

$$\left(\frac{\partial(x,y)}{\partial(x,z)}\right)^{-1} = \frac{\partial(x,z)}{\partial(x,y)} = \begin{vmatrix} 1 & 0 \\ y & x \end{vmatrix} = x. \tag{12.128}$$

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

$$f_{X,Z}(x,z) = f_{X,Y}(x,y)\frac{1}{x} = f(x)f(y)/x = f(x)f(z/x)/x. \tag{12.129}$$

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

$$1 \le z/x < 10 \iff 1 \ge x/z > 1/10 \iff z \ge x > z/10. \tag{12.130}$$

Combining both constraints gives

$$\max\{1, z/10\} < x \le \min\{10, z\}. \tag{12.131}$$

All in all,

$$f_{X,Z}(x,z) = f(x)f(z/x)/x\, I_{\max\{1,z/10\}<x\le\min\{10,z\}}. \tag{12.132}$$

As a test, $z = 110 \implies x > 110/10 > 10$, but the indicator says that $x \le 10$, 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

$$f_Z(z) = \int_1^{10} f_{X,Z}(x,z)\,\mathrm{d}x. \tag{12.133}$$

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,

$$f_Z(z) = \frac{1}{9^2} \int_1^{10} I_{\max\{1, z/10\} < x \leq \min\{10, z\}} \frac{\mathrm{d}x}{x} \tag{12.134}$$

$$= \frac{\log(\min\{10, z\}) - \log(\max\{1, z/10\})}{81} I_{1 \leq z < 100}. \tag{12.135}$$

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.6.1. 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

$$P\{Z = n\} = \sum_{k=0}^{\infty} P\{X = k\} \, P\{Z = n \mid X = k\} \tag{12.136}$$

$$= \sum_{k=0}^{\infty} P\{X = k\} \, P\{Y = n - X \mid X = k\} \tag{12.137}$$

$$= \sum_{k=0}^{n} P\{X = k\} \, P\{Y = n - k\} \tag{12.138}$$

$$= \sum_{k=0}^{n} \frac{e^{-\lambda}}{k!} \lambda^k \cdot \frac{e^{-\mu}}{(n-k)!} \mu^{n-k} \tag{12.139}$$

$$= e^{-(\lambda+\mu)} \sum_{k=0}^{n} \frac{1}{k!(n-k)!} \cdot \lambda^k \mu^{n-k} \tag{12.140}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \cdot \lambda^k \mu^{n-k} \tag{12.141}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \cdot \lambda^k \mu^{n-k} \tag{12.142}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} (\lambda + \mu)^k. \tag{12.143}$$

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

s.6.2. We use mathematical induction. For $n = 1$ we have $T_1 = X_1$, which follows an exponential distribution with rate λ. We get

$$\frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} = \frac{\lambda^1}{0!} t^0 e^{-\lambda t} \tag{12.144}$$

$$= \lambda e^{-\lambda t}, \tag{12.145}$$

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

$$f_{T_{n-1}}(t) = \frac{\lambda^{n-1}}{(n-2)!} t^{n-2} e^{-\lambda t}, \quad t > 0. \tag{12.146}$$

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 $t > 0$:

$$f_{T_n}(t) = \int_{-\infty}^{\infty} f_{T_{n-1}}(t - x) f_{X_n}(x) dx \tag{12.147}$$

$$= \int_0^t f_{T_{n-1}}(t - x) f_{X_n}(x) dx \tag{12.148}$$

$$= \int_0^t \frac{\lambda^{n-1}}{(n-2)!} (t - x)^{n-2} e^{-\lambda(t-x)} \lambda e^{-\lambda x} dx \tag{12.149}$$

$$= \frac{\lambda^n e^{-\lambda t}}{(n-2)!} \int_0^t (t - x)^{n-2} dx \tag{12.150}$$

$$= \frac{\lambda^n e^{-\lambda t}}{(n-2)!} \left[-\frac{(t - x)^{n-1}}{n-1} \right]_{x=0}^t \tag{12.151}$$

$$= \frac{\lambda^n e^{-\lambda t}}{(n-2)!} \left(0 - \frac{-t^{n-1}}{n-1} \right) \tag{12.152}$$

$$= \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}. \tag{12.153}$$

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

s.6.3. Recall the pdf of a $\mathcal{N}(\mu, \sigma^2)$ random variable T:

$$f_T(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(t - \mu)^2}{2\sigma^2} \right\}, \quad t \in \mathbb{R}. \tag{12.154}$$

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

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) dx \tag{12.155}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(z - x)^2}{2} \right\} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{x^2}{2} \right\} dx \tag{12.156}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ -\frac{z^2 - 2zx + 2x^2}{2} \right\} dx \tag{12.157}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ -\left[\frac{1}{2} z^2 - zx + x^2 \right] \right\} dx \tag{12.158}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left\{ -\left[\left(x - \frac{1}{2} z \right)^2 + \frac{1}{4} z^2 \right] \right\} dx \tag{12.159}$$

$$= \frac{1}{2\pi} \exp\left\{ -\frac{1}{4} z^2 \right\} \int_{-\infty}^{\infty} \exp\left\{ -\left(x - \frac{1}{2} z \right)^2 \right\} dx \tag{12.160}$$

$$= \frac{1}{\sqrt{2\pi} \cdot \sqrt{2}} \exp\left\{ -\frac{z^2}{2 \cdot \sqrt{2}^2} \right\} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot (1/\sqrt{2})} \exp\left\{ -\frac{\left(x - \frac{1}{2} z \right)^2}{2 \cdot (1/\sqrt{2})^2} \right\} dx \tag{12.161}$$

$$= \frac{1}{\sqrt{2\pi} \cdot \sqrt{2}} \exp\left\{ -\frac{z^2}{2 \cdot \sqrt{2}^2} \right\}, \tag{12.162}$$

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.7.1. We have by the Bayes' billiard stories that

$$\binom{n}{k}\int_0^1 p^k(1-p)^{n-k}\,\mathrm{d}p = \frac{1}{n+1}, \tag{12.163}$$

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.7.2. 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, $r \le w$, 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

$$\mathsf{P}\{X=k\} = w\int_0^1 \binom{b}{k}p^k(1-p)^{b-k} \times \binom{w-1}{r-1}p^{r-1}(1-p)^{w-r}\,\mathrm{d}p. \tag{12.164}$$

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:

$$\mathsf{P}\{X=k\} = w\binom{b}{k}\binom{w-1}{r-1}\int_0^1 p^{k+r-1}(1-p)^{b-k+w-r}\,\mathrm{d}p \tag{12.165}$$

$$= w\binom{b}{k}\binom{w-1}{r-1}\beta(k+r, b-k+w-r+1) \tag{12.166}$$

$$= w\binom{b}{k}\binom{w-1}{r-1}\frac{(r+k-1)!(b-k+w-r)!}{(w+b)!}. \tag{12.167}$$

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

$$\mathsf{P}\{X=k\} = \frac{\binom{w}{r-1}\binom{b}{k}}{\binom{w+b}{r+k-1}}\frac{w-(r-1)}{w-(r-1)+b-k} \tag{12.168}$$

$$\tag{12.169}$$

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—,

$$P\{X = k\} = \frac{\binom{w}{r-1}\binom{b}{k}}{\binom{w+b}{r+k-1}} \frac{w-r+1}{w+b-r-k+1} \tag{12.170}$$

$$= \binom{w}{r-1}\binom{b}{k}\frac{(r+k-1)!(w+b-r-k+1)!}{(w+b)!}\frac{w-r+1}{w+b-r-k+1} \tag{12.171}$$

$$= \binom{w}{r-1}\binom{b}{k}\frac{(r+k-1)!(w+b-r-k)!}{(w+b)!}(w-r+1) \tag{12.172}$$

$$= w\binom{w-1}{r-1}\binom{b}{k}\frac{(r+k-1)!(w+b-r-k)!}{(w+b)!}, \tag{12.173}$$

where we use that

$$w\binom{w-1}{r-1} = w\frac{(w-1)!}{(w-r)!(r-1)!} = \frac{w!}{(w-r)!(r-1)!} \tag{12.174}$$

$$= \frac{w!}{(w-r+1)!(r-1)!}(w-r+1) = \binom{w}{r-1}(w-r+1). \tag{12.175}$$

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

So, inspecting (12.165) 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.7.3. 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.7.4. 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,

$$N(w,b) = \frac{b}{w+b}(1 + N(w,b-1)). \tag{12.176}$$

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

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

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.7.5. 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 (12.199), 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.8.1. We need that the pdf f_X integrates to one. Hence, we need

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \qquad \Longleftrightarrow \qquad (12.177)$$

$$\int_0^4 cdx = 1 \qquad \Longleftrightarrow \qquad (12.178)$$

$$4c = 1 \qquad \Longleftrightarrow \qquad (12.179)$$

$$c = 1/4. \qquad (12.180)$$

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

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \qquad \Longleftrightarrow \qquad (12.181)$$

$$\int_{-\infty}^{\infty} c \cdot e^{-\frac{(x-4)^2}{8}} dx = 1 \qquad \Longleftrightarrow \qquad (12.182)$$

$$c \cdot \sqrt{2\pi} \cdot 2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot 2} e^{-\frac{1}{2}\frac{(x-4)^2}{2^2}} dx = 1 \qquad \Longleftrightarrow \qquad (12.183)$$

$$c \cdot \sqrt{2\pi} \cdot 2 \cdot 1 = 1 \qquad \Longleftrightarrow \qquad (12.184)$$

$$c = \frac{1}{\sqrt{2\pi} \cdot 2}. \qquad (12.185)$$

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.8.3. See the book.

s.8.4. 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.8.5. See the book.

s.8.6. See the book.

s.9.1. 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

$$\mathsf{E}[X | X \leq \tau] = \int tf(t)\,dt \qquad (12.186)$$

$$= \frac{1}{1 - e^{-\lambda \tau}} \int_0^\tau \lambda t e^{-\lambda t}\,dt. \qquad (12.187)$$

Simplyfing,

$$\int_0^\tau \lambda t e^{-\lambda t}\,dt = -t e^{-\lambda t}\Big|_0^\tau + \int_0^\tau e^{-\lambda t}\,dt \qquad (12.188)$$

$$= -\tau e^{-\lambda \tau} + \frac{e^{-\lambda \tau} - 1}{-\lambda} \qquad (12.189)$$

$$= \frac{1 - e^{-\lambda \tau} - \tau \lambda e^{-\lambda \tau}}{\lambda}, \qquad (12.190)$$

$$\mathsf{E}[X|X \le \tau] = \frac{1 - e^{-\lambda\tau} - \tau\lambda e^{-\lambda\tau}}{\lambda(1 - e^{-\lambda\tau})}. \tag{12.191}$$

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

$$\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X|Y]], \quad \text{Adam's law} \tag{12.192}$$

$$= \mathsf{E}[X|Y = 1]\,\mathsf{P}\{Y = 1\} + \mathsf{E}[X|Y = 0]\,\mathsf{P}\{Y = 0\}, \quad \text{LOTE} \tag{12.193}$$

$$= \mathsf{E}[X|X \le \tau]\,\mathsf{P}\{X \le \tau\} + \mathsf{E}[X|X > \tau]\,\mathsf{P}\{X > \tau\} \tag{12.194}$$

$$= \frac{1 - e^{-\lambda\tau} - \tau\lambda e^{-\lambda\tau}}{\lambda(1 - e^{-\lambda\tau})}(1 - e^{-\lambda\tau}) + (\tau + 1/\lambda)e^{-\lambda\tau} \tag{12.195}$$

$$= 1/\lambda - e^{-\lambda\tau}/\lambda - \tau e^{-\lambda\tau} + \tau e^{-\lambda T} + e^{-\lambda\tau}/\lambda \tag{12.196}$$

$$= 1/\lambda. \tag{12.197}$$

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

s.9.2. Let X_i be the lifetime of machine i. Suppose that the second machine does not break down before τ, then

$$\mathsf{E}[T|X_2 > \tau] = \mathsf{E}[X_1 + \tau + C|X_2 > \tau] = 1/\lambda + \tau + \mathsf{E}[T], \tag{12.198}$$

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,

$$\mathsf{E}[T|X_2 \le \tau] = \mathsf{E}[X_1 + X_2|X_2 \le \tau] = \frac{1}{\lambda} + \frac{1 - e^{-\lambda\tau} - \tau\lambda e^{-\lambda\tau}}{\lambda(1 - e^{-\lambda\tau})}. \tag{12.199}$$

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

$$\mathsf{E}[T] = \mathsf{E}[T|X_2 \le \tau]\,\mathsf{P}\{X_2 \le \tau\} + \mathsf{E}[T|X_2 \ge \tau]\,\mathsf{P}\{X_2 \ge \tau\} \tag{12.200}$$

$$= (1 - e^{-\lambda\tau})\left(\frac{1}{\lambda} + \frac{1 - e^{-\lambda\tau} - \tau\lambda e^{-\lambda\tau}}{\lambda(1 - e^{-\lambda\tau})}\right) + e^{-\lambda\tau}\left(\frac{1}{\lambda} + \tau + \mathsf{E}[T]\right) \tag{12.201}$$

$$= \frac{1}{\lambda} + \frac{1 - e^{-\lambda\tau} - \tau\lambda e^{-\lambda\tau}}{\lambda} + e^{-\lambda\tau}\tau + e^{-\lambda\tau}\,\mathsf{E}[T] \tag{12.202}$$

$$\implies \tag{12.203}$$

$$\left(1 - e^{-\lambda\tau}\right)\mathsf{E}[T] = \frac{2 - e^{-\lambda\tau}}{\lambda} \tag{12.204}$$

$$\implies \tag{12.205}$$

$$\mathsf{E}[T] = \frac{2 - e^{-\lambda\tau}}{\lambda(1 - e^{-\lambda\tau})}. \tag{12.206}$$

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

$$T_n = 1 + T_{n+1} P\{\text{draw a new ball}\} + T_n P\{\text{draw an old ball}\} \tag{12.207}$$

$$= 1 + T_{n+1}\frac{N-n}{N} + T_n\frac{n}{N} \tag{12.208}$$

$$\implies \tag{12.209}$$

$$T_n\frac{N-n}{N} = 1 + T_{n+1}\frac{N-n}{N} \tag{12.210}$$

$$\implies \tag{12.211}$$

$$T_n = \frac{N}{N-n} + T_{n+1}. \tag{12.212}$$

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

$$T_{N-1} = \frac{N}{N-(N-1)} + T_N = N, \tag{12.213}$$

and so on.

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

s.9.4. Here is the 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 |────────────────────────────
1   N = 45
2
3   bigT = function(n) {
4     if(n >= N) {
5       return(0)
6     }
7     return(N / (N - n) + bigT(n + 1))
8   }
9
10  #print(bigT(44))
11  #print(bigT(0))
────────────────────────────────────────────────────────────────
```

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

s.9.5. 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

$$\binom{n}{6-k}\binom{N-n}{k}/\binom{N}{6}. \tag{12.214}$$

Therefore, by LOTE,

$$T_n \approx 1 + \sum_{k=0}^{6} \frac{\binom{n}{6-k}\binom{N-n}{k}}{\binom{N}{6}} T_{n+k}. \tag{12.215}$$

But there is a slight problem, if there are just 2 new balls, we cannot pick $k = 6$ new balls. In general, we can pick $k = \min\{6, N-n\}$ new balls. Hence,

$$T_n = 1 + \sum_{k=0}^{\min\{6,N-n\}} \frac{\binom{n}{6-k}\binom{N-n}{k}}{\binom{N}{6}} T_{n+k} \tag{12.216}$$

$$\implies \tag{12.217}$$

$$T_n - \frac{\binom{n}{6}\binom{N-n}{0}}{\binom{N}{6}} T_n = 1 + \sum_{k=1}^{\min\{6,N-n\}} \frac{\binom{n}{6-k}\binom{N-n}{k}}{\binom{N}{6}} T_{n+k} \tag{12.218}$$

$$\implies \tag{12.219}$$

$$T_n \left(\binom{N}{6} - \binom{n}{6} \right) = \binom{N}{6} + \sum_{k=1}^{\min\{6,N-n\}} \binom{n}{6-k}\binom{N-n}{k} T_{n+k} \tag{12.220}$$

$$\tag{12.221}$$

And this is the final result.

s.9.6. 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
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
1   S = function(n, N) {
2     if(n >= N) {
3       return(0)
4     }
```

```
5      return(1 + S(n + 1, N) + S(n + 2, N))
6    }
7
8    for (N in 1:24) {
9      print(S(0, N))
10   }
```

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 compuation 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
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 ───────────────────────────────────
1    library("memoise")
2
3    S = function(n, N) {
4      if(n >= N) {
5        return(0)
6      }
7      cat("Called for n =", n, "\n")
8      return(1 + S(n + 1, N) + S(n + 2, N))
9    }
10   S = memoise(S)
11
12   S(0, 5)
```

Now the result is this:

Called for n = 0

Called for n = 1

<div align="center">
Called for n = 2

Called for n = 3

Called for n = 4
</div>

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
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
                              ┌─────────┐
                              │ R Code  │
1   S = function(n, N) {
2     if(n >= N) {
3       return(0)
4     }
5     cat("Called for n =", n, "\n")
6     return(1 + S(n + 1, N) + S(n + 2, N))
7   }
8
9   S(0, 5)
```

This is the result:

<div align="center">
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
</div>

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
from math import comb
from functools import lru_cache


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
1   library("memoise")
2
3   N = 45
4
5   bigT = function(n, m) {
6     if(n >= N) {
7       return(0)
8     }
9     res = choose(N, m)
10    for (k in 1:min(m, N - n)) {
11      P = choose(n, m - k)
12      P = P * choose(N - n, k)
13      res = res + P * bigT(n + k, m)
14    }
15    return(res / (choose(N, m) - choose(n, m)))
16  }
17  bigT = memoise(bigT)
18
19  bigT(0, 1)
20  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.10.1. 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

$$\mathsf{E}[S] = \mathsf{E}[\,\mathsf{E}[S|X]\,] \tag{12.222}$$
$$= \mathsf{E}[X/2] \tag{12.223}$$
$$= \mathsf{E}[X]/2 \tag{12.224}$$
$$= 100/4 = 25. \tag{12.225}$$

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

s.10.2. Let X denote the point where the initial stick gets broken; let Y denote the length of the first remaining stick; let Z denote the point where the first remaining stick gets broken; let S denote the length of the final remaining stick. We have

$$X \sim \text{Unif}(0, 100), \tag{12.226}$$
$$Y = \max\{X, 100 - X\}, \tag{12.227}$$
$$Z \sim \text{Unif}(0, Y), \tag{12.228}$$
$$S = \max\{Z, Y - Z\}. \tag{12.229}$$

We have

$$\mathsf{E}[S] = \mathsf{E}[\,\mathsf{E}[S|Y]\,], \tag{12.230}$$

where

$$\mathsf{E}[S|Y = y] = \mathsf{E}[\max\{Z, Y - Z\}|Y = y] \tag{12.231}$$
$$= \mathsf{E}[\max\{Z, y - Z\}|Y = y] \tag{12.232}$$
$$= \mathsf{P}\{Z \geq y|Y = y\}\,\mathsf{E}[Z|Z \geq y, Y = y] + \mathsf{P}\{Z < y|Y = y\}\,\mathsf{E}[y - Z|Z < y, Y = y] \tag{12.233}$$
$$= \frac{1}{2} \cdot \frac{3y}{4} + \frac{1}{2} \cdot \frac{3y}{4} \tag{12.234}$$
$$= \frac{3y}{4}. \tag{12.235}$$

Hence,

$$\mathsf{E}[S] = \mathsf{E}[\,\mathsf{E}[S|Y]\,] \tag{12.236}$$
$$= \mathsf{E}\left[\frac{3Y}{4}\right] \tag{12.237}$$
$$= \frac{3}{4}\mathsf{E}[Y]. \tag{12.238}$$

We have

$$\mathsf{E}[Y] = \mathsf{E}[\max\{X, 100 - X\}] \tag{12.239}$$
$$= \frac{3}{4} \cdot 100 = 75, \tag{12.240}$$

by symmetry with $\mathsf{E}[S|Y = y]$ above. Hence,

$$\mathsf{E}[S] = \frac{3}{4}\mathsf{E}[Y] \tag{12.241}$$

$$= \frac{3}{4} \cdot 75 = 56.25 \tag{12.242}$$

$$\tag{12.243}$$

So we expect to end up with a stick of length 56.25 cm.

s.10.3. 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,

$$\mathsf{E}[Y] = \mathsf{E}[\mathsf{E}[Y|\lambda]]. \tag{12.244}$$

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

$$\mathsf{E}[\mathsf{E}[Y|\lambda]] = \lambda t. \tag{12.245}$$

Hence,

$$\mathsf{E}[Y] = \mathsf{E}[\mathsf{E}[Y|\lambda]] \tag{12.246}$$

$$= \mathsf{E}[\lambda t] \tag{12.247}$$

$$= \mathsf{E}[\lambda]t \tag{12.248}$$

$$= \frac{r_0}{b_0}t. \tag{12.249}$$

s.10.4. See the book.

s.10.5. See the book.