IMM - DTU

02405 Probability
2021-27-01
Dagh Nielsen

These are suggested solutions and explanations for the December 2020 exam in the course 02405 *Sandsynlighedsregning* at DTU. Page references are to the book *Probability* by Jim Pitman.

Problem 1

This is a situation which we know from the book as the "expectation of a function g of x" on page 263. Using $g(x) = \sqrt{x}$, the formula on page 263 and the given limits 0 and 1 for x, we find that

$$\begin{aligned}
E(Y) &= E(g(X)) \\
&= \int_{-\infty}^{\infty} g(x)f(x)\,dx \\
&= \int_{0}^{1} \sqrt{x}\cdot 3x^2\,dx \\
&= \tfrac{6}{7}x^{3.5}\Big|_{x=0}^{x=1} \\
&= \tfrac{6}{7}.
\end{aligned}$$

Answer 4 is correct.

Problem 2

This is a series of 10 independent Bernoulli trials with success parameter $p = \frac{1}{10}$, where we define success as a frog having crystalized filling. Hence the number of frogs with crystalized filling is a $binomial(10, \frac{1}{10})$ distribution (see p. 81). Using this, we see that

$$\begin{aligned}
&P(\textit{at most two frogs with crystalized filling}) \\
&= \sum_{i=0}^{2} \binom{10}{i} p^i (1-p)^i \\
&= \binom{10}{0}\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^{10} + \binom{10}{1}\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^{9} + \binom{10}{2}\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^{8}
\end{aligned}$$

which simplifies to option 4.

Answer 4 is correct.

Problem 3

Let A denote the event that the train is on time, and let B denote the event that the commuter can find a seating. We can now use the "Multiplication Rule" from page 37:

$$P(A \cap B) = P(B|A) \cdot P(A) = \frac{1}{5} \cdot \frac{1}{3} = \frac{1}{15}.$$

Answer 5 is correct.

Problem 4

We use a number of formulas to solve this problem:

$$\begin{aligned}
Var(X+Y) &= Var(X) + Var(Y) + 2Cov(X,Y) \quad \text{(variance of a sum, p. 430)} \\
Var(aX) &= a^2 Var(X) \quad \text{(scaling and shifting, p. 188)} \\
Cov(aX, bY) &= abCov(X,Y) \quad \text{(bilinearity of covariance, p. 444)} \\
Cov(X,Y) &= E(XY) - E(X)E(Y) \quad \text{(alternative formula for covariance, p. 430)}
\end{aligned}$$

Applying these in succession and inserting the given information, we obtain:

$$\begin{aligned}
Var(Z) &= Var(3X + 2Y) \\
&= Var(3X) + Var(2Y) + 2Cov(3X, 2Y) \\
&= 3^2 Var(X) + 2^2 Var(Y) + 2 \cdot 3 \cdot 2Cov(X,Y) \\
&= 9Var(X) + 4Var(Y) + 12[E(XY) - E(X)E(Y)] \\
&= 9 \cdot 4 + 4 \cdot 9 + 12[10 - 3 \cdot 2] \\
&= 120.
\end{aligned}$$

Answer 4 is correct.

Problem 5

The situation is sampling without replacement (see page 125). We let the red balls be "good" and the green balls be "bad". We have 8 red balls ($G = 8$) and 16 green balls ($B = 16$), and a total of 24 balls ($N = 24$). We request the probability of drawing 1 good red and 2 bad green in 3 draws ($g = 1$, $b = 2$, $n = 3$). Inserting all this in the formula on page 125, we obtain:

$$P(g \text{ good and } b \text{ bad}) = \frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}} = \frac{\binom{8}{1}\binom{16}{2}}{\binom{24}{3}}.$$

Answer 5 is correct.

Problem 6

Let A denote the event that the front wheel collapses, and let B denote the event that the rear wheel collapses. We are given that $P(A) = 0.3$ and $P(B) = 0.2$.

We are told that the probability that at least one wheel collapses is 0.5. In terms of A and B, this means that $P(A \cup B)$=0.5 (since "at least one" means one or the other or both, which is the union).

We are asked for the probability that both wheels collapse. In terms of A and B, this means $P(A \cap B)$.

We can now use a slight rearrangement of the inclusion-exclusion formula from p. 22:

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) \\ &= 0.3 + 0.2 - 0.5 \\ &= 0. \end{aligned}$$

Answer 1 is correct.

Problem 7

Knowing just the expected value and the standard deviation, we can use Chebychev's Inequality (p. 191):

$$P(|X - E(X)| \geq k \cdot SD(X)| \leq \frac{1}{k^2}.$$

This problem is a bit tricky though because it involves a random variable X described as "a difference", while we are also going to need to consider the difference and the absolute difference between X and $E(X)$. So let us forget that X is "a difference" and just regard X as a generic random variable.

Chebychev's Inequality gives us an upper bound for the probability that the absolute value of the difference between X and $E(X)$ is greater than some number. Now, if the difference of X from $E(X)$ in one direction is greater (or smaller) than some number, the absolute difference is *also* greater than said number. In other words, the event we are looking at in the question ($X < -5$) is a sub-event of the appropriate event in Chebychev's Inequality ($|X - E(X)| \geq 4$ with $E(X) = -1$) and thus has lower or equal probability. Hence, the bound for Chebychev's event is also a bound for the question's event!

So let us write down the comparison to Chebychev's event, using $E(X) = -1$:

$$\begin{aligned} P(X < -5) &= P(X - (-1) < -4) \\ &\leq P(|X - (-1)| \geq 4). \end{aligned}$$

This fits the left side in Chebychev's Inequality

$$P(|X - E(X)| \geq k \cdot SD(X)| \leq \frac{1}{k^2}$$

with $E(X) = -1$, $SD(X) = 2$ and $k = 2$, and we can continue:

$$\begin{aligned} P(|X - (-1)| \geq 2 \cdot 2) &\leq \frac{1}{k^2} \\ &\leq \frac{1}{2^2} \\ &= \frac{1}{4}. \end{aligned}$$

So in total we can conclude $P(X < -5) \leq \frac{1}{4}$.

Answer 2 is correct.

Problem 8

The situation is a series of $n = 400$ Bernoulli trials with success probability $p = \frac{3}{4}$. We assume the trials are independent.

We can calculate the exact probability of at least 280 successes with the binomial distribution (p. 81):

$$P(X \geq 280) = \sum_{i=280}^{400} \binom{400}{i} \left(\frac{3}{4}\right)^i \left(\frac{1}{4}\right)^{400-1}.$$

This looks like option 1, but the exponents are interchanged. Also option 2 is wrong, since if we reframe the question into counting discarded components, we should have between 0 and 119 discarded components, not 120.

We can instead approximate with the normal distribution (p. 99):

$$P(\text{a to b successes}) = \Phi\left(\frac{b + 0.5 - n \cdot p}{\sqrt{n \cdot p \cdot q}}\right) - \Phi\left(\frac{a - 0.5 - n \cdot p}{\sqrt{n \cdot p \cdot q}}\right)$$

Here the first term is routinely replaced by 1 when we are counting up to all n successes. When we simplify the denominator, the second term evaluates to

$$\Phi\left(\frac{280 - 0.5 - 400 \cdot 0.75}{\sqrt{400 \cdot 0.75 \cdot 0.25}}\right) = \Phi\left(\frac{280 - 0.5 - 300}{5\sqrt{3}}\right).$$

So we obtain the normal approximation

$$P(\text{280 to 400 successes}) = 1 - \Phi\left(\frac{280 - 0.5 - 300}{5\sqrt{3}}\right).$$

Answer 5 is correct.

Problem 9

Since the point is chosen randomly in some region, we have a uniform distribution, and thus the constant density c times the area of the region is 1 (see for example Example 1 on page 351).

The area of the region is $1^2 - 0.5^2 = 3/4$, and c is the reciprocal of that value, so $c = 4/3$.

Answer 4 is correct.

Problem 10

Let R denote the event that there is a predator nearby, and let S denote the event that the bird species screams.

We can use Bayes' Theorem (p. 49). We are given the prior probabilities of a predator or not a predator

$$\begin{aligned} P(R) &= 0.05 \\ P(R^{\complement}) &= 1 - 0.05 = 0.95 \end{aligned}$$

and the likelihoods of a scream given a predator or not a predator

$$\begin{aligned} P(S|R) &= 0.8 \\ P(S|R^{\complement}) &= 0.05. \end{aligned}$$

We can calculate the probability of the evidence, a scream, by

$$\begin{aligned} P(S) &= P(S|R)P(R) + P(S|R^{\complement})P(R^{\complement}) \\ &= 0.8 \cdot 0.05 + 0.05 \cdot 0.95. \end{aligned}$$

Inserting into Bayes' formula to find the posterior probability of a predator given a scream, we obtain:

$$\begin{aligned} P(R|S) &= \frac{P(S|R)P(R)}{P(S)} \\ &= \frac{0.8 \cdot 0.05}{0.8 \cdot 0.05 + 0.05 \cdot 0.95} \\ &= \frac{0.8}{0.8 + 0.95} \\ &= \frac{16/20}{16/20 + 19/20} \\ &= \frac{16}{16 + 19} \\ &= \frac{16}{35}. \end{aligned}$$

Answer 1 is correct.

Problem 11

The exponential distribution has the memoryless property, see page 279. So, in the given setup, the waiting time until the next bus does not depend on the time since the last arrival.

On page 279 we see that if the waiting time is exponentially distributed with rate, or parameter, λ, then the expected waiting time is $\frac{1}{\lambda}$.

Answer 4 is correct.

Problem 12

Since X and Y are independent standard normal distributions, we can use rotational symmetry, see figure 5 page 362. Hence, we can rotate the shaded region so that it becomes a cross along the x-axis and y-axis with the width of the beams given by $\frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}} = 2 \cdot \frac{1}{\sqrt{8}}$. The event we are discussing is then the union of the event that X is in the interval $[-\frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}]$, or that Y is in the same interval. Let's denote these two events A and B.

For the union of these two events, we can use the inclusion-exclusion formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Now, we know that A and B are independent, so $P(A \cap B) = P(A)P(B)$. And we can infer from the symmetry that $P(A) = P(B)$. Inserting this, we obtain:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 2P(A) - P(A)^2. \end{aligned}$$

Now, to calculate $P(A) = P(X \in [-\frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}])$, we can use that X is a standard normal variable. Then we obtain:

$$P(A) = \Phi\left(\frac{1}{\sqrt{8}}\right) - \Phi\left(\frac{-1}{\sqrt{8}}\right) = 2\Phi\left(\frac{1}{\sqrt{8}}\right) - 1.$$

Inserting this in the formula above, we obtain:

$$\begin{aligned} P(A \cup B) &= 2P(A) - P(A)^2 \\ &= 2\left[2\Phi\left(\frac{1}{\sqrt{8}}\right) - 1\right] - \left[2\Phi\left(\frac{1}{\sqrt{8}}\right) - 1\right]^2 \\ &= 8\Phi\left(\frac{1}{\sqrt{8}}\right) - 4\Phi\left(\frac{1}{\sqrt{8}}\right)^2 - 3. \end{aligned}$$

Answer 5 is correct.

Problem 13

We are in a situation with a change of variable, given by $Y = \log(X)$. This function is one-to-one, so we can use the formula on page 304:

$$f_Y(y) = \frac{f_X(x)}{\left|\frac{dy}{dx}\right|}.$$

We obtain the denominator

$$\left|\frac{dy}{dx}\right| = \left|\frac{1}{x}\right| = \frac{1}{x}.$$

We notice that $\frac{1}{x}$ will cancel out the $\frac{1}{x}$ in the numerator, so there is no need to express it in terms of y.

In the numerator, we can substitute $y = \log(x)$ in the exponent:

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{1}{2}(\log(x))^2} \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{1}{2}y^2}. \end{aligned}$$

Inserting, we obtain:

$$f_Y(y) = \frac{f_X(x)}{\left|\frac{dy}{dx}\right|} = \frac{\frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{1}{2}y^2}}{\frac{1}{x}} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}.$$

Answer 1 is correct.

Problem 14

We can use the *average conditional expectation* formula from the bottom of page 425:

$$E(Y) = \int E(Y|X = x)f_X(x)\,dx.$$

We interchange X and Y so it fits our question:

$$E(X) = \int E(X|Y = y)f_Y(y)\,dy.$$

Then let us translate carefully the given information.

"The random variable Y is *exponential*(2) distributed" means (p. 279) that Y has density

$$f_Y(y) = 2e^{-2y}.$$

And we note that this density applies only for $y \geq 0$, as for all exponential distributions. This is the limit we need to use in the integral.

"The conditional distribution of the random variable X given $Y = y$ is *exponential* $\left(\frac{1}{y}\right)$ distributed" means, in terms of expectation, that

$$E(X|Y = y) = 1/\left(\frac{1}{y}\right) = y.$$

Here we have used that the expectation of the exponential distribution is the reciprocal of the rate (p. 279).

Inserting these two things in the formula above yields

$$\begin{aligned} E(X) &= \int E(X|Y = y) f_Y(y)\, dy \\ &= \int_0^\infty y \cdot 2e^{-2y}\, dy. \end{aligned}$$

Integration by parts or Maple then provides the result $E(X) = \frac{1}{2}$.

Answer 2 is correct.

Problem 15

It is useful to argue through an interpretation of X. Let us consider a roof divided into n equally big patches, and let X_i denote the number of raindrops landing on patch i over night. Then this number X_i is usually described as a Poisson process (see page 222).

Now the information given in the question is that m raindrops in total landed on the roof over night. Armed with this information, we can watch a "rewind" of the night and try to predict how many raindrops landed in a specific patch of the roof. Specifically, we can ask: Where will the next raindrop land? Since the patches behave identically, there is $\frac{1}{n}$ chance that the next raindrop will land in patch 1.

So we see that we are actually filling up n patches of roof with m raindrops, with each raindrop having $\frac{1}{n}$ chance to land in patch 1. So to count the number or raindrops in patch 1, we look at m independent Bernoulli trials (raindrops), each with success probability $\frac{1}{n}$ (chance to land in patch 1). The distribution of the number of successes (raindrops in patch 1) is then *binomial* $\left(m, \frac{1}{n}\right)$.

Answer 4 is correct.

Problem 16

The probability to fail in a very small time interval dt, given you survived so far, is provided by a formula involving the hazard rate $h(t)$ (see page 296):

$$P(T \in dt | T > t) = h(t)dt.$$

The hazard rate is given by the ratio of the density and the survival function (see formula 6 on page 297):

$$h(t) = \frac{f(t)}{G(t)}.$$

Looking up the density function and the c.d.f. for the gamma distribution on page 481, and inserting $r = 2$, we see that

$$f(t) = \frac{\lambda^r}{(r-1)!} t^{r-1} e^{-\lambda t} = \lambda^2 t e^{-\lambda t}$$

and

$$G(t) = 1 - F(t) = \sum_{k=0}^{r-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \sum_{k=0}^{1} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = e^{-\lambda t} (1 + \lambda t).$$

Hence the hazard rate is

$$h(t) = \frac{f(t)}{G(t)} = \frac{\lambda^2 t e^{-\lambda t}}{e^{-\lambda t} (1 + \lambda t)} = \frac{\lambda^2 t}{1 + \lambda t} = \frac{\lambda}{1 + \frac{1}{\lambda t}}$$

and the requested probability is

$$P(T \in dt | T > t) = h(t)dt = \frac{\lambda}{1 + \frac{1}{\lambda t}} dt.$$

Answer 1 is correct.

Problem 17

We are in a situation where it is natural to use the average conditional probability. However, the formulas on page 424 and 425 don't directly cover the case here, where we have a combination of a discrete random variable N and a continuous random variable X.

We can, however, either deduce the required formula from our intuitive understanding of the other cases, or reason as follows:

The average conditional probability of an event B is given by (p. 424):

$$P(B) = \sum_{\text{all } x} P(B | Y = y) P(Y = y).$$

Letting B denote the event $X \in dx$ and substituting y with n, we obtain:

$$P(X \in dx) = \sum_{n=1}^{\infty} P(X \in dx | N = n) P(N = n).$$

We can then divide by dx to get the densities (notice, if you like, the first and last line in the derivation under the box on p. 411):

$$\frac{P(X \in dx)}{dx} = \sum_{n=1}^{\infty} \frac{P(X \in dx | N = n)}{dx} P(N = n)$$

and thus

$$f_X(x) = \sum_{n=1}^{\infty} f_X(x | N = n) P(N = n).$$

This is the required formula for the density of X.

Since X given N is a *gamma*(N, λ) distribution, the conditional density is (p. 481)

$$f_X(x | N = n) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}.$$

Inserting this, and the premise $P(N = n) = (1-p)^{n-1} p$, we obtain:

$$\begin{aligned} f_X(x) &= \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} (1-p)^{n-1} p \\ &= \lambda p e^{-\lambda x} \sum_{n=1}^{\infty} \frac{[(1-p)\lambda x]^{n-1}}{(n-1)!} \\ &= \lambda p e^{-\lambda x} \sum_{n=0}^{\infty} \frac{[(1-p)\lambda x]^{n}}{n!} \\ &= \lambda p e^{-\lambda x} e^{(1-p)\lambda x} \\ &= \lambda p e^{-\lambda p x}. \end{aligned}$$

We used the power series expansion of the exponential function to get rid of the sum:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

Answer 2 is correct.

Problem 18

That the numerically greatest value of the coordinates is less that 1 is the same thing as both coordinates being between -1 and 1.

Since both coordinates are standard normal variables, the probability of each event (1st respectively 2nd coordinate between -1 and 1) is equal to $\Phi(1) - \Phi(-1)$. To obtain the probability of the intersection of these two events, we can just multiply the probabilities together, since the events are independent. Hence the requested probability is $(\Phi(1)-\Phi(-1))^2$.

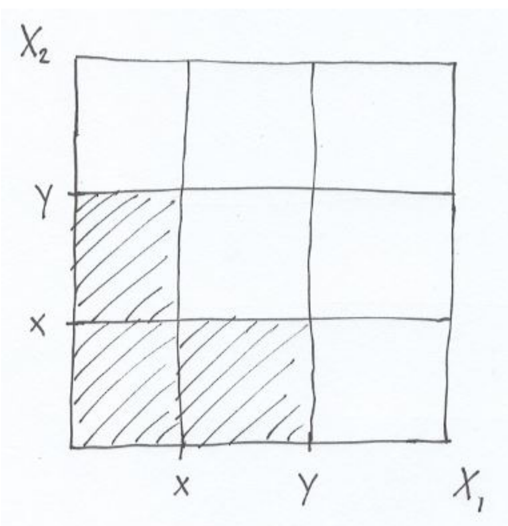
Answer 3 is correct.

Problem 19

An efficient way to get a grasp of an event like

$$[X_{(1)} \leq x, X_{(2)} \leq y] = [\min_i X_i \leq x, \max_i X_i \leq y]$$

is to draw a table divided along lines representing the values x and y.



The shaded regions satisfy that the maximum is lower than y AND that the minimum is lower than x. Now it is quick to see that this total event is equal to the event that both X_1 and X_2 are below y minus the sub-event that they are both between x and y. Using this and expressing things with the independent and identical c.d.f.'s, we obtain:

$$\begin{aligned} F^*(x, y) &= P(X_{(1)} \leq x, X_{(2)} \leq y) \\ &= P(\min_i X_i \leq x, \max_i X_i \leq y) \\ &= P(X_1 \text{ and } X_2 \text{ are both lower than } y) - P(X_1 \text{ and } X_2 \text{ are both between } x \text{ and } y) \\ &= F(y)^2 - (F(y) - F(x))^2. \end{aligned}$$

Answer 1 is correct.

Problem 20

We are asked to find the probability of the first success occurring in the third trial. This event is equal to the event that the first two trials are failures and the third trial is a success. Each trial is independent of the others, so we can multiply the probabilities of failure and success:

$$P(\text{first success in third trial}) = (1-p)(1-p)p = \frac{2}{3}\frac{2}{3}\frac{1}{3} = \frac{4}{27}.$$

Answer 3 is correct.

Problem 21

Since X and Y have standard bivariate normal distribution with correlation $\rho = \frac{3}{5}$, we can rewrite X as follows (p. 451):

$$X = \rho Y + \sqrt{1-\rho^2}Z = \tfrac{3}{5}Y + \tfrac{4}{5}Z.$$

Here, Z is a standard normal variable, and Z and Y are independent.

Inserting this, we obtain:

$$\begin{aligned}
P(\ |X| \leq 1 \mid Y=1) &= P(\ \left|\tfrac{3}{5}Y + \tfrac{4}{5}Z\right| \leq 1 \mid Y=1) \\
&= P(\ \left|\tfrac{3}{5} + \tfrac{4}{5}Z\right| \leq 1 \mid Y=1) \\
&= P(\ \left|\tfrac{3}{5} + \tfrac{4}{5}Z\right| \leq 1) \qquad \text{(since Z and Y are independent)} \\
&= P(-1 \leq \tfrac{3}{5} + \tfrac{4}{5}Z \leq 1) \\
&= P(-\tfrac{8}{5} \leq \tfrac{4}{5}Z \leq \tfrac{2}{5}) \\
&= P(-2 \leq Z \leq \tfrac{1}{2}) \\
&= \Phi\left(\tfrac{1}{2}\right) - \Phi(-2) \\
&= \Phi\left(\tfrac{1}{2}\right) - (1 - \Phi(2)) \\
&= \Phi(2) + \Phi\left(\tfrac{1}{2}\right) - 1.
\end{aligned}$$

Answer 5 is correct.

Problem 22

The random variable Z is defined as the ratio of Y and X, that is, $Z = \frac{Y}{Z}$. In this situation, we can use formula (f) on top of page 383:

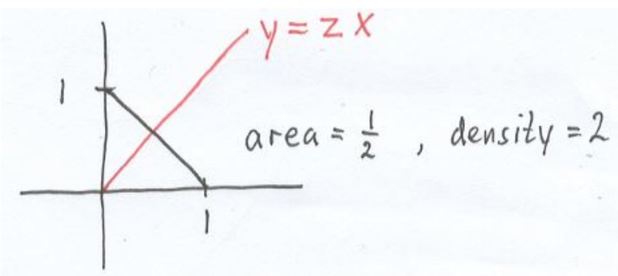
$$f_Z(z) = \int_{-\infty}^{\infty} |x| f(x, zx)\, dx.$$

This formula looks simple, but we have to be quite careful. The tricky thing is to figure out *when* the joint density $f(x, zx)$ evaluates to *what*.

A good way to think about this is to consider z fixed. Then we can phrase the question as "what values of x give what joint density?". And in the case where the joint density is constant (on some region), we can ask "which values of x cause us to be in the region where the density is non-zero?". When we know this, we also know the limits of integration that we should use.

In this problem, we have a uniform distribution on a triangle of area $\frac{1}{2}$. Hence the joint density is 2 whenever we are within this triangle (since they multiply to 1).

So now we can phrase the question like this: "Which values of x cause the point (x, zx) to be within the triangle?"



The points (x, zx) lie on a line through origo with slope z. So in our case, we have to find those x where this line is inside the triangle. So we find the intersection between the line $y = zx$ and the line $y = 1 - x$ from the triangle:

$$\begin{aligned} zx &= 1 - x \\ x &= \frac{1}{1+z} \end{aligned}$$

So, in conclusion: For a fixed $z \geq 0$, whenever x is between 0 and $\frac{1}{1+z}$, the joint density $f(x, zx)$ is 2. (Note that we can assume $z = y/x$ non-negative because X and Y are distributed only on non-negative values.)

Using this, we can finally evaluate the integral:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |x| f(x, zx)\, dx \\ &= \int_{-\infty}^{\infty} |x| \cdot 2 \cdot I_{[0 \leq x \leq \frac{1}{1+z}]}\, dx \\ &= \int_{0}^{\frac{1}{1+z}} 2x\, dx \\ &= x^2 \Big|_{x=0}^{x=\frac{1}{1+z}} \\ &= \frac{1}{(1+z)^2}. \end{aligned}$$

Answer 3 is correct.

Problem 23

The outcome space of X is $\{-3, 0, 2, 4, 7, 10, 12\}$. When this is the case, the event $X^2 < 10$ is equal to the event that $X = -3$ or $X = 0$ or $X = 2$. Or, phrased in a different way, equal to the event $X \in B$, where $B = \{-3, 0, 2\}$.

Since X is discrete, we can add the point probabilities to get the total probability of this event (see p. 141):

$$\begin{aligned} P(X \in B) &= \sum_{x \in B} P(X = x) \\ &= P(X = -3) + P(X = 0) + P(X = 2) \\ &= \tfrac{1}{20} + \tfrac{1}{10} + \tfrac{2}{15} \\ &= \tfrac{17}{60}. \end{aligned}$$

Answer 1 is correct.

Problem 24

X and Y have standard bivariate normal distribution with correlation $\rho = \frac{1}{2}$. Then, according to the "Standard Bivariate Normal Distribution" theorem on p. 451, we can write Y as

$$\begin{aligned} Y &= \rho X + \sqrt{1 - \rho^2} Z \\ &= \tfrac{1}{2} X + \tfrac{\sqrt{3}}{2} Z \end{aligned}$$

where X and Z are *independent* standard normal variables.

We are asked to find the probability that the point (X, Y) lies in the first quadrant between the lines $y = \frac{x}{2}$ and $y = 2x$. Written as inequalities, this is equal to the event $\frac{X}{2} < Y < 2X$ and $X > 0$. Substituting $Y = \frac{1}{2} X + \frac{\sqrt{3}}{2} Z$, we obtain:

$$\begin{aligned} P(\tfrac{1}{2} X < Y < 2X \text{ and } X > 0) &= P(\tfrac{1}{2} X < \tfrac{1}{2} X + \tfrac{\sqrt{3}}{2} Z < 2X \text{ and } X > 0) \\ &= P(Z > 0 \text{ and } Z < \sqrt{3} X \text{ and } X > 0). \end{aligned}$$

As in Example 2 on p. 457, we can now use the rotational symmetry of the joint distribution of X and Z. (The rotational symmetry is due to the fact that X and Z are *independent* standard normal variables.)

The three inequalities correspond to the region in the 1st quadrant under the line through origo with slope $\sqrt{3}$. The angle between this line and the X-axis is $Arctan(\sqrt{3}) = \frac{\pi}{3}$. Due to the rotational symmetry, the probability of landing in this region is given by this angle divided by 2π, so we finally obtain the probability

$$\frac{\frac{\pi}{3}}{2\pi} = \frac{1}{6}.$$

Answer 2 is correct.

Problem 25

This situation is a typical situation where it makes sense to apply a Poisson random scatter, see page 228 and forwards.

We want to apply the Poisson Scatter Theorem (i) on page 230. We first decide on a unit volume of 1000 cubic light years. We are informed that then the parameter (mean) for the Poisson process is $\lambda = 4$. Then, looking at another region B (the box with side length 8 light years), we determine its volume relative to the unit volume. In our case, we have the volume:

$$\frac{8^3}{1000} = \frac{512}{1000}.$$

Now the random scatter theorem tells us that the Poisson parameter λ_B for this region B is

$$\lambda_B = \lambda \cdot \frac{512}{1000} = 4 \cdot \frac{512}{1000} = 2.048.$$

For the number N of relevant planets within B, we can now deduce:

$$\begin{aligned} P(N \geq 1) &= 1 - P(N = 0) \\ &= 1 - \frac{\lambda_B^k}{k!} e^{-\lambda_B} \Big|_{k=0} \\ &= 1 - e^{-\lambda_B} \\ &= 1 - e^{-2.048} \\ &= 0.871. \end{aligned}$$

Answer 2 is correct.

Problem 26

We have 3 independent and identically distributed variables, so we can use the theorem "Density of the kth Order Statistic" on p. 326. (We could also use the specific formula for uniform distributions on p. 327, that would be a bit easier.)

We let one hour be the time unit. Then the c.d.f. and density of the uniform distributions of the arrival times are

$$F(x) = x \quad \text{(for } x \text{ within the hour)}$$
$$f(x) = 1 \quad \text{(for } x \text{ within the hour).}$$

We are looking for the density $g(x)$ of the second smallest of the 3 variables, which translates to $k = 2$ and $n = 3$. Inserting all this in the formula from the theorem, we find that

$$\begin{aligned} g(x) &= nf(x)\binom{n-1}{k-1}(F(x))^{k-1}(1-F(x))^{n-k} \\ &= 3 \cdot 1 \cdot \binom{3-1}{2-1}(x)^{2-1}(1-x)^{3-2} \\ &= 6x(1-x) \\ &= 6(x-x^2) \end{aligned}$$

which applies whenever x is between 0 and 1. We can now integrate this density to obtain the probability $P(A)$ that the second person to arrive arrives after 20 minutes and before 40 minutes (corresponding to 1/3 and 2/3 of an hour):

$$\begin{aligned} P(A) &= \int_{\frac{1}{3}}^{\frac{2}{3}} g(x)\,dx \\ &= \int_{\frac{1}{3}}^{\frac{2}{3}} 6(x-x^2)\,dx \\ &= (3x^2 - 2x^3)\Big|_{x=\frac{1}{3}}^{x=\frac{2}{3}} \\ &= \frac{13}{27}. \end{aligned}$$

Answer 3 is correct.

Problem 27

We are given that two returns X and Y have bivariate normal distribution with

$$\begin{aligned} X &\sim normal(15, 10^2) \\ Y &\sim normal(20, 20^2) \\ \rho &= \frac{-1}{4}. \end{aligned}$$

We are asked to find $P(X + Y < 0)$.

Overall, the strategy to solve this exercise follows 3 main steps:

- Rewrite into 2 *standard* normal variables.
- Rewrite into 2 *independent* standard normal variables.
- Rewrite into 1 normal variable.

We first rewrite X and Y using standardized normal variables X^* and Y^*, cf. box on p. 454:

$$X = \mu_X + \sigma_X X^* = 15 + 10X^*$$
$$Y = \mu_Y + \sigma_Y Y^* = 20 + 20Y^*$$

The standard normal variables X^* and Y^* have the same correlation $\rho = \frac{-1}{4}$ as the normal variables X and Y, according to the box on p. 454.

Using this rewrite, we have

$$\begin{aligned} P(X+Y<0) &= P(15 + 10X^* + 20 + 20Y^* < 0) \\ &= P(10X^* + 20Y^* < -35) \\ &= P(2X^* + 4Y^* < -7). \end{aligned}$$

Since X^* and Y^* are *standardized* bivariate normal variables, we can rewrite Y^* using the formula on p. 451, with X^* and Z^* being *independent* standard normal variables:

$$\begin{aligned} Y^* &= \rho X^* + \sqrt{1-\rho^2} Z^* \\ &= \frac{-1}{4} \cdot X^* + \sqrt{1 - \left(\frac{-1}{4}\right)^2} \cdot Z^* \\ &= \frac{-1}{4} X^* + \frac{\sqrt{15}}{4} Z^*. \end{aligned}$$

Inserting this expression, we obtain

$$\begin{aligned} P(X+Y<0) &= P(2X^* + 4Y^* < -7) \\ &= P(2X^* + 4(\frac{-1}{4}X^* + \frac{\sqrt{15}}{4}Z^*) < -7) \\ &= P(X^* + \sqrt{15}Z^* < -7). \end{aligned}$$

Now, since X^* and Z^* are independent standard normal variables, a linear combination $V = X^* + \sqrt{15}Z^*$ is a normal variable with mean zero and standard deviation given by

$$\sigma_V^2 = 1^2 \cdot 1^2 + (\sqrt{15})^2 \cdot 1^2 = 16 = 4^2.$$

This is according to the formula given on p. 460 (which builds on the result for the variance of a scaling on p. 188 and the theorem about sums of independent normal variables on p. 363).

We can standardize V into V^* by dividing with its SD of 4. Doing this, we finally obtain:

$$\begin{aligned} P(X+Y<0) &= P(X^* + \sqrt{15}Z^* < -7) \\ &= P(V < -7) \\ &= P(V^* < \tfrac{-7}{4}) \\ &= \Phi(\tfrac{-7}{4}). \end{aligned}$$

Answer 3 is correct.

Problem 28

We are given

$$X \sim \text{Poisson}(2)$$
$$P(Y=y|X=x) = \binom{x}{y}\left(\frac{1}{2}\right)^x, \quad y = 0, \ldots, x.$$

We are asked to find $P(E(Y|X) \geq 1)$.

Let us first recognize that, for fixed x, the conditional probability $P(Y=y|X=x) = \binom{x}{y}\left(\frac{1}{2}\right)^x$ actually results in a *binomial*$(x, \frac{1}{2})$ distribution for Y (given $X = x$), since y runs from 0 to x. Hence, the mean of this conditional distribution is $x \cdot \frac{1}{2}$, that is:

$$E(Y|X=x) = \frac{x}{2}.$$

Let us now recall the definition of $E(Y|X)$ on page 402:

"The conditional expectation of Y given X, denoted $E(Y|X)$, is the function of X whose value is $E(Y|X=x)$ if $X = x$."

In our case, this means that the function $E(Y|X)$ of X is the function of X that has value $\frac{x}{2}$ if $X = x$. Or, written briefly:

$$E(Y|X) = \frac{X}{2}.$$

We can now apply the Poisson(2) distribution of X to obtain:

$$\begin{aligned}
P(E(Y|X) \geq 1) &= P(\tfrac{X}{2} \geq 1) \\
&= P(X \geq 2) \\
&= 1 - P(X \leq 1) \\
&= 1 - \sum_{i=0}^{1} \frac{\mu^i}{i!} e^{-\mu} \\
&= 1 - e^{-2}(1+2) \\
&= 1 - \frac{3}{e^2}.
\end{aligned}$$

Answer 2 is correct.

Problem 29

The area of the region is $\frac{9}{2}$, so the joint density in the region is $\frac{2}{9}$, since we have a uniform distribution with constant density in the region.

We can now use the formula for the marginal density given on page 349:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy.$$

To evaluate this integral, we need to figure out the proper limits for y, given a fixed x, so that we are inside the region where the density is actually positive.

The lower limit is 1, no matter the value of x.

The upper limit is on the top line of the region. This top line has equation

$$y = 2 + \tfrac{1}{3}(x-1).$$

So now we can evaluate the integral:

$$\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\, dy \\
&= \int_{1}^{2+\frac{1}{3}(x-1)} \tfrac{2}{9}\, dy \\
&= (2 + \tfrac{1}{3}(x-1) - 1) \cdot \tfrac{2}{9} \\
&= \tfrac{2}{9} + \tfrac{2}{27}(x-1).
\end{aligned}$$

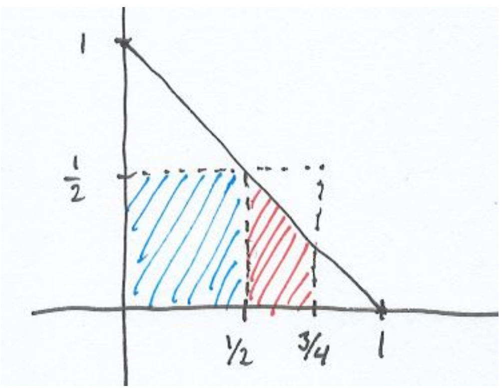
Answer 3 is correct.

Problem 30

We are given the joint density in a region within the first quadrant:

$$f(x,y) = 24xy \quad , \quad 0 < x < 1 - y.$$

The second inequality is equivalent to $y < 1 - x$, so the region is a triangle bounded by the axes and the line $y = 1 - x$.



To find $P(X \leq \frac{3}{4}, Y \leq \frac{1}{2})$, we can integrate the density over this region's intersection with the triangle. This intersection is the blue and red region in the figure.

To make things easier, we split up into two integrals, corresponding to the blue and red parts. We use Maple to evaluate the integrals:

$$P(blue) = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} 24xy\, dy\, dx = \frac{3}{8}.$$

and

$$P(red) = \int_{\frac{1}{2}}^{\frac{3}{4}} \int_0^{1-x} 24xy\, dy\, dx = \frac{67}{256}.$$

Adding these, we obtain $\frac{163}{256}$.

Answer 5 is correct.