Probability Theory
2024/25 Period IIb
Instructor: Gilles Bonnet
Resit
8/7/2024
Duration: 2 hours

Solutions

Question:	1	2	3	4	5	Total
Points:	12	13	18	20	27	90
Score:						

1. (12 points) A machine produces 6-sided dice. The machine is defective: while 99.9% of the dice it produces are normal, the remaining 0.1% have all their faces marked 6. Suppose I take (at random) a die produced by this machine and roll it n times, and then I inform you that all the rolls resulted in 6. For which values of n is it more likely that I took a defective die than that I took a normal die?

Solution:

$$\mathbb{P}(\text{Defective die} \mid n \text{ times } 6) = \frac{0.001 \cdot 1}{0.001 \cdot 1 + 0.999 \cdot (1/6)^n} = \frac{1}{1 + 999/6^n}.$$

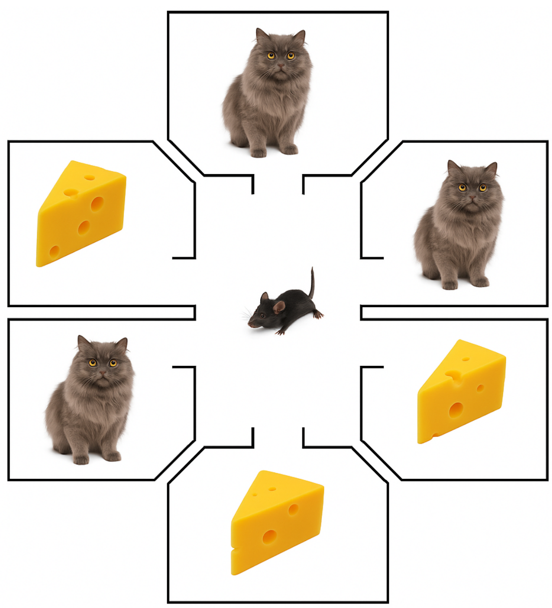
This is larger than $\frac{1}{2}$ if $6^n > 999$, which holds for $\boxed{n \geq 4}$.

2. (13 points) The continuous random variables Z and W are independent, with Z following the exponential distribution with parameter 1 and W following the (continuous) uniform distribution on $(0, 1)$. Compute $\mathbb{P}(Z < W < 3Z)$.

Solution:

$$\begin{aligned}
\mathbb{P}(Z < W < 3Z) &= \int\int \mathbf{1}_{\{z<w<3z\}} f_{Z,W}(z,w)\, dwdz \\
&= \int\int \mathbf{1}_{\{z<w<3z\}} f_Z(z) f_W(w)\, dwdz \qquad \text{by independence} \\
&= \int_0^\infty \int_0^1 \mathbf{1}_{\{z<w<3z\}} e^{-z}\, dwdz \\
&= \int_0^{\frac{1}{3}} \int_z^{3z} e^{-z}\, dwdz + \int_{\frac{1}{3}}^1 \int_z^1 e^{-z}\, dwdz \\
&= \int_0^{\frac{1}{3}} 2ze^{-z} dz + \int_{\frac{1}{3}}^1 (1-z)e^{-z}\, dz \\
&= -3e^{-\frac{1}{3}} + 2 + e^{-1}.
\end{aligned}$$

3. A forgetful mouse is subjected to an experiment. It is placed inside the central room of a maze that has 7 rooms, 3 cats and 3 cheeses as shown below. The mouse always moves into a room chosen uniformly at random from all rooms adjacent to the room it is in, completely independently of all its previous choices.



The experimenters will only remove the mouse from the experiment if it's found all three cheeses. Of course, whenever our mouse enters a room with a cat it will play with the cat and will not leave that room ever again.

(a) (6 points) What is the probability the mouse will find at least one of the cheeses before entering a cat's room?

Solution: Let E_i be the event that the mouse finds at least i cheeses before entering the cat's room. We want to compute $\mathbb{P}(E_1)$. Each of the 6 rooms occupied by either a cheese or the cat are equally likely to be visited first, so $\boxed{\mathbb{P}(E_1) = \frac{3}{6}}$.

(b) (6 points) Given that the mouse has found one cheese before it meets the cat, what is the probability that the mouse will find a second cheese before entering a cat's room?

Solution: We want to compute $\mathbb{P}(E_2 \mid E_1)$. Given that the mouse has found one cheese, there are 5 occupied rooms left (2 cheeses and 3 cats), and the mouse is equally likely to visit any of them. Thus, $\boxed{\mathbb{P}(E_2 \mid E_1) = \frac{2}{5}}$.

(c) (6 points) What is the chance the mouse will find the three cheeses before meeting a cat?

Solution: We a similar argument as in the previous question, we find $\mathbb{P}(E_3 \mid E_2) = \frac{1}{4}$, since after finding two cheeses, there are 4 rooms left (1 cheese and 3 cats), and the mouse is equally likely to visit either of them. Thus, we have

$$\mathbb{P}(E_3) = \mathbb{P}(E_3 \mid E_2) \cdot \mathbb{P}(E_2 \mid E_1) \cdot \mathbb{P}(E_1) = \frac{1}{4} \cdot \frac{2}{5} \cdot \frac{3}{6} = \frac{1}{20}.$$

Alternative direct solution: We can consider the similar experiment where the mouse does not stop when meeting a cat. Eventually the mouse will visit all the rooms at least once. Now we can take note of in which order the rooms are visited for the first times, for instance $Cat_2, Cheese_1, Cat_1, Cheese_2, Cheese_3, Cat_3$. We are interested in the probability that the three first rooms are the ones with cheeses.

There are 6! possible orders of visiting the 6 rooms, out of which $3! \times 3!$ correspond to the orders where the first three rooms visited are all cheeses. The probability that the mouse finds all three cheeses before meeting a cat is the same as the probability that the first three rooms visited are all cheeses. So the probability is $\frac{3! \times 3!}{6!} = \frac{3\times2\times1}{6\times5\times4} = \frac{1}{20}$.

Comment: Again, a serious student would have recognized that this is a small variation of a tutorial exercise, namely Exercise 14.3.

4. (20 points) Let $Z_1 \sim \mathcal{N}(0,1)$ and $Z_2 \sim \mathcal{N}(0,1)$ be independent standard normal random variables and

$$U = \sigma_1 Z_1 + \mu_1, \qquad V = \rho\sigma_2 Z_1 + \sqrt{1-\rho^2} \cdot \sigma_2 Z_2 + \mu_2.$$

Show that (U, V) follows a bivariate normal distribution (and write down the parameters of this distribution).

Solution: See Lemma 15.1.1 and its proof in the lecture notes. (You need to apply the Theorem 10.1.2 which gives you the joint pdf of the image of a random vector under bijective differentiable function.)

5. We have m urns and n balls, where $m \geq 2$ and n are integer numbers. We place the balls successively into the urns, so that any given ball is equally likely to go into any urn. Each placement is independent of the other ones.

   (a) (15 points) Let X and Y be the number of balls that go into urn 1 and 2, respectively. Compute $\mathrm{Cov}(X, Y)$.

   Hint. It can help to write X in the form $X = \sum_{i=1}^{n} X_i$, where X_i is the indicator that ball i goes into urn 1; and similarly for Y. Then you might want to compute $\mathrm{Cov}(X_i, Y_j)$ for $i \neq j$ and $i = j$ separately and use this to answer the question.

Solution: (Those who study correctly would have recognize that this is a small variation of a tutorial exercise, namely Exercise 10.3.)

As per the hint, we can write

$$X = \sum_{i=1}^{n} X_i, \quad Y = \sum_{i=1}^{n} Y_i,$$

where X_i is the indicator variable that ball i goes into urn 1, and Y_i is the indicator variable that ball i goes into urn 2. Then

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j) \\ &= \sum_{i=1}^{n} \mathrm{Cov}(X_i, Y_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, Y_j)\end{aligned}$$

Note that for $i \neq j$, X_i and Y_j are independent, so $\mathrm{Cov}(X_i, Y_j) = 0$. For $i = j$, we have $\mathrm{Cov}(X_i, Y_i) = \mathbb{E}[X_i Y_i] - \mathbb{E}[X_i]\mathbb{E}[Y_i]$. Since X_i and Y_i are indicators for the same ball going into urn 1 and urn 2, respectively, we have $\mathbb{E}[X_i Y_i] = 0$ (the same ball cannot go into both urns). Also, $\mathbb{E}[X_i] = \mathbb{E}[Y_i] = \frac{1}{m}$, so

$$\mathrm{Cov}(X_i, Y_i) = 0 - \frac{1}{m} \cdot \frac{1}{m} = -\frac{1}{m^2}.$$

Therefore, we have

$$\mathrm{Cov}(X, Y) = n \cdot \left(-\frac{1}{m^2}\right) + 0 = \boxed{-\frac{n}{m^2}}.$$

   (b) (6 points) Compute the variance of $X + Y$.

Solution: We observe that $X + Y$ is a binomial random variable with parameters n and $\frac{2}{m}$, since each ball can go into either urn 1 or urn 2 with probability $\frac{1}{m}$ each.

Therefore, we have

$$\mathrm{Var}(X+Y) = n \cdot \frac{2}{m}\left(1-\frac{2}{m}\right) = n \cdot \frac{2}{m} - n \cdot \frac{4}{m^2} = \frac{2n}{m} - \frac{4n}{m^2}.$$

Alternatively, more convoluted solution: Both X and Y are binomial random variables with parameters n and $\frac{1}{m}$, so we have $\mathrm{Var}(X) = \mathrm{Var}(Y) = n \cdot \frac{1}{m}\left(1-\frac{1}{m}\right)$. Thus, we have

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y) \\
&= n \cdot \frac{1}{m}\left(1-\frac{1}{m}\right) + n \cdot \frac{1}{m}\left(1-\frac{1}{m}\right) + 2\left(-\frac{n}{m^2}\right) \\
&= 2n \cdot \frac{1}{m}\left(1-\frac{1}{m}\right) - \frac{2n}{m^2} \\
&= \frac{2n}{m}\left(1-\frac{2}{m}\right).
\end{aligned}$$

So the final answer is $\boxed{\frac{2n}{m} - \frac{4n}{m^2}}$.

(c) (6 points) Now, assume that there are $n = 2000$ balls and $m = 1000$ urns. Show that the probability that there are at least a total of 24 balls in total in the first two urns is less than 1%.
Remark: If you show a (correct) stronger bound, this is obviously also fine.

Solution: We note that $\mathbb{E}(X+Y) = 2n/m = 4$, and $\mathrm{Var}(X+Y) = 2n/m - 4n/m^2 = 4 - \frac{8}{1000} \leq 4$.

$$\begin{aligned}
\mathbb{P}(X+Y \geq 24) &= \mathbb{P}(X+Y-\mathbb{E}(X+Y) \geq 20) \\
&\leq \mathbb{P}\left(|X+Y-\mathbb{E}(X+Y)| \geq 20\right) \\
&\leq \frac{\mathrm{Var}(X+Y)}{20^2} && \text{(Chebyshev's inequality)} \\
&\leq \frac{4}{400} = 0.01.
\end{aligned}$$

Alternatively, using the Central Limit Theorem: We can apply the Central Limit Theorem, since $X+Y$ is the sum of a large number of independent and identically distributed random variables. The mean is $\mathbb{E}[X+Y] = 4$ and the variance is $\mathrm{Var}(X+Y) = 4 - \frac{8}{1000} = 3.992$. By the Central Limit Theorem, we can approximate $X+Y$

by a normal distribution with mean $\mu = 4$ and variance $\sigma^2 = 3.992 \leq 2^2$.

$$
\begin{aligned}
\mathbb{P}(X+Y \geq 24) &= \mathbb{P}\left(\frac{X+Y-\mu}{\sigma} \geq \frac{24-4}{\sigma}\right) \\
&\leq \mathbb{P}\left(\frac{X+Y-\mu}{\sigma} \geq \frac{24-4}{2}\right) \\
&= \mathbb{P}\left(\frac{X+Y-\mu}{\sigma} \geq 10\right) \\
&\approx \mathbb{P}\left(Z \geq 10\right) && \text{(where } Z \text{ is standard normal)} \\
&\leq \mathbb{P}(Z \geq 3.49) && \text{(very crude estimate)} \\
&\simeq 0.0002 && \text{(from standard normal table)} \\
&< 0.01. && \text{(again very crude estimate)}
\end{aligned}
$$