Question 1 On a multiple choice exam there are 100 questions each with 4 possible answers. A student is certain of the correct answer to each question with probability 0.6 and guesses randomly among the four choices otherwise.

(a) [2 points] What is the probability that the student correctly answers question 1.
(b) [3 points] What is the probability that the student was certain of the answer to question 1 given that they got it correct.
(c) [3 points] Find, approximately, the probability that the student gets at least 80 questions correct on the test.

Solutions

(a) The probability the the student gets the question correct is

\[
P[\text{Correct}] = P[\text{Correct} | \text{Certain of answer}]P[\text{Certain of answer}] + P[\text{Correct} | \text{Guesses randomly}]P[\text{Guesses randomly}]
\]

\[
= 1 \cdot 0.6 + \frac{1}{4} \cdot 0.4 = 0.7.
\]

(b)

\[
P[\text{Certain of answer} | \text{Correct}] = \frac{P[\text{Certain of answer, Correct}]}{P[\text{Certain of answer}]}
\]

\[
= \frac{0.6}{0.7} = \frac{6}{7}.
\]
(c) The number of correct solutions is Bin(100, 0.7). By a normal approximation, putting it into standard units,

\[ \mathbb{P}[\#\text{Correct} \geq 80] \approx \mathbb{P}[N(0, 1) \geq \frac{80 - 100 \cdot 0.7}{\sqrt{100 \cdot 0.7 \cdot (1 - 0.7)}}] = 1 - \Phi\left(\frac{10}{\sqrt{21}}\right). \]

Or \(1 - \Phi\left(\frac{10.5}{\sqrt{21}}\right)\) with the continuity correction.
Question 2
In a class of 20 students each student tosses a fair coin 4 times independently.
(a) [2 points] What is the probability that no student gets 4 heads.
(b) [3 points] What is the expected number of students who get 4 heads.

Solutions
(a) The chance that a single student get 4 heads is \((1/2)^4 = 1/16\). Since each students tosses are independent the probability that no student gets 4 heads is

\[ P[\text{No student 4 heads}] = (1 - 1/16)^{20}. \]

(b) Let \(X_i\) be the indicator that student \(i\) gets 4 heads so \(P[X_i = 1] = \mathbb{E}X_i = 1/16\). Let \(X\) be the total number of students with 4 heads so \(X = \sum_{i=1}^{20} X_i\) and

\[ \mathbb{E}X = \sum_{i=1}^{20} \mathbb{E}X_i = 20 \cdot \frac{1}{16} = \frac{5}{4}. \]
Question 3
A standard deck of cards has 52 cards, 4 of each type (Ace, King, Queen, Jack, 10,...,2). From a well shuffled deck, you are dealt a hand of 5 cards (without replacement).
(a) [2 points] What is the probability that you are dealt at least one face card (that is a king, queen or jack)?
(b) [3 points] What is the probability that you are dealt both at least one ace and at least one face card?

Solutions
(a) There are 12 face cards so by the sampling without replacement formula

\[
P[\text{No Face card}] = \frac{\binom{40}{5}}{\binom{52}{5}} = \frac{40}{5}.
\]

So the probability of at least one face card is \(1 - \frac{40}{5} \frac{5}{52} \frac{52}{4}\).

(b) Again by the sampling without replacement formula

\[
P[\text{No Ace or Face card}] = \frac{\binom{36}{5}}{\binom{52}{5}} = \frac{26}{5}.
\]

So the probability of at least one ace or one face card is \(1 - \frac{36}{5} \frac{4}{52} \frac{52}{4}\). Then

\[
P[\text{An Ace and a Face card}] = P[\text{An Ace}] + P[\text{A Face card}] - P[\text{An Ace or a Face card}]
\]

\[
= 1 - \frac{48}{5} + 1 - \frac{40}{5} - 1 (1 - \frac{36}{5}).
\]
Question 4
Let $X$ be uniform $[0, 1]$ and let $Y$ be an independent random variable uniform on $[0, 2]$.

(a) [3 points] Find the density of $W = -\log(X)$ and identify the distribution.

(b) [3 points] Find the density of $Z = X + Y$ and sketch it.

Solutions

(a) Denote the function $g(x) = -\log(x)$ which is one-to-one on $[0, 1]$ and has range $[0, \infty)$. Then by the change of variable formula for $x \in [0, 1]$,

$$f_W(g(x)) = \frac{f_X(g(x))}{|g'(x)|} = \frac{1}{|\frac{-1}{x}|} = x.$$

If $w = g(x)$ then $x = e^{-w}$ and so

$$f_W(w) = e^{-w}.$$

Hence $Z$ is a rate one exponential.

(b) The joint density of $(X, Y)$ is $\frac{1}{2}I(0 < x < 1, 0 < y < 2)$. So by the summation formula

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{2}I(0 < x < 1, 0 < z-x < 2)dx$$

$$= \int_{\max\{0,z-2\}}^{\min\{1,z\}} \frac{1}{2}dx$$

$$= \begin{cases} 
\int_{0}^{\frac{1}{2}} \frac{1}{2}dx & 0 \leq z \leq 1 \\
\int_{0}^{1} \frac{1}{2}dx & 1 \leq z \leq 2 \\
\int_{z-2}^{1} \frac{1}{2}dx & 2 \leq z \leq 3 \\
\end{cases}$$

$$= \begin{cases} 
\frac{z}{2} & 0 \leq z \leq 1 \\
\frac{1}{2} & 1 \leq z \leq 2 \\
\frac{3-z}{2} & 2 \leq z \leq 3 \\
\end{cases}$$
The joint probability density function of \( X \) and \( Y \) is given by \( f(x, y) = C(x + y) \) for \( 0 \leq x \leq 1, 0 \leq y \leq 1 \) and 0 otherwise where \( C \) is a constant.

(a) [3 points] Find the constant \( C \) that ensures that \( f(x, y) \) is a probability density.

(b) [3 points] Find \( f_X(x) \), the marginal density of \( X \).

(c) [3 points] Find the condition distribution of \( Y \) given \( X = x \).

(d) [2 points] Are \( X \) and \( Y \) independent?

Solutions

(a) Since a probability density must integrate to 1,

\[
1 = \int_0^1 \int_0^1 C(x + y) \, dx \, dy = \int_0^1 \frac{1}{2}C + Cy \, dy = \frac{1}{2}C + \frac{1}{2}C = C
\]

so \( C = 1 \).

(b) We can get the marginal density of \( X \) by integrating out \( y \),

\[
f_X(x) = \int_0^1 x + y \, dy = \frac{1}{2} + x.
\]

(c) The conditional distribution of \( Y \) given \( X = x \) is,

\[
f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x + y}{\frac{1}{2} + x}.
\]

(d) The conditional distribution of \( Y \) given \( X = x \) depends on \( x \) so they are not independent.
Question 6
Let $X$ be a uniform random variable on $[0, 1]$ and conditional on $X = x$ let $Y$ be an exponential with rate $1/x$.

(a) [3 points] Find the joint density of $X$ and $Y$.

(b) [3 points] Find $EY$.

(c) [3 points] Find $\text{Var}(Y)$.

Solutions
(a) We have that $f_X(x) = 1$ for $0 < x < 1$ and the conditional density of $Y$ given $X$ is $f_{Y|X}(y|x) = \frac{1}{x} e^{-y/x}$. Then the joint density is

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x) = \frac{1}{x} e^{-y/x}$$

for $0 < x < 1$ and $y \geq 0$.

(b) We could compute $\int \int y f_{X,Y}(x,y) dxdy$ but it is easier to use the conditional distribution of $Y$ given $X$. Since $Y$ is a rate $1/X$ exponential conditional on $X$ we have that $E[Y | X = x] = x$. Hence

$$EY = E[E[Y | X]] = E[X] = \frac{1}{2}$$

since $X$ is a uniform on $[0, 1]$ and so has expected value $\frac{1}{2}$.

(c) If $W$ is an exponential of rate $\lambda$ then $W$ has expected value $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$ and so $E[W^2] = \frac{2}{\lambda^2}$. Since $Y$ is a rate $1/X$ exponential conditional on $X$ we have that $E[Y^2 | X = x] = 2x^2$. Hence

$$EY^2 = E[E[Y^2 | X]] = E[2X^2] = \int_0^1 2x^2 dx = \frac{2}{3}.$$  

Then

$$\text{Var}[Y] = EY^2 - [EY]^2 = \frac{2}{3} - \frac{1}{4} = \frac{5}{12}.$$
Question 7

The midterm and final exam scores \((X, Y)\) of students in a class have a joint normal distribution with means \(\mu_x = \mu_y = 70\) and \(\sigma_x^2 = \sigma_y^2 = 25\) and correlation \(\rho = 0.8\).

(a) [3 points] Find the probability that a student who scored 80 in the first test scored above average in the second.

(b) [3 points] What is the distribution of the improvement from the midterm to the final, that is \(I = Y - X\)?

(c) [3 points] The final score for students is \(Z\), the average score on the two tests. Show that the improvement \(I\) and the final score \(Z\) are independent.

Solutions

We want to calculate the probability that \(Y \geq 70\) given that \(X = 80\). By writing \((X, Y)\) in standard units,

\[
X^* = \frac{X - 70}{5}, \quad Y^* = \frac{Y - 70}{5}
\]

we want to calculate

\[
\mathbb{P}[Y^* > 0 \mid X^* = \frac{80 - 70}{5}] = \mathbb{P}[Y^* > 0 \mid X^* = 2].
\]

Since \(X^*\) and \(Y^*\) have standard bivariate normal distribution with correlation \(\rho = 0.8\) we have that the conditional distribution of \(Y^*\) given \(X^* = 2\) is \(N(\rho \cdot 2, 1 - \rho^2)\),

\[
\mathbb{P}[Y^* > 0 \mid X^* = 2] = \mathbb{P}[N(0.8 \cdot 2, 1 - (0.8)^2) > 0]
\]

\[
= \mathbb{P}[N(1.6, 0.36) > 0] = \mathbb{P}[N(0, 1) > -\frac{1.6}{0.6}]
\]

\[
= 1 - \Phi(-\frac{1.6}{0.6}) = \Phi\left(\frac{1.6}{0.6}\right)
\]
Question 8

People arrive at a store according to a Poisson process of rate 2 per minute. Let $T_i$ denote the arrival time in minutes of the i’th customer.

(a) [3 points] What is the expected value of the third arrival $T_3$.

(b) [2 points] What is the expected value of the third arrival conditional on no customers arriving by time 2?

(c) [3 points] What is the probability that the third arrival $T_3$ is more than 1 minute after $T_1$.

Bonus Questions

The store has 10 types of products. Suppose that each customer buys exactly one of the 10 types of products chosen uniformly at random. Let $M$ be the number of different products purchased by at least one customer in the first 5 minutes.

(d) [3 points] Find $EM$.

(e) [4 points] Find $\text{Var}(M)$.

Solutions

(a) The third arrival time has distribution Gamma(3, 2). We can calculate the expected value of this directly or not that if $W_i = T_i - T_{i-1}$ then the $W_i$ are independent exponentials with rate 2. Hence

$$\mathbb{E}T_3 = \sum_{i=1}^{3} \mathbb{E}W_i = 3 \cdot \frac{1}{2} = \frac{3}{2}.$$  

(b) By the memoryless property of the exponential distribution, conditional on no arrivals in the first 2 minutes, the time until the first arrival is 2 plus a rate 2 exponential

$$\mathbb{E}[T_3 \mid T_1 > 2] = \sum_{i=1}^{3} \mathbb{E}[W_i \mid T_1 > 2] = (2 + \frac{1}{2}) + \frac{1}{2} + \frac{1}{2} = \frac{7}{2}.$$  

(c) The time between the first and third arrivals $T_3 - T_1$ has the same distribution of $T_2$ which is a Gamma(2, 2) and which has density $4te^{-2t}$ so,

$$\mathbb{P}[T_3 - T_1 > 1] = \int_1^\infty 4te^{-2t} dt = -2te^{-2t} \big|_1^\infty - \int_1^\infty -2e^{-2t} dt = 2e^{-2} - e^{-2t} \big|_1^\infty = 3e^{-2}.$$  

(d) Let $X$ be the total number of arrivals by time 5 and let $X_i$ be the number of those customers who bought product $i$ for $1 \leq i \leq 10$. The distribution of $X$ is Poisson with mean 10 and by Poisson thinning the $X_i$ are independent and have distribution Poisson with mean 1. Let $M_i$ be the indicator that $X_i \geq 1$, that is that at least one product of type $i$ was purchased. Then

$$\mathbb{E}M_i = \mathbb{P}[X_i \geq 1] = 1 - \mathbb{P}[X_i = 0] = 1 - e^{-1}.$$  

Then $M$ the number of different types of products sold is $\sum M_i$ and so

$$\mathbb{E}M = \sum_{i=1}^{10} \mathbb{E}M_i = 10(1 - e^{-1}).$$
(e) Since the $X_i$ are independent so are the $M_i$.

$$\text{Var}[M_i] = \mathbb{E}M_i^2 - (\mathbb{E}M_i)^2 = (1 - e^{-1}) - (1 - e^{-1})^2 = e^{-1}(1 - e^{-1})$$

Since the $X_i$ are independent so are the $M_i$ and hence since the sum of the variances is the variance of the sum for independent random variables

$$\text{Var}[M] = \sum_{i=1}^{10} \text{Var}[M_i] = 10\text{Var}[M_i] = 10e^{-1}(1 - e^{-1}).$$