02405 Probability Theory

Exam solutions E18

Technical University of Denmark - 18.11.2018

Introduction

These are the solutions to the exam set of the course "02405 Probability Theory" offered at the Technical University of Denmark in the fall of 2018. All references in this solution manual refer to the course text book *Probability* by Pitman,

Problem 1

We consider three tosses with a fair coin, i.e. there is a 50% of tails for each toss. Let X denote the number of tails in the three tosses, then $X \sim \text{Bin}\left(3, \frac{1}{2}\right)$. The probability of getting exatcly one tails in the three tosses is then given by

$$\mathbb{P}(X = 1) = \binom{3}{1}\left(\frac{1}{2}\right)\left(1 - \frac{1}{2}\right)^{3-1} = \frac{3}{8}.$$

Answer 5 is correct.

Problem 2

There are three different car categories, say 1, 2, and 3. Define then X_i as the number of claims for cars in category $i \in \{1, 2, 3\}$ in a given week. We then have that

$$X_1 \sim \text{Pois}\left(\frac{3}{2}\right), \quad X_2 \sim \text{Pois}(1), \quad \text{and} \quad X_3 \sim \text{Pois}\left(\frac{5}{2}\right).$$

Define then $X = X_1 + X_2 + X_3$ as the total number of claims in a given week. Since the random variables X_1, X_2, and X_3 are independent, the theorem "Sum of Independent Poisson Variables are Poisson" (p. 226) applies and yields that $X \sim \text{Pois}(\mu)$, where $\mu = 3/2 + 1 + 5/2 = 5$. Therefore,

$$\mathbb{P}(X > 8) = \sum_{i=9}^{\infty} \mathbb{P}(X = i) = \sum_{i=9}^{\infty} \frac{\mu^i}{i!} e^{-\mu} = \sum_{i=9}^{\infty} \frac{5^i}{i!} e^{-5},$$

cf. the probability mass function for Poisson random variables.

Answer 2 is correct.

Problem 3

Let X and Y denote the deviations from the target point measured on the first and second axes of the coordinate system, respectively. From the problem description, we know that X and Y are independent and identically distributed according to a standard normal distribution, i.e. $X, Y \sim \mathcal{N}(0,1)$.

The distance to the target point, say R, is evaluated as $R = \sqrt{X^2+Y^2}$. Consequently, R is distributed according to a Rayleigh distribution, cf. p. 359. From the distribution summaries on p. 477, we find that

$$\mathbb{E}[R] = \sqrt{\frac{\pi}{2}}.$$

Alternatively, the expected value can be found using the probability density function given on p. 359:

$$\mathbb{E}[R] = \int_0^\infty r f_R(r) dr = \int_0^\infty r^2 e^{-\frac{1}{2}r^2} dr = \sqrt{\frac{\pi}{2}}.$$

Answer 4 is correct.

Problem 4

We are given that $X \sim \text{Exp}(\lambda)$ and $Y = \sqrt[3]{X}$. Since X is non-negative and the function $g(x) = \sqrt[3]{x}$ is strictly increasing for $x \geq 0$, the "One-to-One Change of Variable for Densities"-theorem (p. 304) applies and yields that

$$f_Y(y) = \frac{f_X(x)}{|g'(x)|} = \frac{\lambda e^{-\lambda x}}{\left|\frac{1}{3}x^{-2/3}\right|} = \frac{\lambda e^{-\lambda x}}{\frac{1}{3}x^{-2/3}}, \quad x > 0,$$

cf. the probability density function for exponentially distributed random variables. As $y = \sqrt[3]{x}$, it follows that $x = y^3$, which is inserted in the above expression:

$$f_Y(y) = \frac{\lambda e^{-\lambda y^3}}{\frac{1}{3}(y^3)^{-2/3}} = 3\lambda y^2 e^{-\lambda y^3}, \quad y > 0.$$

The condition that $x = y^3 > 0$ is equivalent with $y > 0$, which justifies the condition above.

Answer 4 is correct.

Problem 5

First note that $Z = 2X - Y = 2X + (-Y)$. Therefore,

$$\mathbb{V}[Z] = \mathbb{V}[2X + (-Y)] = \mathbb{V}[2X] + \mathbb{V}[-Y] + 2\text{Cov}(2X, -Y),$$

cf. the theorem "Variance of a Sum" on p. 430. Using the linearity of the variance operator and the bilinearity of the covariance operator (see. "Covariance is Bilinear" on p. 444), we obtain that

$$\mathbb{V}[Z] = 2^2\mathbb{V}[X] + (-1)^2\mathbb{V}[Y] + (-1)(2)2\text{Cov}(X, Y) = 4\mathbb{V}[X] + \mathbb{V}[Y] - 4\text{Cov}(X, Y).$$

To calculate the variances we apply the "Computational Formula for Variance" on p. 186, while the "Alternative Formula" in the "Definition of Covariance" on p. 430 is used to calculate the covariance. Thus,

$$\begin{aligned}
\mathbb{V}[Y] &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = 8 - 2^2 = 4,\\
\mathbb{V}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 10 - (-3)^2 = 1,\\
\text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = (-4) - (-3)2 = 2.
\end{aligned}$$

These quantities are inserted into the calculation for the variance of Z to get that

$$\mathbb{V}[Z] = 4 \cdot 1 + 4 - 4 \cdot 2 = 0.$$

Answer 5 is correct.

Problem 6

Let X_i denote the lifetime of component $i \in \{1, ..., 1000\}$ and define $W = \sum_{i=1}^{1000} X_i$ as the total lifetime of the system. For exponentially distributed random variables, the mean is equal to the inverse of the rate parameter, i.e. $X_i \sim \text{Exp}(2)$ for $i \in \{1, ..., 1000\}$. Therefore, since the lifetimes of the components can be assumed independent, we have that $W \sim \text{Gamma}(1000, 2)$, cf. "Poisson Arrival Times (Gamma Distribution)" on p. 286. However, as this is not a possibility, we seek an approximate result.

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. Therefore, define $S_{1000} = \sum_{i=1}^{1000} X_i$. According to the CLT, S_{1000} is approximately distributed according to a normal distribution with mean $\mathbb{E}[S_{1000}] = 1000\mathbb{E}[X_i] = 500$ and standard deviation $\text{SD}(S_{1000}) = \text{SD}(X_i)\sqrt{1000} = \sqrt{1000}/2 = \sqrt{250}$ (as $\mathbb{E}[X_i] = \text{SD}(X_i)$). Hence, $S_{1000} \sim \mathcal{N}(500, 250)$.

Answer 2 is correct.

Problem 7

The results follows directly from an application of the "Inclusion-Exclusion" theorem on p. 22:

$$\mathbb{P}(AB) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B) = \frac{2}{3} + \frac{3}{4} - 1 = \frac{5}{12}.$$

Answer 3 is correct.

Problem 8

The marginal distribution of Y can be found with the boxed result "Marginal Probabilities" on p. 145:

$$\mathbb{P}(Y = y) = \sum_{x=1}^{6} \mathbb{P}(X = x, Y = y).$$

For a given y, the number of x-values exceeding y is $6 - y$, while the number of x-values below y is $y - 1$. Consequently, we get that

$$\mathbb{P}(Y = y) = \sum_{x=1}^{6} \mathbb{P}(X = x, Y = y) = (6 - y)\frac{1}{18} + \frac{1}{36} = \frac{13 - 2y}{36},$$

as $\mathbb{P}(X = x, Y = y) = 1/18$ for $x > y$ and $\mathbb{P}(X = x, Y = y) = 1/36$ for $x = y$.

Answer 5 is correct.

Problem 9

This problem can be solved in many ways, but the most straightforward method invokes the "Integral Conditioning Formula" on p. 417. Applying the theorem yields

$$\mathbb{P}(Y > X) = \int_0^\infty \mathbb{P}(Y > X | X = x) f_X(x) dx = \int_0^\infty \mathbb{P}(Y > x) f_X(x) dx = \int_0^\infty e^{-\lambda x} x e^{-\frac{1}{2}x^2} dx,$$

cf. the survival function for $Y \sim \text{Exp}(\lambda)$ and the probability density function for $X \sim$ Rayleigh.

Answer 2 is correct.

Problem 10

Define the random variable X as the number of heads in the 400 tosses with the perfectly fair coin. Then $X \sim \text{Bin}\left(400, \frac{1}{2}\right)$ and the exact probability is given by

$$\mathbb{P}(X = 150) = \binom{400}{150} \left(\frac{1}{2}\right)^{150} \left(1 - \frac{1}{2}\right)^{250} = \binom{400}{150} \left(\frac{1}{2}\right)^{400}.$$

However, as this answer is not an option, an approximative probability is found. We therefore apply the "Normal Approximation to the Binomial Distribution" on p. 99, which states that

$$\mathbb{P}(a \leq X \leq b) \approx \Phi\left(\frac{b + \frac{1}{2} - \mu}{\sigma}\right) - \Phi\left(\frac{a - \frac{1}{2} - \mu}{\sigma}\right),$$

where $\mu = np$ and $\sigma = \sqrt{np(1-p)}$. Here n and p refer to the parameters in the binomial distribution, i.e. $n = 400$ and $p = 1/2$. Hence,

$$\mathbb{P}(X = 150) = \mathbb{P}(150 \leq X \leq 150) \approx \Phi\left(\frac{150 + \frac{1}{2} - 200}{10}\right) - \Phi\left(\frac{150 - \frac{1}{2} - 200}{10}\right).$$

Answer 4 is correct.

Problem 11

Let $R \sim$ Rayleigh. We shall calculate the hazard rate from "Hazard from density and survival" in the box on p. 297. The probability density function of R is given on p. 359 as

$$f_R(r) = re^{-\frac{1}{2}r^2}, \ \ r > 0,$$

while the cumulative distribution function for R is found just below on the same page:

$$F_R(r) = 1 - e^{-\frac{1}{2}r^2}, \ \ r > 0.$$

The survival function follows directly as $G_R(r) = 1 - F_R(r) = e^{-\frac{1}{2}r^2}$. Hence, the hazard rate is found as

$$\lambda(t) = \frac{f_R(t)}{G_R(t)} = \frac{te^{-\frac{1}{2}t^2}}{e^{-\frac{1}{2}t^2}} = t, \ \ t > 0.$$

The function $\lambda(t) = t$ is strictly increasing.

Answer 1 is correct.

Problem 12

From the problem description we have that $B \sim \mathcal{N}(\mu, \sigma^2)$ and $\mathbb{E}[A|B = b] = \phi + \kappa b + \gamma b^2$. To find the unconditional expectation of A, we apply the theorem "Average conditional expectation" on p. 425:

$$\mathbb{E}[A] = \int_{-\infty}^{\infty} \mathbb{E}[A|B = b] f_B(b) db = \int_{-\infty}^{\infty} \left(\phi + \kappa b + \gamma b^2\right) f_B(b) db,$$

where f_B is the probability density function of B. After multiplying out the brackets, we apply the linearity of the integral to obtain:

$$\mathbb{E}[A] = \phi \int_{-\infty}^{\infty} f_B(b) db + \kappa \int_{-\infty}^{\infty} b f_B(b) db + \gamma \int_{-\infty}^{\infty} b^2 f_B(b) db.$$

Recall that integrating a probability density function over its domain yields one. Furthermore, the expectation of a function of B can be evaluated using "Expectation of a Function" on the bottom of p. 263 as $\mathbb{E}[g(B)] = \int_{-\infty}^{\infty} g(b) f_B(b) db$. Thus,

$$\mathbb{E}[A] = \phi + \kappa \mathbb{E}[B] + \gamma \mathbb{E}[B^2].$$

From the definition it follows easily that $\mathbb{E}[B] = \mu$. The expectation of B^2 is found using the "Computational Formula for Variance" on p. 186:

$$\mathbb{V}[B] = \mathbb{E}[B^2] - \mathbb{E}[B]^2 \Leftrightarrow \mathbb{E}[B^2] = \mathbb{V}[B] + \mathbb{E}[B]^2 = \sigma^2 + \mu^2.$$

In conclusion, we get that

$$\mathbb{E}[A] = \phi + \kappa\mu + \gamma(\mu^2 + \sigma^2).$$

Answer 4 is correct.

Problem 13

As every spin on the American roulette is independent and the probability distribution over the different colours is the same at every spin, we are in the context of repeated measurements (we repeat an experiment multiple times under identical conditions). Therefore, every spin can be considered a Bernoulli trial with success probability $18/(18 + 18 + 2) = 18/38 = 9/19$, i.e. the probability of hitting a red number. The waiting time until the third success, say T_3, is then distributed according to a negative binomial distribution, specifically $T_3 \sim \text{NB}\left(3, \frac{9}{19}\right)$. The probability that the third success occurs on the fifth spin is then given as

$$\mathbb{P}(T_3 = 5) = \binom{4}{2} \left(\frac{9}{19}\right)^3 \left(1 - \frac{9}{19}\right)^2 = \binom{4}{2} \left(\frac{9}{19}\right)^3 \left(\frac{10}{19}\right)^2,$$

cf. the probability mass function of negatively binomially distributed random variables.

Answer 5 is correct.

Problem 14

From the problem description we have that (X, Y) is uniformly distributed on the disc $D = \{(x, y)|x^2 + y^2 \leq 1\}$, which has an area of π. Consequently, the joint probability density function of (X, Y) is given as

$$f_{XY}(x, y) = \frac{1}{\pi}, \ \ (x, y) \in D,$$

and zero elsewhere. For points $(x, y) \in D$, we have that $-1 \leq x, y \leq 1$. Therefore, the condition that $y > x + 1$ implies that $y > 0$ and that $x < 0$. For such values of x and y, we get that

$$x^2 + y^2 \leq 1 \Leftrightarrow 0 < y \leq \sqrt{1 - x^2}.$$

We now calculate the wanted probability by integrating over the proper limits:

$$\mathbb{P}(Y > X + 1) = \int_{-1}^{0} \int_{x+1}^{\sqrt{1-x^2}} \frac{1}{\pi} dy dx = \frac{1}{\pi} \int_{-1}^{0} \left(\sqrt{1 - x^2} - (x + 1) \right) dx = \frac{1}{4} - \frac{1}{2\pi}.$$

Alternatively, this problem can be solved graphically by evaluating the proportion of the disc D, which satisfies the condition that $y > x + 1$.

Answer 2 is correct.

Problem 15

Let the events A, B, and C denote that the associated person is guilty. The initial estimates (prior probabilities) that the persons are guilty given by the police are $\mathbb{P}(A) = 9/10$, $\mathbb{P}(B) = 9/100$, and $\mathbb{P}(C) = 1/100$. Since there is only one perpetrator, A, B, and C constitute a partition. Let D_C denote the event that the police dog selects person C. We then apply Bayes' Rule (p. 49) to evaluate the probability that person C is guilty given that the police dog selected person C:

$$\begin{aligned}
\mathbb{P}(C|H_C) &= \frac{\mathbb{P}(H_C|C)\mathbb{P}(C)}{\mathbb{P}(H_C|A)\mathbb{P}(A) + \mathbb{P}(H_C|B)\mathbb{P}(B) + \mathbb{P}(H_C|C)\mathbb{P}(C)} \\
&= \frac{\frac{19}{20}\frac{1}{100}}{\frac{1}{40}\frac{90}{100} + \frac{1}{40}\frac{9}{100} + \frac{19}{20}\frac{1}{100}} = \frac{\frac{38}{40}}{\frac{90}{40} + \frac{9}{40} + \frac{38}{40}} = \frac{38}{137} \approx 0.277.
\end{aligned}$$

Answer 5 is correct.

Problem 16

The problem description states that $V \sim \mathcal{N}(\mu_V, \sigma_V^2)$ and $W \sim \mathcal{N}(\mu_W, \sigma_W^2)$ follow a bivariate normal distribution with $\mu_V = 1$, $\mu_W = 2$, $\sigma_V = 2$, $\sigma_W = 3$ and correlation coefficient $\rho = -1/4$. Let V^* and W^* denote the standardized versions of V and W, respectively. By the boxed result on p. 454, "Bivariate Normal Distribution", we find that

$$V = \sigma_V V^* + \mu_V = 2V^* + 1 \text{ and } W = \sigma_W W^* + \mu_W = 3W^* + 2.$$

Furthermore, according to the theorem, (V^*, W^*) follows a standard bivariate normal distribution with correlation coefficient $\rho = -1/4$. We now introduce the auxiliary random variable Z such that $W^* = \rho V^* + \sqrt{1-\rho^2} Z$, where V^* and Z are independent standard normal random variables, cf. the "Standard Bivariate Normal Distribution" on p. 451. Hence,

$$\mathbb{P}(V - W \leq 0) = \mathbb{P}(2V^* + 1 - (3W^* + 2) \leq 0) = \mathbb{P}(2V^* - 3W^* \leq 1)$$

can be recast as

$$\mathbb{P}(V - W \leq 0) = \mathbb{P}\left(2V^* - 3\left(-\frac{1}{4}V^* + \sqrt{\frac{15}{16}}Z\right) \leq 1\right) = \mathbb{P}\left(\frac{11}{4}V^* - \frac{3\sqrt{15}}{4}Z \leq 1\right).$$

As V^* and Z are independent standard normal variables, any linear combination of V^* and Z is normally distributed, cf. "Sums of Independent Normal Variables" on p. 363. Therefore, let a second auxiliary random variable, say X, be defined as $X = (11/4)V^* - (3\sqrt{15}/4)Z \sim \mathcal{N}(\mu_X, \sigma_X^2)$, where

$$\mu_X = \mathbb{E}[X] = \mathbb{E}\left[\frac{11}{4}V^* - \frac{3\sqrt{15}}{4}Z\right] = \frac{11}{4}\mathbb{E}[V^*] - \frac{3\sqrt{15}}{4}\mathbb{E}[Z] = 0,$$

$$\sigma_X^2 = \mathbb{V}[X] = \mathbb{V}\left[\frac{11}{4}V^* - \frac{3\sqrt{15}}{4}Z\right] = \left(\frac{11}{4}\right)^2 \mathbb{V}[V^*] + \left(\frac{3\sqrt{15}}{4}\right)^2 \mathbb{V}[Z] = \left(\frac{11}{4}\right)^2 + \left(\frac{3\sqrt{15}}{4}\right)^2 = 16,$$

since the covariance between independent random variables is zero. Thus, $X \sim \mathcal{N}(0, 4^2)$, which implies that $X^* = X/4 \sim \mathcal{N}(0,1)$. In conclusion:

$$\mathbb{P}(V - W \leq 0) = \mathbb{P}(X \leq 1) = \mathbb{P}\left(X^* \leq \frac{1}{4}\right) = \Phi\left(\frac{1}{4}\right).$$

Answer 3 is correct.

Problem 17

To find the probability density function of $Z = Y/X$, we shall invoke the the result:

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_{XY}(x, xz) dx$$

from p. 383. Before doing so, we need to establish the proper integration limits. From the problem description we have that $0 < x < y < 1$. Inserting $z = y/x$, we obtain that $y = xz > x$, which leads to $z > 1$, i.e. the random variable Z is positive. Therefore, $y = xz < 1$ implies that $x < 1/z$. Conclusively, we should integrate x over the interval $[0, 1/z]$. Thus,

$$f_Z(z) = \int_0^{1/z} |x| f_{XY}(x, xz) dx = \int_0^{1/z} 2x dx = 2\left[\frac{1}{2}x^2\right]_0^{1/z} = \frac{1}{z^2}, \ \ z > 1.$$

Answer 2 is correct.

Problem 18

Let R denote the event that the traveller finds a suitable restaurant and let E denote the event that the traveller can be served and enjoy a meal prior to departure. We apply the "Multiplication Rule" on p. 37 to find the probability that the traveller finds a suitable restaurant and can be served and enjoy a meal before departure as

$$\mathbb{P}(ER) = \mathbb{P}(E|R)\mathbb{P}(R) = \frac{1}{2}\frac{4}{5} = \frac{4}{10} = \frac{2}{5}.$$

Answer 3 is correct.

Problem 19

Let X_1, X_2, and X_3 denote the growths of the different colonies, which are independent and identically distributed according to a normal distribution with mean 6 and variance 4. Furthermore, let $X = \max(X_1, X_2, X_3)$. According to the bottom formula on p. 316,

$$\mathbb{P}(X \leq x) = F_{\max}(x) = F_1(x)F_2(x)F_3(x), \ \ x \in \mathbb{R},$$

where F_i refers to the cumulative distribution function for X_i, $i \in \{1, 2, 3\}$. The cumulative distribution function for $Y \sim \mathcal{N}(\mu, \sigma^2)$ is given by $F_Y(y) = \Phi\left((y - \mu)/\sigma\right)$, cf. the distribution summary on p. 477. Since X_1, X_2, and X_3 are identically distributed, they have the same cumulative distribution function, which in this case is $F_1(x) = F_2(x) = F_3(x) = \Phi\left((x - \mu)/\sigma\right)$. Consequently,

$$\mathbb{P}(X > 10) = 1 - \mathbb{P}(X \leq 10) = 1 - F_1(10)F_2(10)F_3(10) = 1 - \Phi\left(\frac{10-6}{2}\right)^3 = 1 - \Phi(2)^3.$$

Answer 3 is correct.

Problem 20

We note that the probability density function takes the form of a trapezoid, which implies that the total area can be calculated as $A = \frac{1}{2}(4+2)h$, where h denotes the height of the trapezoid. Since a probability density function integrates to 1, we can conclude that $h = 1/3$. The probability density function can written as

$$f_X(x) = \begin{cases} \frac{1}{3}(x+2), & -2 \leq x < -1 \\ \frac{1}{3}, & -1 \leq x < 1 \\ -\frac{1}{3}(x-2), & 1 \leq x \leq 2 \end{cases}.$$

In the first interval, the probability density is a linear function with slope 1/3 and a root in -2. In the second interval, the probability density is constant with a value equal to the height of the trapezoid, i.e. 1/3. Finally, in the last interval, the probability density is linear with slope $-1/3$ and a root in 2. The cumulative distribution function can then be found by integrating over the probability density function. In the first interval, the cumulative distribution function becomes

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-2}^{x} \frac{1}{3}(t+2)dt = \frac{1}{3}\int_{0}^{x+2} udu = \frac{1}{3}\left[\frac{1}{2}u^2\right]_0^{x+2} = \frac{1}{6}(x+2)^2, \;\; -2 \leq x < -1.$$

Similarly for the second interval:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-2}^{-1} \frac{1}{3}(t+2)dt + \int_{-1}^{x} \frac{1}{3}dt = \frac{1}{6} + \frac{1}{3}(x+1) = \frac{2x+3}{6}, \;\; -1 \leq x < 1.$$

For the last interval, we get:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{-2}^{-1} \frac{1}{3}(t+2)dt + \int_{-1}^{1} \frac{1}{3}dt + \int_{1}^{x} -\frac{1}{3}(t-2)dt = \frac{6-(x-2)^2}{6}, \;\; 1 \leq x \leq 2.$$

Alternatively, we could have used a symmetry argument to derive the form of the cumulative distribution function in the last interval.

Answer 1 is correct.

Problem 21

Let W_i denote the wire time for package $i \in \{1, 2, 3, 4\}$. According to the problem description, $W_i \sim \text{Exp}(\lambda)$ and $\mathbb{E}[W_i] = 3/2$, which implies that $\lambda = 2/3$, cf. the "Exponential Survival Function" box on p. 279. Since the wire times are independent and exponentially distributed with same intensity, we know that $W_1 + W_2 + W_3 + W_4 \sim \text{Gamma}(4, 2/3)$, cf. "Poisson Arrival Times (Gamma Distribution)" on p. 286. We define $T_4 = W_1 + W_2 + W_3 + W_4$ and use the "Right tail probability" from the same box, which yields:

$$\mathbb{P}(T_4 \leq 8) = 1 - \mathbb{P}(T_4 > 8) = 1 - \sum_{i=0}^{4-1} e^{-\frac{2}{3}\cdot 8}\frac{(2/3 \cdot 8)^i}{i!} = 1 - \sum_{i=0}^{3} e^{-\frac{16}{3}}\frac{(16/3)^i}{i!}.$$

Answer 2 is correct.

Problem 22

Let U_1, U_2, U_3, and U_4 be independent and identically uniformly distributed on the interval $[0,1]$. Furthermore, let $X = \min(U_1, U_2, U_3, U_4)$ and $Y = \max(U_1, U_2, U_3, U_4)$. To solve this problem, first note that

$$\mathbb{P}(Y \leq y) = \mathbb{P}(Y \leq y, X \leq x) + \mathbb{P}(Y \leq y, X > x) = F_{XY}(x,y) + \mathbb{P}(Y \leq y, X > x),$$

which yields that

$$F_{XY}(x,y) = \mathbb{P}(Y \leq y) - \mathbb{P}(Y \leq y, X > x).$$

For a random variable $Z \sim U(0,1)$, the cumulative distribution function is given by $F_Z(z) = z$ in its domain, while probabilities can be evaluated as $\mathbb{P}(a \leq Z \leq b) = b - a$ for $b > a$ and $a, b \in [0,1]$. Using the cumulative distribution functions for maxima and minima of independent random variables on p. 316-317, we get that

$$\mathbb{P}(Y \leq y) = F_{\max}(y) = F_{U_1}(y)F_{U_2}(y)F_{U_3}(y)F_{U_4}(y) = y^4, \ \ y \in [0,1].$$

Furhtermore, we find that

$$\begin{aligned}
\mathbb{P}(Y \leq y, X > x) &= \mathbb{P}(U_1 \leq y, U_2 \leq y, U_3 \leq y, U_4 \leq y, U_1 > x, U_2 > x, U_3 > x, U_4 > x) \\
&= \mathbb{P}(x < U_1 \leq y, x < U_2 \leq y, x < U_3 \leq y, x < U_4 \leq y) \\
&= \mathbb{P}(x < U_1 \leq y)\mathbb{P}(x < U_2 \leq y)\mathbb{P}(x < U_3 \leq y)\mathbb{P}(x < U_4 \leq y) \\
&= (y - x)^4, \ \ x \leq y.
\end{aligned}$$

In conclusion,

$$F_{XY}(x,y) = \mathbb{P}(Y \leq y) - \mathbb{P}(Y \leq y, X > x) = y^4 - (y-x)^4, \ \ x \leq y \leq 1.$$

Answer 4 is correct.

Problem 23

Let X and Y denote the first and second measurement of vitamin D, respectively. We then know that (X, Y) follows a standard bivariate normal distribution with correlation coefficient $\rho = 4/5$. First, we shall apply the "General Formula for $\mathbb{P}(A|B)$" on p. 36:

$$\mathbb{P}(Y < X|X < 0) = \frac{\mathbb{P}(Y < X, X < 0)}{\mathbb{P}(X < 0)},$$

since zero is the mean of X (any standard normal random variable). Furthermore, as non-skewed normal distributions are symmetric around their mean, we get that $\mathbb{P}(X < 0) = 1/2$. To evaluate the numerator in the above expression, we introduce the auxiliary random variable Z according to the box on p. 451. Since (X, Y) has a standard bivariate normal distribution, we can write $Y = \rho X + \sqrt{1-\rho^2}Z$, where X and Z are independent and both distributed according to a standard normal distribution. Thus,

$$\mathbb{P}(Y < X|X < 0) = 2\mathbb{P}\left(\frac{4}{5}X + \sqrt{1-\left(\frac{4}{5}\right)^2}Z < X, X < 0\right) = 2\mathbb{P}\left(\frac{3}{5}Z < \frac{1}{5}X, X < 0\right)$$
$$= 2\mathbb{P}\left(Z < \frac{1}{3}X, X < 0\right).$$

The two conditions in the latter expression describe an area in the (X, Z)-plane and the probability we seek is equal to the proportion of that area to the entire (X, Z)-plane (We recommend that you make a drawing of this area to visualize the solution). This proportion is equivalent to the ratio between the angle span by the area and 2π. Hence, denote now the angle which the area span around the origin by α. By inscribing a right triangle in the area, we see that $\tan(\alpha) = 3$, which implies that $\alpha = \arctan(3)$. The solution is therefore given as

$$\mathbb{P}(Y < X|X < 0) = 2\mathbb{P}\left(Z < \frac{1}{3}X, X < 0\right) = 2\frac{\arctan(3)}{2\pi} = \frac{\arctan(3)}{\pi}.$$

Since we do not find this solution among the possible answers immediately, we invoke the trigonometric identity that $\arctan(x) = \pi/2 - \arctan(x^{-1})$ for $x > 0$. Thus,

$$\mathbb{P}(Y < X|X < 0) = \frac{\arctan(3)}{\pi} = \frac{\pi/2 - \arctan(1/3)}{\pi} = \frac{1}{2} - \frac{\arctan(1/3)}{\pi}.$$

Alternatively, the solution can be found with the same methods by considering the angle spanned between the area and the first axis in the (X, Z)-plane.

Answer 5 is correct.

Problem 24

Let X denote the waiting time until the first toss that yields a heads (number of tosses). Since all tosses are independent and there is an identical probability that the coin lands heads in each toss, the tosses constitute a sequence of Bernoulli trials. As there is a 50% chance that the coin lands heads, we conclude that $X \sim \text{Geo}(1/2)$. We then define the random variable Y as the number of tosses in the game described in the problem. We know that Y is the number of tosses until the first heads, but censored at four tosses, i.e. the coin is tossed a maximum of four times. Consequently, we conclude that $Y = \min(X, 4)$, i.e. Y is a function of X. We then apply "Expectation of a Function of X" from p. 175:

$$\mathbb{E}[Y] = \mathbb{E}[\min(X,4)] = \sum_{x=1}^{\infty} \min(x,4)\mathbb{P}(X=x) = \mathbb{P}(X=1)+2\mathbb{P}(X=2)+3\mathbb{P}(X=3)+4\sum_{x=4}^{\infty}\mathbb{P}(X=x).$$

We first note that $\sum_{x=4}^{\infty}\mathbb{P}(X=x) = 1-\sum_{x=1}^{3}\mathbb{P}(X=x)$ and use the probability mass function of the geometric distribution to calculate the probabilities in the above equation. Thus,

$$\begin{aligned}
\mathbb{E}[Y] &= \frac{1}{2} + 2\frac{1}{2}\left(1-\frac{1}{2}\right) + 3\frac{1}{2}\left(1-\frac{1}{2}\right)^2 + 4\left(1-\frac{1}{2}-\frac{1}{2}\left(1-\frac{1}{2}\right)-\frac{1}{2}\left(1-\frac{1}{2}\right)^2\right) \\
&= \frac{1}{2}+\frac{2}{4}+\frac{3}{8}+4\left(1-\frac{1}{2}-\frac{1}{4}-\frac{1}{8}\right) = \frac{11}{8}+\frac{4}{8} = \frac{15}{8}.
\end{aligned}$$

Answer 1 is correct.

Problem 25

Firstly, note that the only way to get exactly 1.75 kr. with three coins is by choosing one of each type of coin. The number of ways we can select one of each type of coin is given as

$$\binom{5}{1}\binom{5}{1}\binom{5}{1} = \binom{5}{1}^3 = \left(\frac{5!}{(5-1)!1!}\right)^3 = 5^3 = 125.$$

Similarly, the number of ways we can select three coins out of the fifteen is given by

$$\binom{15}{3} = \frac{15!}{(15-3)!3!} = 455.$$

If we denote the event that we select coins worth exactly 1.75 kr. by A, we can calculate the probability of A with the "Equally likely Outcomes" formula on p. 3:

$$\mathbb{P}(A) = \frac{125}{455} = \frac{25}{91}.$$

Alternatively, we can give some heuristic arguments for this solution. The first coin can be any of the fifteen coins. The second coin has to be of a different type than the first, which leaves 10 out of the 14 remaining coins. The final coin has to be of a different type than the two previous coins, which leaves 5 out of the remaining 13. Hence, $\mathbb{P}(A) = 1 \cdot 10/14 \cdot 5/13 = 25/91$.

Answer 1 is correct.

Problem 26

We define the 5 independent, standard normal random variables as $X_1,...,X_5$. Recall that for standard normal random variable, say X, the probability density function and the cumulative distribution function are given as

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \text{ and } F_X(x) = \Phi(x).$$

We apply the "Density of the k'th Order Statistic" on p. 326 to formulate the probability density function of the third order statistic of $X_1,...,X_5$:

$$f_{(3)}(x) = 5\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \binom{4}{2} \Phi(x)^2 \left(1 - \Phi(x)\right)^2 = 30\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Phi(x)^2 \left(1 - \Phi(x)\right)^2.$$

Answer 1 is correct.

Problem 27

We can find the conditional probabilities by invoking the "General Formula for $\mathbb{P}(A|B)$" on p. 36:

$$\mathbb{P}(X = x|Y = 1) = \frac{\mathbb{P}(X = x, Y = 1)}{\mathbb{P}(Y = 1)}.$$

The numerator in the above expression is given in the problem, while the denominator can be evaluated using "Marginal Probabilities" on p. 145:

$$\mathbb{P}(Y = 1) = \sum_{x=0}^{2} \mathbb{P}(X = x, Y = 1) = \frac{3!}{2!}\left(\frac{1}{4}\right)^3 + 3!\left(\frac{1}{2}\right)\left(\frac{1}{4}\right)^2 + \frac{3!}{2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{4}\right) = \frac{3}{64} + \frac{3}{16} + \frac{3}{16} = \frac{27}{64}.$$

We sum from zero to two because X is a non-negative random variable and $X + Y \leq 3$. We then calculate the probabilities as follow:

$$\mathbb{P}(X = 0|Y = 1) = \frac{\mathbb{P}(X = 0, Y = 1)}{\mathbb{P}(Y = 1)} = \frac{\frac{3!}{2!}\left(\frac{1}{4}\right)\left(\frac{1}{4}\right)^2}{\frac{27}{64}} = \frac{\left(\frac{3}{64}\right)}{\left(\frac{27}{64}\right)} = \frac{3}{27} = \frac{1}{9}$$
$$\mathbb{P}(X = 1|Y = 1) = \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(Y = 1)} = \frac{3!\left(\frac{1}{2}\right)\left(\frac{1}{4}\right)\left(\frac{1}{4}\right)}{\frac{27}{64}} = \frac{\left(\frac{3}{16}\right)}{\left(\frac{27}{64}\right)} = \frac{12}{27} = \frac{4}{9}$$
$$\mathbb{P}(X = 2|Y = 1) = \frac{\mathbb{P}(X = 2, Y = 1)}{\mathbb{P}(Y = 1)} = \frac{\frac{3!}{2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{4}\right)}{\frac{27}{64}} = \frac{\left(\frac{3}{16}\right)}{\left(\frac{27}{64}\right)} = \frac{12}{27} = \frac{4}{9}.$$

Alternatively, one can recognize that the probability mass function given in the problem description is equivalent with a multinomial distribution, specifically $\text{MN}(3, 1/2, 1/4, 1/4)$. Therefore, given $Y = 1$, we know that $X \sim \text{Bin}(2, 2/3)$, which yields the same probabilities.

Answer 1 is correct.

Problem 28

From the problem description we know that (X, Y) follows a standard bivariate normal distribution with correlation coefficient $\rho = 3/5$. According to the "Standard Bivariate Normal Distribution" box on p. 451, the conditional distribution of X given $Y = y$ is a normal distribution with mean ρy and variance $1 - \rho^2$, i.e. given $Y = y$, $X \sim \mathcal{N}(\rho y, 1 - \rho^2)$. Thus, for $Y = 1$, $X \sim \mathcal{N}(3/5, 16/25)$. Hence, if we define $Z \sim \mathcal{N}(3/5, 16/25)$ and denote the standardized version by Z^*, we get

$$\mathbb{P}(X \leq 1|Y = 1) = \mathbb{P}(Z \leq 1) = \mathbb{P}\left(\frac{Z - 3/5}{\sqrt{16/25}} \leq \frac{1 - 3/5}{\sqrt{16/25}}\right) = \mathbb{P}\left(Z^* \leq \frac{1}{2}\right) = \Phi\left(\frac{1}{2}\right).$$

Answer 3 is correct.

Problem 29

We shall apply Chebychev's inequality (p. 191) for $X - Y$. Therefore, we calculate the expectation and the standard deviation of the difference:

$$\mathbb{E}[X - Y] = \mathbb{E}[X] - \mathbb{E}[Y] = 0 \text{ and } \mathbb{V}[X - Y] = \mathbb{V}[X] + \mathbb{V}[Y] = 8 + 8 = 16.$$

The first result follows from the fact that X and Y are identically distributed, while the second result follows from the independence between X and Y, which implies that the covariance between X and Y is zero. Hence, $\mathrm{SD}(X - Y) = \sqrt{16} = 4$. Consequently,

$$\mathbb{P}(|X - Y - \mathbb{E}[X - Y]| \geq 20) = \mathbb{P}(|X - Y| \geq 20) = \mathbb{P}(|X - Y| \geq 5 \cdot \mathrm{SD}(X - Y)) \leq \frac{1}{5^2} = \frac{1}{25}.$$

Answer 3 is correct.

Problem 30

From the problem we have that $0 < x/2 < y < 2x$. Furthermore, for the probability of interest, we have that $x \leq 1$ and $y \leq 2$ (We recommend that you make a drawing to visualize the solution). We can paramterize the area of interest by: $x \in [0, 1]$ and $y \in [x/2, 2x]$. Hence,

$$\begin{aligned}
\mathbb{P}(X \leq 1, Y \leq 2) &= \int_0^1 \int_{x/2}^{2x} f_{XY}(x, y) dy dx = \int_0^1 \int_{x/2}^{2x} 3e^{-(x+y)} dy dx = 3 \int_0^1 e^{-x} \int_{x/2}^{2x} e^{-y} dy dx \\
&= 3 \int_0^1 e^{-x}\left(e^{-x/2} - e^{2x}\right) dx = 3 \int_0^1 e^{-3x/2} - e^{-3x} dx = 2 - 2e^{-3/2} + e^{-3} - 1 \\
&= e^{-3} + 1 - 2e^{-3/2} = \left(1 - e^{-3/2}\right)^2.
\end{aligned}$$

Answer 4 is correct.