Exams Statistics and Probability: Probability Part

Master Statistics and Data Science

Instructor: Roula Tsonaka

1. The joint frequency function for the number of audio equipment sales X_1 and number of electronic equipment sales X_2 per hour for a wholesale retailer are given below:

	$X_2 = 0$	$X_2 = 1$	$X_2 = 2$
$X_1 = 0$	0.1	0	0
$X_1 = 1$	0.1	0.2	0
$X_1 = 2$	0.1	?	0.15

Choose the correct answer in the following questions:

(a) Which is the missing value?

(i) 0

(ii) 0.8

(iii) 0.35

(iv) 0.45

(b) What is the probability that both $X_1 \leq 1$ and $X_2 < 1$?

(i) 0

(ii) 0.1

(iii) 0.2

(iv) 0.4

(c) Are X_1, X_2 correlated? Explain.

(i) Yes, because $P(X_2 = j \mid X_1 = 2) = P(X_2 = j)$ for all j.

(ii) No, because $P(X_2 = j \mid X_1 = 2) \neq P(X_2 = j)$ for all j.

(iii) No, because $E(X_1 \cdot X_2) = E(X_1)E(X_2)$.

(iv) Yes, because $Cov(X_1, X_2) \neq 0$.

(d) Which of the following statements is correct?

(i) If X_1, X_2 are independent, then they are correlated.

(ii) If X_1, X_2 are not correlated, then they are independent.

(iii) If X_1, X_2 are independent, then they are not correlated.

(e) What is the value of $E(X_1^2 \cdot X_2)$?

(i) 1.5.

(ii) 2.8.

(iii) 0.21.

(f) What is the expected value of the random variable X_1 given that the random variable X_2 takes the value 0?

(i) 0.3.

(ii) 1.

(iii) 0.4.

<u>*Solution:*</u>

(a) We know that: $\sum_{i=0}^{2}\sum_{j=0}^{2} P(X_1 = i, X_2 = j) = 1$. Thus. $? = 0.35$. The correct answer is (iii).

(b) $P(X_1 \leq 1, X_2 < 1) = P(X_1 = 0, X_2 = 0) + P(X_1 = 1, X_2 = 0) = 0.1 + 0.1 = 0.2$ The correct answer is (iii).

(c) We need to compute $Cov(X_1, X_2) = E(X_1 \cdot X_2) - E(X_1)E(X_2)$.
$E(X_1 \cdot X_2) = \sum_{i=0}^{2}\sum_{j=0}^{2} i \cdot jP(X_1 = i, X_2 = j) = 1 \cdot 1 \cdot P(X_1 = 1, X_2 = 1) + 2 \cdot 1 \cdot P(X_1 = 2, X_2 = 1) + 2 \cdot 2 \cdot P(X_1 = 2, X_2 = 2) = 1 \cdot 0.2 + 2 \cdot 0.35 + 4 \cdot 0.15 = 1.5$
Then $E(X_1) = \sum_{i=0}^{2} i \cdot P(X_1 = i) = 1 \cdot 0.3 + 2 \cdot 0.6 = 1.5$. And $E(X_2) = \sum_{j=0}^{2} j \cdot P(X_2 = i) = 0 \cdot 0.3 + 1 \cdot 0.55 + 2 \cdot 0.15 = 0.85$. Thus $Cov(X_1, X_2) = 1.5 - 1.5 \cdot 0.85 = 0.225$. The correct answer is (iv).

(d) The correct answer is (iii).

(e) $E(X_1^2 \cdot X_2) = \sum_{i=0}^{2}\sum_{j=0}^{2} i^2 \cdot j \cdot P(X_1 = i, X_2 = j) = 1 \cdot 1 \cdot P(X_1 = 1, X_2 = 1) + 2^2 \cdot 1 \cdot P(X_1 = 2, X_2 = 1) + 2^2 \cdot 2 \cdot P(X_1 = 2, X_2 = 2) = 1 \cdot 0.2 + 4 \cdot 0.35 + 8 \cdot 0.15 = 2.8$. The correct answer is (ii).

(f) We need first $P(X_1 = i \mid X_2 = 0)$, $i = 0, 1, 2$. The marginal frequency function for X_2 has been computed in (b). We have: $P(X_1 = 0 \mid X_2 = 0) = 0.1/0.3$, $P(X_1 = 1 \mid X_2 = 0) = 0.1/0.3$ and $P(X_1 = 2 \mid X_2 = 0) = 0.1/0.3$. Then $E(X_1 \mid X_2 = 0) = 1 \cdot 0.1/0.3 + 2 \cdot 0.1/0.3 = 1$. The correct answer is (ii).

2. The general practitioners' office at LUMC receives on average 4 phone calls per minute on a Friday night. The manager of the unit wishes to know the following probabilities:

   (a) What is the probability that 5 or 6 or 7 phone calls are received in a minute?

   (b) What is the probability that no phone calls are received between 10pm and 10.10pm?

   (c) What is the expected number of phone calls between 10pm and 10.10pm? And the standard deviation?

   (d) What is the probability that we have to wait more than 20 seconds for the next call?

   <u>*Solution:*</u>

   - Let X the number of calls per minute on a Friday night. We know that $X \sim Poisson(\lambda = 4)$. We want $P(5 \leq X \leq 7) = P(X = 5) + P(X = 6) + P(X = 7) = \frac{4^5}{5!} \times \exp^{-4} + \frac{4^6}{6!} \times \exp^{-4} + \frac{4^7}{7!} \times \exp^{-4}$
   - Let Y the number of calls in 10 minutes. Then $Y \sim Poisson(\lambda' = 10 \times 4 = 40)$. We want $P(Y = 0) = \exp^{-\lambda'} = \exp^{-40} = 4.248354e - 18$. The probability is very small.
   - $E(Y) = 40$ and the standard deviations is $\sqrt{Var(Y)} = \sqrt{40}$.
   - Let Z the waiting time between calls. Then $Z \sim Exponential(\lambda = 4)$ or $Z \sim Exponential(\beta = 1/4)$. Then $P(Z > 20/60) = 1 - P(Z \leq 1/3) = 1 - (1 - \exp^{-4\times 1/3}) = \exp^{-4\times 1/3} = 0.264$.

3. Suppose that a fair coin is tossed 900 times.

   (a) Compute approximately the probability of obtaining more than 495 heads.

   (b) Let Y be the number of heads in 900 tosses of the coin. How big needs m to be such that $P(440 \leq Y \leq m) \approx 0.5$?

   <u>*Solution:*</u>

- We shall approximate the probability of obtaining more than 495 heads. For $i = 1, \ldots, 900$, let $X_i = 1$ if a head is obtained on the ith toss and let $X_i = 0$ otherwise. Then $E(X_i) = 1/2$ and $Var(X_i) = 1/4$. Therefore, the values $X_1, \ldots, X_{900}$ form a random sample of size $n = 900$ from a distribution with mean $1/2$ and variance $1/4$. It follows from the central limit theorem that the distribution of the total number of heads $H = \sum_{i=1}^{900} X_i$ will be approximately the normal distribution for which the mean is $(900) \times (1/2) = 450$, the variance is $(900) \times (1/4) = 225$, and the standard deviation is $\sqrt{(225)} = 15$. Therefore, the variable $Z = (H450)/15$ will have approximately the standard normal distribution. Thus,

$$P(H > 495) = P(\frac{H - 450}{15} > \frac{495 - 450}{15}) = P(Z > 3) = 1 - \Phi(3) = 0.0013.$$

- We want to find m such that $P(440 \leq H \leq m) = 0.5$. Since $\mu = E(X_i) = 0.5$ and $\sigma^2 = Var(X_i) = 0.25$ by central limit theorem, $Z = \frac{H-450}{15}$ is approximately standard normal. So,

$$\begin{aligned} P(440 \leq H \leq m) &= P(\frac{440 - 450}{15} \leq \frac{H - 450}{15} \leq \frac{m - 450}{15}) \\ &= P(\frac{-10}{15} \leq Z \leq \frac{m - 450}{15}) = P(-0.6666667 \leq Z \leq \frac{m - 450}{15}) \\ &= \Phi(\frac{m - 450}{15}) - 0.2524925. \end{aligned}$$

Thus $\Phi(\frac{m-450}{15}) = 0.5 + 0.2524925$ and $\Phi(\frac{m-450}{15}) = 0.7524925$. Thus, from the table we find that $\Phi^{-1}(0.7524925) = 0.6823542$ and therefore $m = 0.6823542 \times 15 + 450 = 460.2353$.

4. The instructor of the Probability course prepares two sets of exams: Exam A and Exam B for the final examination in September. The probability that a student will get Exam A is 0.80. The probability that the first question on the exam is difficult is 0.90 for exam B and 0.15 for exam A. Compute the following:

   (a) What is the probability that the first question on your exam is marked as difficult?

   (b) What is the probability that your exam is B given that the first question on the exam is marked as difficult?

Solution:

(a) Let C be the event that the first question on the exam is marked as difficult. Let B_1 be the event that the exam is A and B_2 be the event that the exam is B. Applying the formula $P(C) = \sum_{i=1}^{2} P(C \mid B_i)P(B_i)$ gives $P(C) = 0.15 \times 0.8 + 0.9 \times 0.2 = 0.3$.

(b) The probability that the exam is B given that the first question on the exam is marked as difficult is equal to

$$\begin{aligned} P(B_2 \mid C) &= \frac{P(C, B_2)}{P(C)} = \frac{P(C \mid B_2)P(B_2)}{P(C)} \\ &= \frac{0.9 \times 0.2}{0.3} = 0.6. \end{aligned}$$

Motivate your answers!

Good luck!