IMM - DTU

02405 Probability
2024-20-11
Hannah Nielsen

These are suggested solutions and explanations for the December 2023 exam in the course 02405 *Sandsynlighedsregning* at DTU. Page references are to the book *Probability* by Jim Pitman.

Problem 1

Let B denote the event that there is a beetle attack, and let M denote the event that the coffee bean is discolored.

We can use Bayes' Theorem (p. 49). We are given the prior probabilities of a beetle attack and not beetle attack

$$P(B) = 0.001$$
$$P(B^{\complement}) = 1 - 0.001 = 0.999$$

and the likelihoods of a discoloration given a beetle attack or not beetle attack

$$P(M|B) = 0.8$$
$$P(M|B^{\complement}) = 0.01.$$

We can calculate the probability of discoloring, by

$$\begin{aligned} P(M) &= P(M|B)P(B) + P(M|B^{\complement})P(B^{\complement}) \\ &= 0.8 \cdot 0.001 + 0.01 \cdot 0.999. \end{aligned}$$

Inserting into Bayes' formula to find the posterior probability of a beetle attack given discolored beans, we obtain:

$$\begin{aligned} P(B|M) &= \frac{P(M|B)P(B)}{P(M)} \\ &= \frac{0.8 \cdot 0.001}{0.8 \cdot 0.001 + 0.01 \cdot 0.999} \\ &= 0.074 \end{aligned}$$

Answer 4 is correct.

Problem 2

Let EB be the event customer buys an electric car which happens with probability $p_{EB} = 0.45$, BB be the event customer buys a petrol car which happens with probability $p_{BB} = 0.15$ and IB the event that no car is bought with $p_{IB} = 0.40$.

Let n be the number of customers on a day and N_{EB} be the number customers on a day who buys electric cars and similarly for N_{BB} and N_{IB}. If $n = 5$ and we want to find the probability of observing exactly $N_{EB} = 2, N_{BB} = 1, N_{IB} = 2$ the multinomial distribution can be used(p. 155):

$$\begin{aligned} P(N_{EB} = 2, N_{BB} = 1, N_{IB} = 2) &= \frac{5!}{2!1!2!}(0,45)^2(0,15)^1(0,40)^2 \\ &= \frac{5!}{2!2!}(0.45)^2(0.15)(0.40)^2. \end{aligned}$$

Answer 3 is correct.

Problem 3

Let X denote maximal daily wave height. Since we know the expected value $E(X) = 2$ and the standard deviation of X, $Var(X) = 1 \leftarrow SD(X) = 1$ as well as a bounding probability, we can try using Chebychev's Inequality (p. 191):

$$P[|X - E(X)| \geq kSD(X)] \leq \frac{1}{k^2}$$

We want to find the probability the on a given day the maximal wave height is over 6 meter, so the dike is flooded. Note that since $X \geq 0$ and $E(X) = 2$ it holds that $X - E(X) \geq -2$ and there is no probability mass at -4 or below so taking the absolute is allowed. Hereby the bound can be found as;

$$\begin{aligned} P(X \geq 6) &= P(X - E(X) \geq 6 - E(X)) \\ &= P(X - E(X) \geq 4) \\ &= P(|X - E(X)| \geq 4 \cdot 1) \\ &= P(|X - E(X)| \geq 4SD(X)) \leq \frac{1}{4^2} \\ &= \frac{1}{16} \end{aligned}$$

So in total we can conclude $P(X \geq 6) \leq \frac{1}{16}$. Answer 5 is correct.

Problem 4

Let X denote stock-rate 1 and Y stock-rate 2. 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 5 is correct.

Problem 5

X_1 and X_2 are independent with $P(X_i \leq x) = F(x),\ i = 1,2$ and $X_{(1)} = \min X_i,\ X_{(2)} = \max X_i$. For $x \leq y$ we have the joint distribution function as $F^*(x,y) = P(X_{(1)} \leq x, X_{(2)} \leq y)$ which we want to express in terms of $F(x)$ and $F(y)$. Because it is only the probability in the cases where $y \geq$ both X_i's and at least one of them must be greater or equal x so

$$\begin{aligned} F^*(x,y) =& P(X_{(1)} \leq x, X_{(2)} \leq y) \\ =& P(x < X_1 \leq y, X_2 \leq x) + P(X_1 \leq y, x < X_2 \leq x) + P(X_1 \leq x, X_2 \leq x) \\ =& (F(y) - F(x))F(x) + (F(y) - F(x))F(x) + F(x)^2 \\ =& (F(y) - F(x) + F(y) - F(x) + F(x))F(x) \\ =& (2F(y) - F(x))F(x) \\ =& 2F(y)F(x) - F(x)^2 + F(y)^2 - F(y)^2 \\ =& F(y)^2 - (F(y) - F(x))^2. \end{aligned}$$

Alternatively it is reached more easily by;

$$\begin{aligned} P(X_{(1)} \leq x, X_{(2)} \leq y) &= P(X_1 \leq y, X_2 \leq y) - P(x < X_1 \leq y, x < X_2 \leq y) \\ &= F(y)^2 - (F(y) - F(x))^2 \end{aligned}$$

Answer 1 is correct.

Problem 6

This situation is a typical situation where it makes sense to apply a Poisson random scatter, see page 228 and forwards. We want to determine $P(Q_2 \leq 1)$ which is the probability of at most finding 1 eastern Quoll in an area of 2 km^2. The appearance is a Poisson process where we are informed that $Q_1 \sim Pois(\lambda = 3)$ on an area of 1 km^2(p.121) and hereby via Poisson Scatter Theorem (p. 230) we get that for an area of i km^2 we have the appearance described by $Q_i \sim Pois(\lambda_i = 3i)$.

$$\begin{aligned} P(Q_2 \leq 1) &= e^{-3\cdot 2}(3 \cdot 2)^0/0! + e^{-3\cdot 2}(3 \cdot 2)^1/1! \\ &= e^{-3\cdot 2}(1+6) = e^{-6}7 \end{aligned}$$

Answer 3 is correct.

Problem 7

$G(x)$ is the survival function for $X \geq 0$. $G(x) = P(X > x) = 1 - (X \leq x) = 1 - F(x)$. G must be between 0 and 1 and must not increase.

Answer 2 is correct.

Problem 8

We have exponentially distributed time intervals between particle arrivals with a mean of $\mu_i = 0.2$s, giving $\sigma = 0.2$. The start time is t_0 but from the memoryless property can this be disregarded. Each interval T_i is independent of the rest. We want to find $P(\sum_{i=1}^{10000} T_i < 33min) = P(\sum_{i=1}^{10000} T_i < 33 \cdot 60s)$. Since there is a large number of component lifetimes that are independent and identically distributed, the central limit theorem ("The Normal Approximation", p. 196) applies. So the sum $S_{10000} \sim \mathbb{N}(10000 \cdot \mu, (\sigma\sqrt{10000})^2)$.

$$\begin{aligned} P(\sum_{i=1}^{10000} T_i < 33 \cdot 60) =& 1 - P(S_{10000} \geq 1980) \\ &= 1 - P(\frac{S_{10000} - 10000 \cdot 0.2}{0.2\sqrt{10000}} \geq \frac{1980 - 10000 \cdot 0.2}{0.2\sqrt{10000}}) \\ &= 1 - P(\frac{S_{10000} - 10000 \cdot 0.2}{0.2\sqrt{10000}} \geq -1) \\ &\approx 1 - (1 - \Phi(-1)) \text{ from symmetry of standard normal distribution} \\ &= 1 - \Phi(1) \end{aligned}$$

Answer 1 is correct.

Problem 9

Let S be the event of sunshine, V the event that it's a winter day and F the event that it is freezing. Then we are told $P(S|V) = 0.45$ and $P(S, F|V) = 0.35$.

We want to determine $P(S, F^{\complement}|V)$. Which we can find by rules of partition (p. 21).

$$P(S|V) = P(S, F^{\complement}|V) + P(S, F|V) \leftarrow P(S, F^{\complement}|V) = P(S|V) - P(S, F|V) = 0.45 - 0.35$$

Answer 5 is correct.

Problem 10

Claim: only 5% too low concentration. 20 pills are chosen at random, where 3 of the 20 pills have a too low concentration. If $p_l = 0.05$ we need to find $P(N_l \geq 3)$, where N_l is the number of pills with too low concentration out of $n_l = 20$, use the binomial distribution (p. 81):

$$\begin{aligned} P(N_l \geq 3) &= \sum_{i=3}^{20} \binom{20}{i} p_l^i (1 - p_l)^{20-i} \\ &= \sum_{i=3}^{20} \binom{20}{i} 0,05^i (0,95)^{20-i} \end{aligned}$$

Answer 4 is correct.

Problem 11

An area is limited by $y = 3, y = 0, y = \frac{3}{2}(x + 3), y = -\frac{3}{2}(x - 3)$. The first coordinates to the point is denoted by the stochastic variable X. We are asked to find $P(X \geq 1)$. Firstly the boundaries of the right triangle, which is described by the line $y = -\frac{3}{2}(x - 3)$, is found

$$\begin{aligned} 3 &= -\frac{3}{2}(x - 3) => x = -2 + 3 = 1 \\ 0 &= -\frac{3}{2}(x - 3) => x = 3 \end{aligned}$$

Therefore the triangle starts in $x = 1$ and ends in 3. So we can determine the bottom of the triangle to be 2 wide and from symmetry we must get the density normalization constant c

as

$$\begin{aligned}\frac{1}{2} &= P(0 \le X \le 3)\\ &= P(0 \le X \le 1) + P(1 \le X \le 3)\\ &= (1 \cdot 3 + \frac{1}{2}(2 \cdot 3))c\\ &= 3c + 3c = 6c\\ => c &= 1/12\end{aligned}$$

Now to find $P(X \ge 1)$ we now know that this is the area of the right triangle normalized with c i.e.

$$P(X \ge 1) = \frac{c}{2}(2 \cdot 3) = \frac{3}{12} = \frac{1}{4}.$$

Answer 2 is correct.

Problem 12

So since the calls come randomly with a mean of λ calls pr. minute, we can assume that calls come as a Poisson arrival process and as described in the box on p. 284 the time between arrivals/calls then follow an exponential distribution with rate λ. Answer 5 is correct.

Problem 13

Given $f(x, y) = 6(x - y)$, $X = max(U_i)$, $Y = min(U_i)$, $i = 1, 2, 3$ and $U_i \sim Unif(0, 1)$, where U_i independent. We are asked to determine $E(Y|X = x)$. Ref p. 423 on "Conditional Expectations" and p. 349 on "Joint Distribution Defined by a Density" and

$$\begin{aligned}E(Y|X = x) &= E(min(U_i)|max(U_i) = x)\\ &= \int_0^x y f_{Y|X}(y|X = x)dy\\ &= \int_0^x y f(y, x)/f_X(x)dy\\ &= \int_0^x y6(x - y)/f_X(x)dy\\ &= \frac{6}{f_X(x)} \int_0^x y(x - y)dy\\ &= \frac{6}{f_X(x)} \left[xy^2/2 - y^3/3\right]_0^x\\ &= \frac{6}{f_X(x)}(xx^2/2 - x^3/3 - (x0^2/2 - 0^3/3))\\ &= \frac{3x^3 - 2x^3}{f_X(x)}\end{aligned}$$

We can determine the marginal distribution as

$$\begin{aligned} f_X(x) &= \int_0^x f(x,y)dy = \int_0^x 6(x-y)dy \\ &= 6\left[xy - y^2/2\right]_0^1 = 6x^2 - 3x^2 \\ &= 3x^2 \end{aligned}$$

Lets insert the expression found for $f_X(x)$ to get the mean of Y given X.

$$E(Y|X=x) = \frac{x^3}{f_X(x)} = \frac{x^3}{3x^2} = \frac{x}{3}$$

Answer 2 is correct.

Problem 14

X: $f_x(x) = 2x$ for $x \in [0;1]$ $Y = X^2$: $f_Y(y)$ y must also lie in [0;1] We are in a situation with a change of variable, given by $Y = X^2$. 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

$$\left|\frac{dy}{dx}\right| = \left|\frac{d(x^2)}{dx}\right| = |2x| = 2x.$$

We notice that $2x$ will cancel out the $2x$ in the numerator, so there is no need to express it in terms of y.

Inserting, we obtain:

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

Answer 3 is correct.

Problem 15

The situation described is exactly a Geometric distribution with $p = 1/20$ (p. 482) and $T = 3$. Therefor the probability of the success occurs in exactly the 3. try is

$$P(T=3) = (1-p)^{3-1}p = (19/20)^2 1/20 = 19^2/20^3.$$

Answer 5 is correct.

Problem 16

Let O be the event of an overrun water-plant, W the event of a powerful rainfall, and J is a random day in July. We are the given the following probabilities;

$$\begin{aligned}
P(W|J) &= \frac{1}{9} \\
P(W^{\complement}|J) &= 1 - \frac{1}{9} = \frac{8}{9} \\
P(O|W) &= \frac{1}{4} \\
P(O|W^{\complement}) &\approx 0
\end{aligned}$$

Then by "Rule of Average Conditional Probabilities" on p. 396 we get

$$P(O|J) = P(O|W,J)P(W|J) + P(O|W^{\complement},J)P(W^{\complement}|J) = \frac{1}{36}$$

Answer 3 is correct.

Problem 17

Given the joint density function:

$$f(x,y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), \quad 0 < x < 1, \quad 0 < y < 2$$

We need to find $P(Y > \frac{1}{2} \mid X < \frac{1}{2})$. To do this we need to find the conditional density function $f_{Y|X}(y|x)$, which is $f(x,y)/f_X(x)$ by Multiplication Rule (p. 151). So firstly we need to find the marginal of X by $\int_y f(x,y)dy$ (p. 349);

$$\begin{aligned}
f_X(x) &= \int_0^2 \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) dy \\
&= \frac{6}{7}\left[x^2 y + \frac{xy^2}{4}\right]_0^2 \\
&= \frac{6}{7}\left(2x^2 + x\right).
\end{aligned}$$

With the marginal we can now find the conditional density function to

$$\begin{aligned}
f_{Y|X}(y|x) &= \frac{f(x,y)}{f_X(x)} \\
&= \frac{\frac{6}{7}\left(x^2 + \frac{xy}{2}\right)}{\frac{6}{7}(2x^2 + x)} \\
&= \frac{x^2 + \frac{xy}{2}}{2x^2 + x}.
\end{aligned}$$

The conditional probability $P(Y > \frac{1}{2}|X < \frac{1}{2})$ can now be found be integrating y from 1/2 to 2 and x from 0 to 1/2.

$$\begin{aligned}
P(Y > \frac{1}{2} \mid X < \frac{1}{2}) &= \int_0^{1/2} \int_{1/2}^2 \frac{x^2 + \frac{xy}{2}}{2x^2 + x} dy dx \\
&= \int_0^{1/2} \int_{1/2}^2 \frac{x^2 + \frac{xy}{2}}{2x^2 + x} dy\, dx \\
&= \int_0^{1/2} (\int_{1/2}^2 \frac{x^2}{2x^2 + x} dy + \int_{1/2}^2 \frac{\frac{xy}{2}}{2x^2 + x} dy) dx \\
&= \int_0^{1/2} \left(\frac{3x^2}{2(2x^2 + x)} + \frac{15x}{16(2x^2 + x)} \right) dx
\end{aligned}$$

Using a computational tool to evaluate the final integral, we get:

$$P(Y > \frac{1}{2} \mid X < \frac{1}{2}) \approx 0.8625$$

Answer 2 is correct.

Problem 18

Inner bull in radius interval [0;0.25] outer bull with radius in the interval [0.25,1.25/2]. The coordinates hit (X, Y) are two independent normal distributed stochastic variables with mean 0 and variance 1. We have to find the probability of hitting the outer bull. To do this we can use the Rayleigh distribution (p. 358-359). So the probability must be

$$\begin{aligned}
P(0.25 < r < 1.25/2) &= F(1.25/2) - F(0.25) \\
&= 1 - e^{-\frac{1}{2}(0.625)^2} - (1 - e^{\frac{1}{2}(0.25)^2}) \\
&= e^{\frac{1}{2 \cdot 16}} - e^{-\frac{1}{2}\frac{25}{64}} \\
&= e^{\frac{1}{32}} - e^{-\frac{25}{128}}.
\end{aligned}$$

Answer 1 is correct.

Problem 19

For a joint density function $f(x, y) = 2$ for $0 < x < y < 1$ and 0 else we want to find the covariance. We want to use the alternative formula on p.430. Therefore, we first need to find both marginal distributions and their means;

$$\begin{aligned} f_X(x) &= \int_x^1 2dy && = 2(1-x) \\ E(X) &= \int_0^1 x(2-2x)dx && = 1^2 - \frac{2}{3}1^3 = \frac{1}{3} \\ f_Y(y) &= \int_0^y 2dx && = 2y \\ E(Y) &= \int_0^1 2y^2 dy && = \frac{2}{3}(1-0^2) = \frac{2}{3}. \end{aligned}$$

Then we also need the joint mean;

$$E(XY) = \int_0^1 \int_0^y xyf(x,y)dxd = \int_0^1 \int_0^y xy2dxdy = \int_0^1 2y(\frac{1}{2}y^2)dy = \frac{1}{4}$$

Now the covariance can be computed from the formula.

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{4} - \frac{1}{3}\frac{2}{3} = \frac{9-8}{36} = \frac{1}{36}.$$

Answer 3 is correct.

Problem 20

The probability that N of the 1000 go to the one theater can be described with a binomial distribution with $p = \frac{1}{2}$;

$$P(arrivalsinonebio = N) = \binom{1000}{N}\frac{1}{2}^N \frac{1}{2}^{1000-N} = \binom{1000}{N}\frac{1}{2}^{1000}$$

This can then aalso be extended to get the probability that the number of arrivals is above N and some arrivals are rejected;

$$\begin{aligned} 0.01 &< P(\text{refuse entrance}) \\ &= P(\text{arrivals in one bio} > N) \\ &= 1 - P(\text{arrivals in one bio} \leq N) \\ &= (\frac{1}{2})^{1000} \sum_{i=0}^{N} \binom{1000}{i}. \end{aligned}$$

We might also consider using the "Normal Approximation to the Binomial Distribution" on p. 99 since the exponent is so big. In this case, we would get;

$$
\begin{aligned}
0.01 &< P(\text{refuse entrance}) \\
&= P(\text{arrivals in one bio} > N) \\
&= 1 - P(\text{arrivals in one bio} \leq N) \\
&\approx 1 - (\Phi(\frac{N + \frac{1}{2} - 1000\frac{1}{2}}{\sqrt{1000\frac{1}{2}\frac{1}{2}}}) - \Phi(\frac{0 - \frac{1}{2} - 1000\frac{1}{2}}{\sqrt{1000\frac{1}{2}\frac{1}{2}}})) \\
&= 1 - (\Phi(\frac{\frac{1}{2}(2N - 999)}{\sqrt{250}}) - \Phi(\frac{-\frac{1}{2}(1001)}{\sqrt{250}})) \\
&\approx 1 - (\Phi(\frac{N - 499.5}{\sqrt{250}}) - 0).
\end{aligned}
$$

Answer 2 is correct.

Problem 21

We have that $X \sim Bin(6, \frac{1}{2})$ and $Y = |X - 3|$ Since $X \in [0; 6]$ when $X < 3$, then $X - 3 < 0$ and they will be mirrored. When $X = 6$ then $Y = 3$. So Y can take the values $0, 1, 2, 3$;

$$P(Y = 0) = P(X = 3) = \binom{6}{3}\frac{1}{2}^6 = \frac{20}{64} = \frac{5}{16}$$

$$P(Y = 1) = P(X = 2) + P(X = 4) = \binom{6}{2}\frac{1}{2}^6 + \binom{6}{2}\frac{1}{2}^6 = 2\frac{15}{64} = \frac{15}{32}$$

$$P(Y = 2) = P(X = 1) + P(X = 5) = \binom{6}{1}\frac{1}{2}^6 + \binom{6}{5}\frac{1}{2}^6 = 2\frac{6}{64} = \frac{3}{16}$$

$$P(Y = 3) = P(X = 0) + P(X = 6) = \binom{6}{0}\frac{1}{2}^6 + \binom{6}{6}\frac{1}{2}^6 = 2\frac{1}{64} = \frac{1}{32}$$

Answer 3 is correct.

Problem 22

$X_1 + X_2 + X_3 = 20$, $P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{20!}{x_1!x_2!x_3!}(\frac{1}{2})^{x_1}(\frac{1}{5})^{x_2}(\frac{3}{10})^{x_3}$ It is observed that $X_1 = 10$ Find $P(X_2 = x_2, X_3 = x_3 | X_1 = 10)$ $P(X_2 = x_2, X_3 = x_3 | X_1 = 10) = P(X_1 = 10, X_2 = x_2, X_3 = x_3)/P(X_1 = 10)$ Lets observe that stochastic variable (X_1, X_2, X_3) follows a multinomial distribution (p. 155) and we know that the marginal of a multinomial is a binomial (exercise 3.1.12) with for X_1 the parameters $n_1 = 20$ $p_1 = \frac{1}{2}$ $k_1 = 10$ giving

$$P(X_1 = 10) = \binom{20}{10}\frac{1}{2}^{20}$$

$$
\begin{aligned}
P(X_2 = x_2, X_3 = x_3 | X_1 = 10) &= \frac{P(X_1 = 10, X_2 = x_2, X_3 = x_3)}{P(X_1 = 10)} \\
&= \frac{\frac{20!}{10! x_2! x_3!} (\frac{1}{2})^{10} (\frac{1}{5})^{x_2} (\frac{3}{10})^{x_3}}{\frac{20!}{10!10!} \frac{1}{2}^{20}} \\
&= \frac{\frac{1}{x_2! x_3!} (\frac{1}{5})^{x_2} (\frac{3}{10})^{x_3}}{\frac{1}{10!} \frac{1}{2}^{10}} \\
&= \frac{\frac{10!}{x_2! x_3!} \frac{1}{5^{x_2}} (\frac{3}{10})^{x_3}}{2^{-10}} \\
&= \binom{10}{x_2!} \frac{2^{x_2 + x_3}}{5^{x_2}} (\frac{3}{10})^{x_3} \\
&= \binom{10}{x_2!} (\frac{2}{5})^{x_2} (\frac{2 \cdot 3}{10})^{x_3} \\
&= \binom{10}{x_2!} (\frac{2}{5})^{x_2} (\frac{3}{5})^{x_3}
\end{aligned}
$$

Answer 4 is correct.

Problem 23

$X \sim Unif(1, 3)$ and $Y \sim Unif(2, 4)$ the densities of X and Y can be found as on p. 264

$$
\begin{aligned}
f_X(x) = \frac{1}{3-1} &= \frac{1}{2} \text{ if } 1 \le x \le 3 \text{ else } 0 \\
f_Y(y) = \frac{1}{4-2} &= \frac{1}{2} \text{ if } 2 < y < 4 \text{ else } 0.
\end{aligned}
$$

From independence, we have that the joint is the product of the marginals

$$
f(x, y) = f_X(x) f_Y(y) = \frac{1}{2}\frac{1}{2} = \frac{1}{4},
$$

if $1 \le x \le 3$ and $2 < y < 4$ else $f(x, y) = 0$ Answer 1 is correct.

Problem 24

Let R be a stochastic variable which follows a Rayleigh distribution. To find the median use the c.d.f. of the Rayleigh on p. 359 and the median is found by $r = F^{-1}(\frac{1}{2})$ (p.319),

$$\begin{aligned}
&\frac{1}{2} = F_R(r) = 1 - e^{-\frac{1}{2}r^2} \\
=>&\frac{1}{2} = e^{-\frac{1}{2}r^2} \\
=>&\ln\frac{1}{2} = -\frac{1}{2}r^2 \\
=>& -2\ln\frac{1}{2} = r^2 \\
=>&\sqrt{-2\ln 1 + 2\ln 2} = r \\
=>&r = \sqrt{2\ln 2}.
\end{aligned}$$

Answer 1 is correct.

Problem 25

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:

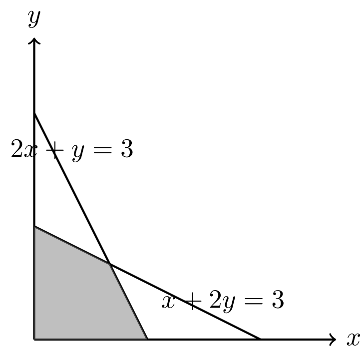
$$f_Z(z) = \int_{-\infty}^{\infty} |x| f(x, zx)\, dx.$$

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 trapeze of area $\frac{3}{2}$. Hence, the joint density is $\frac{2}{3}$ whenever we are within this trapeze (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 trapeze?"



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 trapeze. By visual inspection it can be seen that we have two intersecting lines one when $z \leq 1$ and one when $z \leq 1$. First we consider the case where $z \leq 1$. So we find the intersection between the line $y = zx$ and the line $2x + y = 3 => y = 3 - 2x$:

$$zx = 3 - 2x$$
$$x = \frac{3}{2+z}$$

So, for a fixed $1 \geq z \geq 0$, whenever x is between 0 and $\frac{3}{2+z}$, the joint density $f(x, zx)$ is $2/3$. (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 evaluate the integral:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |x| f(x, zx)\, dx \\ &= \int_{-\infty}^{\infty} |x| \cdot \frac{2}{3} \cdot I_{[0 \leq x \leq \frac{3}{2+z}]}\, dx \\ &= \int_0^{\frac{3}{2+z}} \frac{2}{3} x\, dx \\ &= \frac{1}{3} x^2 \Big|_{x=0}^{x=\frac{3}{2+z}} \\ &= \frac{3}{3(2+z)^2} \\ &= \frac{1}{(2+z)^2}. \end{aligned}$$

Now we consider the case where $z \geq 1$. So we find the intersection between the line $y = zx$ and the line $x + 2y = 3 => y = \frac{3-x}{2}$:

$$zx = \frac{3-x}{2}$$
$$x = \frac{3}{2z+1}$$

So, for a fixed $z \geq 1$, whenever x is between 0 and $\frac{3}{2z+1}$, the joint density $f(x, zx)$ is $2/3$.

Using this, we can evaluate the integral:

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} |x| f(x, zx)\, dx \\
&= \int_{-\infty}^{\infty} |x| \cdot \frac{2}{3} \cdot I_{[0 \leq x \leq \frac{3}{2z+1}]}\, dx \\
&= \int_{0}^{\frac{3}{2z+1}} \frac{2}{3} x\, dx \\
&= \frac{1}{3} x^2 \Big|_{x=0}^{x=\frac{3}{2z+1}} \\
&= \frac{3}{3(2z+1)^2} \\
&= \frac{1}{(2z+1)^2}.
\end{aligned}$$

Answer 4 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.

Then the c.d.f. and density of the Rayleigh distributions of wind speeds are

$$\begin{aligned}
F(x) &= 1 - e^{-\frac{1}{2}x^2} \quad x > 0 \\
f(x) &= x e^{-\frac{1}{2}x^2} \quad x > 0.
\end{aligned}$$

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

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

which applies whenever $x > 0$. Answer 4 is correct.

Problem 27

Let the pair (X, Y) be bivariate normal distributed with $E(X) = 0$, $E(Y) = \mu_Y$, $Var(X) = \sigma_X^2$, $Var(Y) = \sigma_Y^2$ and $Cov(X, Y) = \rho \sigma_X \sigma_Y$ $P(y_1 \leq Y \leq y_2 | X = x)$

Define $U = (X - E(X))/\sigma_X = X/\sigma_X$ and $V = (Y - E(Y))/\sigma_Y = (Y - \mu_Y)/\sigma_Y$ so that U, V have a standard bivariate normal distribution with correlation ρ (p. 454).

Now we can define $V = \rho U + \sqrt{1-\rho^2}W$ where U and W are two independent standard normal variables

$$\begin{aligned}
P(y_1 \le Y \le y_2 | X = x) =& P(\frac{y_1 - \mu_Y}{\sigma_Y} \le V \le \frac{y_2 - \mu_Y}{\sigma_Y} | U = x/\sigma_X) \\
&= P(\frac{y_1 - \mu_Y}{\sigma_Y} \le \rho U + \sqrt{1-\rho^2}W \le \frac{y_2 - \mu_Y}{\sigma_Y} | U = x/\sigma_X) \\
&= P(\frac{\frac{y_1-\mu_Y}{\sigma_Y} - \frac{\rho x}{\sigma_X}}{\sqrt{1-\rho^2}} \le W \le \frac{\frac{y_2-\mu_Y}{\sigma_Y} - \frac{\rho x}{\sigma_X}}{\sqrt{1-\rho^2}}) \\
&= \Phi(\frac{\frac{y_2-\mu_Y}{\sigma_Y} - \frac{\rho x}{\sigma_X}}{\sqrt{1-\rho^2}}) - \Phi(\frac{\frac{y_1-\mu_Y}{\sigma_Y} - \frac{\rho x}{\sigma_X}}{\sqrt{1-\rho^2}}) \\
&= \Phi(\frac{y_2 - \mu_Y - \frac{\rho x \sigma_Y}{\sigma_X}}{\sigma_Y \sqrt{1-\rho^2}}) - \Phi(\frac{y_1 - \mu_Y - \frac{\rho x \sigma_Y}{\sigma_X}}{\sigma_Y \sqrt{1-\rho^2}})
\end{aligned}$$

Answer 4 is correct.

Problem 28

Let X be a stochastic variable which follows a $beta(2, 1)$ and Y be one which given $X = x$ follows a $binomial(4, x)$. We are asked to find $P(Y = 3)$ Then by "Integral Conditioning Formula" on p. 417 we get

$$\begin{aligned}
P(Y = 3) = \int_x P(Y = 3|X = x) f_X(x) dx = \int_0^1 \binom{4}{3} x^3 (1-x)^{4-3} \cdot \frac{(2+1-1)!}{(2-1)!(1-1)!} x^{2-1}(1-x)^{1-1} dx \\
= 2\binom{4}{3}[\frac{1}{5}x^5 - \frac{1}{6}x^6]_0^1 \\
= 8(\frac{1}{5} - \frac{1}{6}) \\
= \frac{8}{30} \\
= \frac{4}{15}
\end{aligned}$$

Answer 1 is correct.

Problem 29

We are given 4 independent which all can be described by $F(X) = 1 - \exp -(x-a)b$ for $x \ge a$ We need to determine the survival of $2a$ for the minimum i.e. $1 - F_{min}(2a)$ we can use the results on p. 319. We are looking for the c.d.f. $F_{min}(x)$ of the minimum of the 4 variables. Inserting all this in the formula, we find that

$$\begin{aligned} F_{min}(x) &= 1 - (1 - F(x))^4 \\ &= 1 - (1 - 1 + \exp -(x-a)b)^4 \\ &= 1 - (\exp -(x-a)b)^4 \end{aligned}$$

Now we can determine the probability that the public institution must pay at least $2a$,

$$\begin{aligned} P(X_{min} \geq 2a) &= 1 - F_{min}(2a) \\ &= 1 - (1 - (\exp -(2a-a)b)^4) \\ &= \exp -ab^4 \\ &= \exp -4ab. \end{aligned}$$

Answer 2 is correct.

Problem 30

We are given that the number of X (20 foot containers) and Y (40 foot containers) have bivariate normal distribution with

$$\begin{aligned} X &\sim normal(1000, 100^2) \\ Y &\sim normal(500, 50^2) \\ \rho &= \frac{-4}{5}. \end{aligned}$$

We are asked to find $P(33X + 66Y > 69000)$.

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:

$$\begin{aligned} X &= \mu_X + \sigma_X X^* = 1000 + 100X^* \\ Y &= \mu_Y + \sigma_Y Y^* = 500 + 50Y^* \end{aligned}$$

The standard normal variables X^* and Y^* have the same correlation $\rho = \frac{-4}{5}$ as the normal variables X and Y, according to the box on p. 454.

Using this rewrite, we have

$$\begin{aligned}P(33X+66Y>69000) &= P(33000+3300X^*+33000+3300Y^*>69000)\\ &= P(3300(X^*+Y^*)>3000)\\ &= P(X^*+Y^*>\frac{10}{11}).\end{aligned}$$

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:

$$\begin{aligned}Y^* &= \rho X^* + \sqrt{1-\rho^2}Z^*\\ &= \frac{-4}{5}\cdot X^* + \sqrt{1-\left(\frac{-4}{5}\right)^2}\cdot Z^*\\ &= \frac{-4}{5}X^* + \frac{3}{5}Z^*.\end{aligned}$$

Inserting this expression, we obtain

$$\begin{aligned}P(33X+66Y>69000) &= P(X^*+Y^*>\frac{10}{11})\\ &= P(X^*+\frac{-4}{5}X^*+\frac{3}{5}Z^*>\frac{10}{11})\\ &= P(X^*+3Z^*>\frac{50}{11}).\end{aligned}$$

Now, since X^* and Z^* are independent standard normal variables, a linear combination $V = X^* + 3Z^*$ is a normal variable with mean zero and standard deviation given by

$$\sigma_V^2 = 1^2\cdot 1^2 + 3^2\cdot 1^2 = 10.$$

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 $\sqrt{10}$. Doing this, we finally obtain:

$$\begin{aligned}P(33X+66Y>69000) &= P(X^*+3Z^*>\frac{50}{11})\\ &= P(V>\frac{50}{11})\\ &= P(V^*>\tfrac{50}{11\sqrt{10}})\\ &= 1-\Phi(\tfrac{5\sqrt{10}}{11})\\ &= \Phi(-\tfrac{5\sqrt{10}}{11}).\end{aligned}$$

Answer 1 is correct.