

These are suggested solutions and explanations to the exam December 2019. Page references are to the book Probability by Jim Pitman.

Problem 1

Let X denote the number of persons from the ethnic group in the sample.

The situation is sampling without replacement, and thus X follows a hypergeometric distribution. However, since the population is big compared to the sample size $n = 100$, it is reasonable to regard the sampling as a series of Bernoulli trials with fixed parameter $p = 0.05$. We can then approximate the hypergeometric distribution with a binomial distribution, cf. the discussion on p. 127.

The probability of getting at least 7 persons from the group is then given by

$$\begin{aligned} P(X \geq 7) &= 1 - P(X \leq 6) \\ &= 1 - \sum_{i=0}^6 \binom{100}{i} (0.05)^i (1 - 0.05)^{100-i} \end{aligned}$$

which is answer 4.

We might also consider using the "Normal Approximation to the Binomial Distribution" on p. 99. In this case, we would get

$$\begin{aligned} P(X \geq 7) &= 1 - \Phi\left(\frac{7 - 0.5 - np}{\sqrt{npq}}\right) \\ &= 1 - \Phi\left(\frac{7 - 0.5 - 100 \cdot 0.05}{\sqrt{100 \cdot 0.05 \cdot 0.95}}\right) \\ &= 1 - \Phi\left(\frac{3}{\sqrt{19}}\right) \end{aligned}$$

which is not one of the options.

Answer 4 is correct.

Problem 2

Let A denote the event that the target is destroyed. We are given the conditional probability

$$P(A|R = r) = e^{-r^2}$$

and the density

$$f_R(r) = r e^{-\frac{1}{2}r^2}$$

for the distance R to the center.

Intuitively, we cannot have a negative distance to the center, so the density applies only for $r > 0$. Alternatively, we can recognize the given density as the density of the Rayleigh distribution (p. 359), which we know integrates to 1 for $r > 0$. Hence the given density must apply only for $r > 0$ (and be 0 for negative r).

We can now use the "Integral Conditioning Formula" on p. 417:

$$\begin{aligned} P(A) &= \int P(A|R=r)f_R(r)dr \\ &= \int_0^\infty e^{-r^2} r e^{-\frac{1}{2}r^2} dr \\ &= \int_0^\infty r e^{-\frac{3}{2}r^2} dr. \end{aligned}$$

To evaluate this integral, we can use integration by substitution with $u = r^2$, or we can use Maple to obtain

$$P(A) = \frac{1}{3}.$$

Answer 2 is correct.

Problem 3

There are $N = 18$ swimmers, out of which $G = 5$ are "good" (best at butterfly) and $B = 13$ are "bad" (not best at butterfly). Out of the 18 swimmers, $n = 4$ are chosen without replacement. Since it is sampling without replacement, the number of chosen "good" swimmers follows a hypergeometric distribution.

We are asked to find the probability that 2 among the 4 chosen are best at butterfly, and implicitly, that 2 are not. Using the formula from the theorem "Sampling With and Without Replacement" on p. 125, we get

$$P(g \text{ good and } b \text{ bad}) = \frac{\binom{G}{g} \binom{B}{b}}{\binom{N}{n}} = \frac{\binom{5}{2} \binom{13}{2}}{\binom{18}{4}}$$

which is the 4th option.

Answer 4 is correct.

Problem 4

Since the times between particles are described by independent random variables with the same exponential distribution, we have a Poisson Arrival Process (cf. p. 284).

The mean value of the waiting time is 2 minutes, so the rate of the waiting times is $\lambda = \frac{1}{2}$ (the mean and the rate of the exponential distribution are inverses of each other, cf. page 279).

Let T_3 denote the arrival time of the 3rd particle. T_3 has a $\text{Gamma}(3, \frac{1}{2})$ distribution as explained in the "Poisson Arrival Times (Gamma Distribution)" box on p. 286.

We are asked to find the probability that the 3rd particle arrives after 3 minutes and before 4 minutes. We can calculate this as the probability that the particle arrives after 3 minutes minus the probability that it arrives after 4 minutes:

$$P(3 < T_3 < 4) = P(T_3 > 3) - P(T_3 > 4).$$

The two terms in this expression are right tail probabilities which we can evaluate according to the formula in the box on p. 286:

$$P(T_r > t) = P(N_t \leq r - 1) = \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Inserting $r = 3$, $\lambda = \frac{1}{2}$, and respectively $t = 3$ and $t = 4$, we obtain:

$$\begin{aligned} P(T_3 > 3) &= \sum_{k=0}^2 \frac{\left(\frac{3}{2}\right)^k}{k!} e^{-\frac{3}{2}} \\ &= e^{-\frac{3}{2}} \left(1 + \frac{3}{2} + \frac{9}{8}\right) \\ &= \frac{29}{8} e^{-\frac{3}{2}} \end{aligned}$$

and

$$\begin{aligned} P(T_3 > 4) &= \sum_{k=0}^2 \frac{\left(\frac{4}{2}\right)^k}{k!} e^{-\frac{4}{2}} \\ &= e^{-2} (1 + 2 + 2) \\ &= 5e^{-2}. \end{aligned}$$

Combining, we get

$$\begin{aligned} P(3 < T_3 < 4) &= P(T_3 > 3) - P(T_3 > 4) \\ &= \frac{29}{8} e^{-\frac{3}{2}} - 5e^{-2}, \end{aligned}$$

which is answer 2.

Answer 2 is correct.

Problem 5

We are given the joint density of X and Y , and we are asked to find the density function of the ratio between them. We can use the formula for the ratio density given on page 383, while taking care to notice that X and Y are switched in our situation:

$$f_Z(z) = \int_{-\infty}^{\infty} |y| f(y, yz) dy$$

The given formula $f(x, y) = 2\lambda^2 e^{-\lambda(x+y)}$ for the joint density only applies when both $x > 0$ and $y > 0$ since X and Y cannot be negative. Using this, we see that $f(y, yz)$ is 0 for negative z no matter the sign of y . We conclude that $f_Z(z) = 0$ for negative z .

Assuming z is not negative, $f(y, yz)$ will evaluate to 0 for negative y . Hence we only need to integrate from 0 to infinity with regard to y :

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |y| f(y, yz) dy \\ &= \int_0^{\infty} y \cdot 2\lambda^2 e^{-\lambda(y+yz)} dy \\ &= 2\lambda^2 \int_0^{\infty} y e^{-\lambda(1+z)y} dy \end{aligned}$$

Integrating by parts or using Maple, we obtain the ratio density

$$f_Z(z) = \frac{2}{(1+z)^2}.$$

Since X is no bigger than Y , the ratio $Z = X/Y$ is at most 1. So the ratio density we have found only applies for $0 \leq z \leq 1$.

Answer 1 is correct.

Problem 6

On p. 42, it is noted that the condition $P(A \cap B) = P(A)P(B)$, which is the defining condition for independence, is equivalent to the condition $P(A|B) = P(A)$, which is option 3.

We could also deduce the second condition from the first by using the definition of

December 14, 2020

conditional probability:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(A)P(B)}{P(B)} \\ &= P(A). \end{aligned}$$

Note: We assume $P(B) \neq 0$ to answer this question.

Answer 3 is correct.

Problem 7

We can use the "Alternative Formula" for covariance from the box on p. 430:

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Since X is uniformly distributed, we notice from the symmetry of the outcome space $\{-2, -1, 1, 2\}$ around 0 that the expectation of X is 0. The calculation from the definition of expectation on p. 163 would be:

$$E(X) = \sum_{\text{all } x} xP(x) = \frac{1}{4}((-2) + (-1) + 1 + 2) = 0$$

where we have used the fact that each of the 4 outcomes are equally likely and thus have probability $1/4$.

Using $E(X) = 0$, we see that

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(XY)$$

and using $Y = X^2$, we obtain:

$$Cov(X, Y) = E(X^3).$$

We can calculate this expectation by the formula from the "Expectation of a Function of X" box on p. 175:

$$Cov(X, Y) = E(X^3) = \sum_{\text{all } x} x^3P(x) = \frac{1}{4}((-2)^3 + (-1)^3 + 1^3 + 2^3) = 0.$$

Answer 2 is correct.

Problem 8

We have 400 passengers, with each passenger bringing too much hand luggage with a probability of 20%, independently of each other. This situation can be seen as a series of 400 independent $Bernoulli(0.2)$ trials, and the number of "successes" (too much hand luggage) thus follows a $binomial(400, 0.2)$ distribution.

Let X denote the number of passengers with too much hand luggage. Using the binomial probability formula on p. 81, we get:

$$\begin{aligned} P(X \leq 60) &= \sum_{i=0}^{60} \binom{n}{i} (p)^i (1-p)^{n-i} \\ &= \sum_{i=0}^{60} \binom{400}{i} (0.2)^i (1-0.2)^{400-i}. \end{aligned}$$

However, this is not one of the options.

We can then try to use the "Normal Approximation to the Binomial Distribution" on p. 99. In this case, we get

$$\begin{aligned} P(X \leq 60) &= \Phi\left(\frac{60 + 0.5 - np}{\sqrt{npq}}\right) - 0 \\ &= \Phi\left(\frac{60 - 0.5 - 400 \cdot 0.2}{\sqrt{400 \cdot 0.2 \cdot 0.8}}\right) \\ &= \Phi(-2.4) \end{aligned}$$

which is answer 5.

Answer 5 is correct.

Problem 9

X and Y have standard bivariate normal distribution with correlation $\rho = \frac{3}{5}$. Then, according to the "Standard Bivariate Normal Distribution" theorem on p. 451, we can write Y as

$$\begin{aligned} Y &= \rho X + \sqrt{1 - \rho^2} Z \\ &= \frac{3}{5}X + \frac{4}{5}Z \end{aligned}$$

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

We are asked to find $P(\frac{1}{2}X < Y < 2X)$. Substituting Y and splitting explicitly into two

inequalities, we get

$$\begin{aligned}
 P\left(\frac{1}{2}X < Y < 2X\right) &= P\left(\frac{1}{2}X < \frac{3}{5}X + \frac{4}{5}Z < 2X\right) \\
 &= P\left(\frac{1}{2}X < \frac{3}{5}X + \frac{4}{5}Z \text{ and } \frac{3}{5}X + \frac{4}{5}Z < 2X\right) \\
 &= P(5X < 6X + 8Z \text{ and } 3X + 4Z < 10X) \\
 &= P(Z > \frac{-X}{8} \text{ and } Z < \frac{7X}{4}).
 \end{aligned}$$

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 two inequalities correspond to the region in the 1st and 4th quadrant between the two lines with slope $7/4$ and $-1/8$. The angle between these two lines is given by

$$\operatorname{Arctan}\left(\frac{7}{4}\right) - \operatorname{Arctan}\left(-\frac{1}{8}\right) = \operatorname{Arctan}\left(\frac{7}{4}\right) + \operatorname{Arctan}\left(\frac{1}{8}\right).$$

Due to the rotational symmetry, the probability of landing in this region is given by this angle divided by 2π , so we finally get:

$$P\left(\frac{1}{2}X < Y < 2X\right) = P\left(Z > \frac{-X}{8} \text{ and } Z < \frac{7X}{4}\right) = \frac{\operatorname{Arctan}\left(\frac{7}{4}\right) + \operatorname{Arctan}\left(\frac{1}{8}\right)}{2\pi},$$

which is answer 4.

Answer 4 is correct.

Problem 10

We first note that X_1 and $Y = X_1 + X_2 + X_3$ are not independent, so we cannot just multiply the marginal densities to get the joint density (as done in the "Independence" formula on p. 349).

However, we can use the conditional "Multiplication Rule for Densities" on p. 416:

$$f(x, y) = f_X(x) \cdot f_Y(y|X = x).$$

Translating to our case, we get

$$f_{(X_1, Y)}(x, y) = f_{X_1}(x) \cdot f_Y(y|X_1 = x).$$

We are given that $X_1 \sim \exp(\lambda)$, so we know that

$$f_{X_1}(x) = \lambda e^{-\lambda x}.$$

Let us then consider the second factor $f_Y(y|X_1 = x)$. The key is to note that for fixed

$X_1 = x$, the whole of $Y = X_1 + X_2 + X_3$ follows the distribution of $X_2 + X_3$, but just moved x to the right. Expressed in densities, this means that

$$f_Y(y|X_1 = x) = f_{X_2+X_3}(y - x).$$

So what is the density of $X_2 + X_3$? Well, we know that they are exponentially distributed, so this can be seen as the sum of two waiting times in a Poisson Arrival process, which is $\text{gamma}(2, \lambda)$ distributed. Using the density formula on p. 286 with $t = y - x$ and $r = 2$, we obtain

$$\begin{aligned} f_Y(y|X_1 = x) &= f_{X_2+X_3}(y - x) \\ &= e^{-\lambda t} \frac{(\lambda t)^{r-1}}{(r-1)!} \lambda \\ &= e^{-\lambda(y-x)} \frac{(\lambda(y-x))^{2-1}}{(2-1)!} \lambda \\ &= e^{-\lambda(y-x)} (\lambda(y-x)) \lambda. \end{aligned}$$

Combining our two factors, we get

$$f_{(X_1, Y)}(x, y) = f_{X_1}(x) \cdot f_Y(y|X_1 = x) = \lambda e^{-\lambda x} \cdot e^{-\lambda(y-x)} (\lambda(y-x)) \lambda = \lambda^3 (y-x) e^{\lambda y}$$

which is option 3.

Answer 3 is correct.

Problem 11

We are given that the patients arrive independently at an average rate of 24 per day. In addition to this, we are to assume that the rate is constant throughout the day. Given this, it is reasonable to assume that the arrivals come at random times, and we thus have a Poisson Arrival process with a rate λ of 24 per day.

The distribution of the number N of arrivals at a specific time of t days is then a $\text{Poisson}(\lambda t)$ distribution, cf. box on p. 284. At time $t = 1/6$ days, we can calculate

$P(N_{1/6} \leq 2)$ with the "Right tail probability" formula on p. 286:

$$\begin{aligned} P(N_{1/6} \leq 2) &= \sum_{k=0}^2 e^{-\lambda t} \frac{(\lambda t)^k}{k!} \\ &= \sum_{k=0}^2 e^{-24 \cdot \frac{1}{6}} \frac{(24 \cdot \frac{1}{6})^k}{k!} \\ &= \sum_{k=0}^2 e^{-4} \frac{4^k}{k!} \\ &= e^4 \left(1 + 4 + \frac{4^2}{2!}\right) \end{aligned}$$

which is option 5.

Answer 5 is correct.

Problem 12

We are given that X and Y are independent. We will use the fact that then also X^2 and Y^2 are independent, which is intuitively clear (and can be shown to hold).

We first apply the "Computational Formula for Variance" on p. 186 on the product XY :

$$\begin{aligned} Var(XY) &= E((XY)^2) - [E(XY)]^2 \\ &= E(X^2Y^2) - [E(XY)]^2 \end{aligned}$$

We can now take advantage of the independence and apply the "Multiplication Rule for Expectation" on p. 177 on both terms:

$$\begin{aligned} E(X^2Y^2) - [E(XY)]^2 &= E(X^2)E(Y^2) - [E(X)E(Y)]^2 \\ &= E(X^2)E(Y^2) - [E(X)]^2[E(Y)]^2 \end{aligned}$$

and then use the "Computational Formula for Variance" backwards on the new factors in the first term, multiply the brackets, and reduce:

$$\begin{aligned} &= E(X^2)E(Y^2) - [E(X)]^2[E(Y)]^2 \\ &= [Var(X) + [E(X)]^2][Var(Y) + [E(Y)]^2] - [E(X)]^2[E(Y)]^2 \\ &= Var(X)Var(Y) + [E(X)]^2Var(Y) + [E(Y)]^2Var(X) + [E(X)]^2[E(Y)]^2 - [E(X)]^2[E(Y)]^2 \\ &= Var(X)Var(Y) + [E(X)]^2Var(Y) + [E(Y)]^2Var(X) \\ &= \sigma_X^2\sigma_Y^2 + \mu_X^2\sigma_Y^2 + \mu_Y^2\sigma_X^2 \end{aligned}$$

which is answer 4.

December 14, 2020

Answer 4 is correct.

Problem 13

We are given the joint density of X and Y and are asked to find the probability of landing in a very small rectangle of length and width 0.01 at the place $x = 0.9$ and $y = 0.1$. In this situation, we can use the "infinitesimal probability formula" on p. 347 (or the box on p. 349):

$$\begin{aligned} P(X \in dx, Y \in dy) &= f(x, y)dxdy \\ &= 6(x - y)dxdy \\ &= 6(0.9 - 0.1) \cdot 0.01 \cdot 0.01 \\ &= 6 \cdot 0.8 \cdot 0.01^2 \end{aligned}$$

which is option 1.

Answer 1 is correct.

Problem 14

We have a change of variable from X to $Y = X^2$, so we want to use the formula for such a change. We are careful to note that $Y = X^2$ is many-to-one in the range of X which is $[-\frac{1}{2}; \frac{1}{2}]$, so we should use the version of the formula on top of p. 307:

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

Since X is uniform on an interval of length 1, its density is just 1:

$$f_X(x) = 1.$$

For the function $y = x^2$, we have

$$\left| \frac{dy}{dx} \right| = |2x| = |2\sqrt{y}| = 2\sqrt{|y}|.$$

The range of $Y = X^2$ is $[0; \frac{1}{4}]$, since the range of X is $[-\frac{1}{2}; \frac{1}{2}]$. Outside this range $f_Y(y) = 0$.

Finally, for $0 < y \leq \frac{1}{4}$, there are two values x such that $x^2 = y$, so we get

$$\begin{aligned} f_Y(y) &= \sum_{x:g(x)=y} f_X(x) / \left| \frac{dy}{dx} \right| \\ &= \sum_{x:x^2=y} \frac{1}{2\sqrt{|y|}} \\ &= 2 \cdot \frac{1}{2\sqrt{|y|}} \\ &= \frac{1}{\sqrt{|y|}}. \end{aligned}$$

This is option 2 apart from the single point $y = 0$. A single point does not really matter for the density function, and all the other options are wrong.

Answer 2 is correct.

Problem 15

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

$$X \sim \text{normal}(1, 2^2)$$

$$Y \sim \text{normal}(2, 3^2)$$

$$\rho = \frac{-1}{4}$$

We are asked to find $P(X - Y \leq 0)$.

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:

$$X = \mu_X + \sigma_X X^* = 1 + 2X^*$$

$$Y = \mu_Y + \sigma_Y Y^* = 2 + 3Y^*$$

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

$$\begin{aligned} P(X - Y \leq 0) &= P(1 + 2X^* - (2 + 3Y^*) \leq 0) \\ &= P(2X^* - 3Y^* \leq 1). \end{aligned}$$

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

$$\begin{aligned} P(X - Y \leq 0) &= P(2X^* - 3Y^* \leq 1) \\ &= P\left(2X^* - 3\left(\frac{-1}{4}X^* + \frac{\sqrt{15}}{4}Z^*\right) \leq 1\right) \\ &= P(11X^* - 3\sqrt{15}Z^* \leq 4). \end{aligned}$$

Now, since X^* and Z^* are independent standard normal variables, a linear combination $V = 11X^* - 3\sqrt{15}Z^*$ of them is a normal variable with mean zero and standard deviation given by

$$\sigma_V^2 = 11^2 \cdot 1^2 + (3\sqrt{15})^2 \cdot 1^2 = 256 = 16^2.$$

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 this linear combination V into V^* by dividing with its SD of 16. Doing this, we finally obtain:

$$\begin{aligned} P(X - Y \leq 0) &= P(11X^* - 3\sqrt{15}Z^* \leq 4) \\ &= P(V \leq 4) \\ &= P(V^* \leq 1/4) \\ &= \Phi\left(\frac{1}{4}\right) \end{aligned}$$

which is answer 4.

Answer 4 is correct.

Problem 16

We know the inclusion-exclusion formula for 3 events from exercise 1.3.11 on p. 31:

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

December 14, 2020

Adding $P(B \cap C)$ and subtracting $P(A \cup B \cup C)$ from both sides gives us

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

which is answer 3.

Answer 3 is correct.

Problem 17

We want to express $P(H)$ for the event $H = (F_0 \cup F_1 \cup F_2)$, where F_i is the event that there is i errors in the material. This union H is the event that there is either 0, 1 or 2 errors in the material. Given that N is a random variable describing the number of errors in the material, and given that the number of errors follows a Poisson distribution and thus only has whole number outcomes 0, 1, 2, 3, ..., this event H is the same as the event $N \leq 2$. We thus have

$$P(H) = P(N \leq 2)$$

which is answer 1.

Answer 1 is correct.

Problem 18

The number of eggs laid by a specific fish species is a Poisson process with mean value (or intensity) μ . Every egg has a female fish larva with probability p , and we are asked to find the expected number of eggs with a female fish larva.

Solution 1. This situation can be seen as a thinning of a Poisson scatter where we "keep" the female eggs with probability p , and where we ignore the positional aspect and consider the unit area to be one egg laying event. According to the theorem "Thinning a Poisson Scatter" on p. 232, the intensity (or mean value or expected value) of the "kept" eggs is μp , which is answer 2.

Solution 2. We can use the "Rule of Average Conditional Expectations" on p. 402. Let N denote the number of laid eggs, and let F denote the number of female eggs among them.

Now, given $N = n$, and assuming the genders of the eggs are independent, the distribution of the number of female eggs is a $\text{binomial}(n, p)$ distribution which has expected value np . So we have $E(F|N = n) = np$.

Applying the average conditional expectation formula, we now see:

$$\begin{aligned}
 E(F) &= \sum_{\text{all } n} E(F|N=n)P(N=n) \\
 &= \sum_{\text{all } n} npP(N=n) \\
 &= p \sum_{\text{all } n} nP(N=n) \\
 &= pE(N) \\
 &= p\mu
 \end{aligned}$$

which is answer 2.

Answer 2 is correct.

Problem 19

Since the aiming point is at the center of the coordinatsystem, and the coordinates of the hit are independent standard normal variables, the distance R from the aiming point to the hit follows a Rayleigh distribution, cf. p. 357-359.

The c.d.f. of the Rayleigh distribution (p. 359) is

$$F_R(r) = 1 - e^{-\frac{1}{2}r^2}$$

and inserting the interval endpoints, we obtain:

$$\begin{aligned}
 P(1 \leq R \leq 2) &= F_R(2) - F_R(1) \\
 &= 1 - e^{-\frac{1}{2}\cdot 2^2} - (1 - e^{-\frac{1}{2}\cdot 1^2}) \\
 &= e^{-\frac{1}{2}} - e^{-2}
 \end{aligned}$$

which is answer 3.

Answer 3 is correct.

Problem 20

Let X denote the round trip time. Since we know the expected value and the standard deviation of X as well as a bounding probability, we can try using Chebychev's Inequality on p. 191:

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

We try setting the bounding probability of $\frac{1}{9}$ equal to the right side $\frac{1}{k^2}$ to see if we get something we can use. In that case $k = 3$, so we use Chebychev's inequality with this value. Inserting $E(X) = 12$ and $SD(X) = 6$, we obtain:

$$\begin{aligned} P[|X - E(X)| \geq kSD(X)] &\leq \frac{1}{k^2} \\ P[|X - 12| \geq 3 \cdot 6] &\leq \frac{1}{3^2} \\ P[|X - 12| \geq 18] &\leq \frac{1}{9}. \end{aligned}$$

Now, since the round trip time X is positive, $|X - 12| \geq 18$ is equivalent to $X \geq 30$, so we obtain

$$P(X \geq 30) \leq \frac{1}{9}.$$

In other words, the probability of X exceeding the value 30 is at most $\frac{1}{9}$. This property of the value is exactly what we were looking for, so option 1 (30ms) is the right answer.

Answer 1 is correct.

Problem 21

We are given that the electrical component has reached the age t , and we are asked to find the probability that the component fails in an infinitesimal interval $[t, t + dt]$ after t .

Solution 1. We know that the exponential function has constant failure rate λ , cf. remark on p. 296. So substituting the general failure rate $\lambda(t)$ with λ in formula 3 on p. 297 yields:

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

which is answer 1.

Solution 2. Since the exponential distribution is memoryless, the probability to survive a further time dt is simply the probability to survive to time dt in the first place, cf. the theorem "Memoryless Property of the Exponential Distribution" on p. 279. Conversely, the probability to fail is the same. And the probability to fail from time 0 to dt is just the

density at 0 multiplied with the infinitesimal interval length dt . So we get:

$$\begin{aligned} P(t < T < t + dt | T > t) &= P(0 < T < dt) \\ &= f(0)dt \\ &= \lambda e^{-\lambda \cdot 0} dt \\ &= \lambda dt \end{aligned}$$

which is answer 1.

Answer 1 is correct.

Problem 22

We are given that X is positive and that the c.d.f of X is

$$F(X) = P(X \leq x) = 1 - e^{-\lambda x}.$$

We can recognize this as the c.d.f. of the exponential distribution. The survival function is obtained by:

$$G(X) = P(X \geq x) = 1 - F(X) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}.$$

Given $Y = \frac{1}{X}$, we can now directly find:

$$\begin{aligned} P(Y \leq y) &= P\left(\frac{1}{X} \leq y\right) \\ &= P\left(X \geq \frac{1}{y}\right) \\ &= G\left(\frac{1}{y}\right) \\ &= e^{-\lambda \cdot \frac{1}{y}} \end{aligned}$$

which is answer 5.

Answer 5 is correct.

Problem 23

Let us look at the situation before the second roll of the dice:

We have $n = 6 - v$ dice left.

We are asking for the probability to get exactly $k = w - v$ sixes in the second roll (since that means we in total get w sixes after getting v sixes in the first roll).

This is a binomial situation with n and k as above and $p = \frac{1}{6}$. Inserting this in the binomial formula (p. 81), we obtain

$$\binom{n}{k} p^k (1-p)^{n-k} = \binom{6-v}{w-v} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{6-w}$$

which is option 5.

Answer 5 is correct.

Problem 24

Since the point within the circle is chosen at random, the probability of landing in a certain region of the circle is given by the area of that region divided by the area of the circle, cf. p. 340.

We are looking for the region of points where the slope of the line through the center of the circle and the point is numerically less than 1. These lines are exactly those lines that make an angle with the first axis of numerically less than 45 degrees.

So which points give such lines? We should take care to include both of these two sections of the circle:

Points to the right of the center in the circle section from -45 to +45 degrees.

Points to the left of the center in the circle section from 135 to 225 degrees.

These are two circle sections of 90 degrees, so they have a total area of half the circle area. So the probability of landing in this region (which is the probability of meeting the condition) is $\frac{1}{2}$.

Answer 5 is correct.

Problem 25

The last sentence in the text lets us assume two things:

1. That at most one failure cause occurs. This means that the 3 failure causes and the event of no failure form a partition of the outcome space. Hence we can use Bayes' rule on p. 49.
2. That a crash only happens if one of the 3 failure causes occurred. This means that the probability of a crash, given no failure, is 0.

Let us use Bayes' formula. We introduce:

December 14, 2020

H denotes human error.

M denotes mechanical error.

T denotes terror.

N denotes no failure.

C denotes the event of a crash.

Our priors are:

$$P(H) = \frac{1}{50}$$

$$P(M) = \frac{1}{500}$$

$$P(T) = \frac{1}{500000}$$

$P(N)$ = whatever (it doesn't matter after multiplication with 0)

Our likelihoods are:

$$P(C|H) = \frac{1}{10000}$$

$$P(C|M) = \frac{1}{1000}$$

$$P(C|T) = \frac{2}{3}$$

$$P(C|N) = 0$$

Applying Bayes' rule, we get

$$\begin{aligned} P(T|C) &= \frac{P(C|T)P(T)}{P(C)} \\ &= \frac{P(C|T)P(T)}{P(C|H)P(H) + P(C|M)P(M) + P(C|T)P(T) + P(C|N)P(N)} \\ &= \frac{\frac{2}{3} \frac{1}{500000}}{\frac{1}{10000} \frac{1}{50} + \frac{1}{1000} \frac{1}{500} + \frac{2}{3} \frac{1}{500000} + 0 \cdot P(N)} \\ &= \frac{\frac{2}{3} \frac{1}{500000}}{\frac{8}{3} \frac{1}{500000}} \\ &= \frac{1}{4} \end{aligned}$$

which is answer 1.

Answer 1 is correct.

Problem 26

We have 5 independent and identically distributed variables, so we can use the theorem "Density of the kth Order Statistic" on p. 326.

The c.d.f. and density of the exponential distribution are

$$F(x) = 1 - e^{-\mu x}$$

$$f(x) = \mu e^{-\mu x}$$

and we are looking for the density $g(x)$ of the second largest of the 5 variables, which translates to $k = 4$ and $n = 5$. Inserting all this in the formula from the theorem, we get

$$\begin{aligned} g(x) &= nf(x) \binom{n-1}{k-1} (F(x))^{k-1} (1 - F(x))^{n-k} \\ &= 5\mu e^{-\mu x} \binom{5-1}{4-1} (1 - e^{-\mu x})^{4-1} (1 - (1 - e^{-\mu x}))^{5-4} \\ &= 20\mu e^{-2\mu x} (1 - e^{-\mu x})^3 \\ &= 20\mu e^{-2\mu x} (1 - 3e^{-\mu x} + 3e^{-2\mu x} - e^{-3\mu x}) \end{aligned}$$

and multiplying into the bracket gives us option 5.

Answer 5 is correct.

Problem 27

Quote from p. 278:

"These exponential and gamma distributions, studied in this section, are the continuous analogs of the geometric and negative binomial distributions of Section 3.4."

See also p. 283 for a discussion and p. 299-289 for a summary.

Briefly put:

1. The exponential distribution is the continuous waiting time between arrivals in a Poisson arrival process, and the total continuous waiting time until the r 'th arrival is the gamma distribution.
2. The geometric distribution is the discrete waiting time between successes in a series of Bernoulli trials, and the total discrete waiting time until the r 'th success is the negative binomial distribution.

Hence the gamma distribution corresponds to the negative binomial distribution in this comparison.

Answer 4 is correct.

Problem 28

We don't need the standard deviations and the covariance to solve this question.

We note that $U = 3X - 2Y + 1$ is formed by multiplications with a constant factor, a summation and a linear function. The formulas for the expectation after these operations are (p. 167 and p. 175):

$$\begin{aligned} E(X + Y) &= E(X) + E(Y) \\ E(cX) &= cE(X) \\ E(aX + b) &= aE(X) + b \end{aligned}$$

Applying these formulas in succession, we obtain

$$\begin{aligned} E(U) &= E(3X - 2Y + 1) \\ &= E(3X - 2Y) + 1 \\ &= E(3X) + E(-2Y) + 1 \\ &= 3E(X) - 2E(Y) + 1 \\ &= 3 \cdot 4 - 2 \cdot 7 + 1 \\ &= -1 \end{aligned}$$

which is answer 2.

Answer 2 is correct.

Problem 29

We follow example 1 p. 455.

Let U denote the catch of tobis, and let V denote the catch of herring. Let X denote the catch of tobis in standard units, and let Y denote the catch of herring in standard units.

Since the simultaneous distribution of the repetitive catches U and V is assumed bivariate normal, also the standardized variables X and Y have a bivariate normal distribution, with the same correlation, according to the box on p. 454.

The natural prediction for Y given $X = x$ is

$$E(Y|X = x) = \rho x$$

where ρ is the correlation, cf. the remark in the example on p. 455.

We are given a catch of tobis of $U = 14$. Converting to standard units, this corresponds to:

$$X = \frac{14 - \mu_U}{\sigma_U} = \frac{14 - 10}{2} = 2.$$

December 14, 2020

Inserting this, we get:

$$E(Y|X = 2) = \rho x = 0.7 \cdot 2 = 1.4.$$

This is the catch Y of herring in standard units. Converting back to V , we obtain the answer:

$$E(V|U = 14) = \mu_V + \sigma_V \cdot E(Y|X = 2) = 1 + 0.3 \cdot 1.4 = 1.42$$

which is answer 3.

Answer 3 is correct.

Problem 30

We are given the joint density

$$f(x, y) = K(y - x)(1 - y) \text{ for } 0 < x < y < 1.$$

We can find the marginal density with the formula on p. 349:

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

The main challenge is to find the right integration limits with regard to x . We see from the given joint density that for fixed y , the joint density is only non-zero for x between 0 and y . So we need only integrate from 0 to y , and we get:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \int_0^y K(y - x)(1 - y) dx \\ &= K(1 - y) \int_0^y (y - x) dx \\ &= K(1 - y)[yx - \frac{1}{2}x^2]_{x=0}^{x=y} \\ &= K(1 - y)(y^2 - \frac{1}{2}y^2) \\ &= \frac{K}{2}(1 - y)y^2 \end{aligned}$$

which is answer 3.

Answer 3 is correct.