NATIONAL UNIVERSITY OF SINGAPORE
DEPARTMENT OF STATISTICS AND DATA SCIENCE

ST2334 PROBABILITY AND STATISTICS

FINAL EXAM SAMPLE PAPER 2

(SEMESTER I, AY 2023/2024)

TIME ALLOWED: 120 MINUTES

Suggested Solutions

INSTRUCTIONS TO STUDENTS

1. Please write your student number only. **Do not write your name.**
2. This assessment contains 30 questions and comprises 11 printed pages.
3. The total marks is 60; marks are equal distributed for all questions.
4. Please answer ALL questions.
5. Calculators may be used.
6. This is an **OPEN BOOK** assessment. Only **HARD COPIES** of materials are allowed.

1. **TRUE/FALSE**
   Let X be a discrete random variable; then $E(X)$ always exists.

   - TRUE
   - FALSE

   SOLUTION
   False

2. **TRUE/FALSE**
   Let $\{X_1, X_2, \ldots, X_{1000}\}$ be a random sample from a population with expectation μ. Then both the sample mean $\bar{X}$ and X_1 are unbiased estimators for μ.

   - TRUE
   - FALSE

   SOLUTION
   TRUE

3. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
   Which of the following is a random sample:

   (a) In order to estimate the probability of getting heads for a biased coin, flip the coin 100 times and collect the results for all flips.

   (b) In order to study the scores of undergraduate students of Singapore, 1000 registered undergraduates in NUS were randomly sampled.

   (c) In order to study the average studying hours of students in NUS every week. A random survey with size 1000 was conducted at all libraries of NUS.

   (d) In order to study the average life time of a brand of bulbs, all bulbs' life times in LT32 over a whole semester were recorded.

   SOLUTION
   (a)

4. **FILL IN THE BLANK**
   Let $X_1, X_2, \ldots, X_{10}$ be independent and identically distributed random variables having the exponential distribution $\text{Exp}(1)$. Let $T = \min\{X_1, X_2, \ldots, X_{10}\}$. Find $E(T)$.

   Answer: ________

   (Provide your answer in decimal form and round it to two decimal places if necessary.)

   SOLUTION
   For any $t \geq 0$,

   $$P(T > t) = \prod_{i=1}^{10} P(X_i > t) = \prod_{i=1}^{10} e^{-t} = e^{-10t};$$

   so $T \sim \text{Exp}(10)$ and $E(T) = 1/10 = 0.1$.

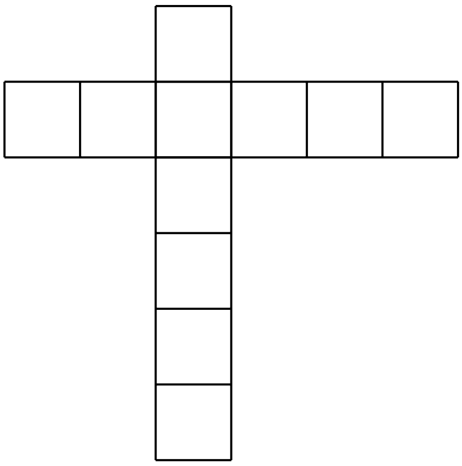
5. **FILL IN THE BLANK**
Suppose that X_1 is Poisson with expectation 1, X_2 is Poisson with expectation 1, and X_3 is Poisson with expectation 2 and assume that the three random variables are independent. Let $Y_1 = X_1 + X_2$ and let $Y_2 = X_2 + X_3$. The conditional probability that $Y_1 = 1$, given that $Y_2 = 2$, is equal to ________.

Answer: ________

(Provide your answer in decimal form and round it to three decimal places if necessary.)

SOLUTION
0.3272 (= 0.0733/0.2240)
$\{(Y_1 = 1) \cap (Y_2 = 2)\}$ is equivalent to (X_1, X_2, X_3) in $\{(0,1,1),(1,0,2)\}$. $P((Y_1 = 1) \cap (Y_2 = 2)) = P(X_1 = 0)P(X_2 = 1)P(X_3 = 1) + P(X_1 = 1)P(X_2 = 0)P(X_3 = 2) = 0.0733$.
$Y_2 = 2$ is equivalent to (X_2, X_3) in $\{(0,2),(1,1),(2,0)\}$. $P(X_2 = 0)P(X_3 = 2) + P(X_2 = 1)P(X_3 = 1) + P(X_2 = 2)P(X_3 = 0) = 0.2240$.

6. **FILL IN THE BLANK**
In how many different ways, we can put 11 different real numbers $x_1, x_2, \ldots, x_{11}$ in the boxes blow (one box each; and each number will be used once and only once), such that for any two neighbors (boxes sharing a common side), the left is always smaller than the right, and the above is always smaller than the blow?



Answer: ________ (Provide your answer in numerical form.)

SOLUTION
105

without loss of generality, assume the numbers are $1, 2, 3, \ldots, 11$. The intersected box must contain the number "4" (called box4 hereafter), since the boxes to its right and bottom are all greater than it and those to its left and above are all smaller than it. 1,2,3 must be put in boxes on the left and above box4, the number of ways to fill in is $\binom{3}{1} = 3$: choose one number to fill in the box on the top of box4, the other two are ordered and fill in the boxes on the left of box4. Similarly, 7 numbers, i.e., 5, 6, …, 11, need to be put in boxes to the right and below box4. The number of ways to fill in is $\binom{7}{3} = 35$: choose 3 numbers out 7 and order them to fill in boxes at the right of the box4, the rest four numbers are ordered and put in boxes below box4. As a consequence, the total number of ways is $3 \times 35 = 105$.

7. **FILL IN THE BLANK**
Let A and B be independent events in the sample space S. If $P(A) = 0.4$, $P(A \cup B) = 0.7$, then $P(B) =$?

Answer: ________

(Provide your answer in decimal form and round it to two decimal places if necessary.)

SOLUTION
0.5.

Using independence and the inclusive exclusive formula,

$$P(A \cup B) = P(A) + P(B) - P(A)P(B),$$

or equivalently

$$0.7 = 0.4 + P(B) - 0.4P(B),$$

the solution of which is $P(B) = 0.5$.

8. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Let A and B be events in the sample space, where $P(A) = 0$. Which of the following must be **TRUE**?

(a) A and B are independent.
(b) A and B are mutually exclusive.
(c) A must be a subset of B.
(d) None of the given options.

SOLUTION
(a).

9. **TRUE/FALSE**
Let $f(x)$ be the probability function of random variable X. If $f(x) = 0$ for $x \in (0, 10)$, then $P(X \leq 0 \text{ or } X \geq 10) = 1$.

- TRUE
- FALSE

SOLUTION
TRUE

No matter whether X is discrete or continuous, $f(x) = 0$ for $x \in (0, 10)$ implies that $P(0 < X < 10) = 0$. Therefore

$$P(X \leq 0 \text{ or } X \geq 10) = 1 - P(0 < X < 10) = 1.$$

10. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Let X be a random variable, such that $V(X)$ exists. Which of the following may **NOT** hold?

(a) $E(3X + 2X^2 + e^X) = 3E(X) + 2E(X^2) + E(e^X)$.
(b) If $E(X) = 0$, then $V(X) = E(X^2)$.

(c) $E(e^X/X) = e^{E(X)}/E(X)$.

(d) $V(X + E(X) + E(e^X)) + (E(X))^2 = E(X^2)$.

SOLUTION

(c).

(a) is correct based on the basic property of expectation.

(b) is correct, since if $E(X) = 0$, $V(X) = E(X^2) - (E(X))^2 = E(X^2)$.

(c) is wrong.

(d) is correct, since bot $E(X)$ and $E(e^X)$ under the variance are constants (non-random), therefore

$$V(X + E(X) + E(e^X)) + (E(X))^2 = V(X) + (E(X))^2 = E(X^2) - (E(X))^2 + (E(X))^2 = E(X^2).$$

11. **MULTIPLE RESPONSE: CHOOSE ALL ANSWERS THAT APPLY**

Suppose A, B, C are events in the sample space S. Which of the following may **NOT** hold?

(a) $(A \cup B) \cup C = A \cup (B \cup C)$

(b) $(A \cap C) \cap B = (A \cap C) \cap (B \cap C)$

(c) $A \cap B = (A \cap B \cap C) \cup (A \cap B \cap C')$

(d) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

SOLUTION

(a), (b), (c), (d).

Based on the computational rules of sets, we can check that all are correct.

12. **FILL IN THE BLANK**

The number of power outrages at a power plant has a Poisson distribution with a mean of 0.06 outrages per day. What is the expected number of power outrages at this power plant in a year? Suppose that a year has 365 days.

Answer: ________

(Provide your answer in decimal form and round it to two decimal places if necessary.)

SOLUTION

21.9.

Let $X =$ number of power outrages at this power plant in a year. Then $X \sim \text{Poisson}(365 \times 0.06) = \text{Poisson}(21.9)$. Therefore $E(X) = 21.9$.

13. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**

Let X be a random variable with cumulative distribution function given by

$$F(x) = \begin{cases} 0 & x < 0 \\ x/4 & 0 \le x < 1 \\ (3+x)/12 & 1 \le x < 3 \\ x/6 & 3 \le x < 6 \\ 1 & x \ge 6 \end{cases}$$

Then $P(1 \le X \le 2) =$?

(a) 1/12

(b) 1/6

(c) 1/4

(d) None of the given options

SOLUTION
(b).

$P(1 \le X \le 2) = P(X \le 2) - P(X < 1) = F(2) - F(1-) = (3+2)/12 - 1/4 = 1/6.$

14. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Assume random variables $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Which of the following statements must be **CORRECT**?

(a) If $\sigma_X = \sigma_Y$, then $Cov(X+Y, X-Y) = 0$.

(b) If $\mu_X = \mu_Y$, then $Cov(X+Y, X-Y) = 0$.

(c) If X and Y are independent, then $Cov(X+Y, X-Y) = 0$.

(d) If X and Y are independent, then $X+Y$ and $X-Y$ are independent.

SOLUTION
(a)

By the property of covariance,

$$\begin{aligned} Cov(X+Y, X-Y) &= E[(X+Y)(X-Y)] - E(X+Y)E(X-Y) \\ &= E(X^2) - E(Y^2) - [E(X)]^2 + [E(Y)]^2 = V(X) - V(Y). \end{aligned}$$

15. **FILL IN THE BLANK**
For a multiple response question with four possible choices (1), (2), (3) and (4), the number of correct answers could be 1 or 2 or 3 or 4. Only when a student answers the question exactly correct, s/he can get the mark. During the exam, suppose one particular student has no time to work on the question, and s/he decides to answer the question by independently flipping a fair coin four times: if a head shows on the ith flip, the choice (i) is included in her/his answer, otherwise, the choice is excluded; if the student gets all tails, s/he leaves the question un-answered. What is the probability that this student gets the mark?

Answer: ________

(Provide your answer in decimal form and round it to four decimal places if necessary.)

SOLUTION
0.0625.

The number of possible answers by this student is $2 \times 2 \times 2 \times 2 = 2^4 = 16$. All these answers are equally likely. But there is only one correct, therefore the probability is $1/16 = 0.0625$.

16. **FILL IN THE BLANK**
The probability function for random variable X is given by

$$f(x) = x^2/10, \quad \text{for } x = -2, -1, 0, 1, 2,$$

and $f(x) = 0$ elsewhere. Compute the variance of X.

Answer: ________

(Provide your answer in decimal form and round it to two decimal places if necessary.)

SOLUTION
3.4

$E(X) = 0$, and thus

$$\begin{aligned} V(X) &= E(X^2) = (-2)^2(-2)^2/10 + (-1)^2(-1)^2/10 + 0 + (2)^2(2)^2/10 + 1^2(1^2)/10 \\ &= 3.4 \end{aligned}$$

17. **FILL IN THE BLANK**
Let (X,Y) be a random vector, whose joint probability function is given by

$$f(x,y) = \begin{cases} \pi e^{-\pi(x^2+y)} & \text{for } -\infty < x < \infty; y \geq 0 \\ 0 & \text{elsewhere} \end{cases}.$$

Compute $Cov(X,Y)$.

Answer: ________

(Provide your answer in decimal form and round it to two decimal places if necessary.)

SOLUTION
0

By identifying that X and Y are independent, we conclude $Cov(X,Y) = 0$.

18. **FILL IN THE BLANK**
Let (X,Y) be a discrete random vector, whose joint probability function is given by:

Table of the joint probability function for (X,Y)

y	x					
	0	1	2	3	4	5
0	0	0.01	0.02	0.05	0.06	0.08
1	0.01	0.03	0.04	0.05	0.05	0.07
2	0.02	0.03	0.05	0.06	0.06	0.07
3	0.02	0.04	0.03	0.04	0.06	0.05

Compute $P(X+Y > 3)$.

Answer: ________

(Provide your answer in decimal form and round it to two decimal places if necessary.)

SOLUTION
0.77.

19. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Let X be the random variable following a normal distribution, where $E(X) = 10$, $E(X^2) = 125$. Suppose $P(-2.5 \leq X < 22.5) = 2\Phi(c) - 1$, where $\Phi(\cdot)$ denotes the cumulative distribution function for the standard normal distribution. What is the value for c?

(a) 2.5
(b) 3
(c) 3.5
(d) Unable to tell

SOLUTION
(a).

$X \sim N(10, 5^2)$. We have

$$P(-2.5 \le X < 22.5) = P\left(\frac{-2.5-10}{5} \le \frac{X-10}{5} \le \frac{22.5-10}{5}\right)$$
$$= P(-2.5 < Z < 2.5) = \Phi(2.5) - P(Z < -2.5) = 2\Phi(2.5) - 1.$$

20. **TRUE/FALSE**
Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with $E(X_1) = 10$ and $\text{var}(X_1) = 1$; denote by $\overline{X}$ the sample mean of these n random variables. Then when n is sufficiently large (approaching infinity say), we can always have

$$P\left(\left|\overline{X} - 10\right| > 1/\sqrt{n}\right) < 0.0000000000000001.$$

- TRUE
- FALSE

SOLUTION
FALSE

Based on Central Limit Theorem, the term on the left hand side is

$$P\left(\left|\overline{X} - 10\right| > 1/\sqrt{n}\right) = P\left(\frac{\left|\overline{X} - 10\right|}{1/\sqrt{n}} > 1\right) \to P(Z > 1) = 1 - \Phi(1),$$

which is much greater than the small quantity on the right hand side.

21. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Roll a fair die until the number 6 appears 6 times. What is the expected number of rolls needed?

(a) 36
(b) 24
(c) 6
(d) None of the given options

SOLUTION
(a)

Consider 6 appears as a success in rolling die; we have $p = 1/6$. Take $k = 6$; let $X =$ number of rolls needed to get k successes. X follows a negative binomial distribution $E(X) = k/p = 6/(1/6) = 36$.

22. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**

Let $A = \{1, 2, 4, 6, 8\}$, $B = \{3, 4, 6, 8, 10\}$, and $C = \{2, 4, 8, 9\}$ be events in the sample space $S = \{1, 2, 3, 4, 6, 8, 9, 10\}$, which of the following is **WRONG**?

(a) $A \cap B \cap C = \{4, 8\}$

(b) $A \cup B \cup C = \{1, 2, 3, 4, 6, 8, 9, 10\}$

(c) $A' \cap C' = \{3, 10\}$

(d) $B' \cap C' = \{1, 5, 7\}$

SOLUTION
(d)

23. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
Let (X,Y) be a random vector. If $\text{var}(X+Y) = 10$, $\text{var}(X) = 4$, $\text{var}(Y) = 8$, then $\text{var}(X-Y) = ?$

(a) 12

(b) 13

(c) 14

(d) not enough information to compute

SOLUTION
(c).

$\text{var}(X+Y) = \text{var}(X) + \text{var}(Y) + 2Cov(X,Y)$, therefore $Cov(X,Y) = -1$. $\text{var}(X-Y) = \text{var}(X) + \text{var}(Y) - 2Cov(X,Y) = 4 + 8 - 2(-1) = 14$.

24. **TRUE/FALSE**
Let $f(x,y)$ be the probability function for the random vector (X,Y); let $f_X(x)$ be the marginal probability function for X. Then, for any real numbers x and y, we must have $f(x,y) \leq f_X(x)$.

- TRUE
- FALSE

SOLUTION
FALSE

It is true if (X,Y) is a discrete random vector, since $f_X(x) = \sum_y f(x,y)$.

But it is not necessarily true if Y is continuous. For example, if X and Y are independent; and $Y \sim U(0, 1/100)$.

25. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
A random sample from $N(\mu, \sigma^2)$ is given below:

$$x_1 = 7, \quad x_2 = 7, \quad x_3 = 3, \quad x_4 = 7.$$

What is the value of b, such that $(2.82, b)$ is a 95% confidence interval for μ?

Note: $t(3, 0.05) = 2.35$; $t(3, 0.025) = 3.18$; $t(4, 0.05) = 2.13$; $t(4, 0.025) = 2.78$.

(a) 7.85

(b) 8.11

(c) 8.76

(d) 9.18

SOLUTION
(d).

$n = 4$ and we can compute $\bar{x} = 6$ and $s^2 = 4$. With normality, 95% CI for μ is given by

$$\bar{x} \pm t(3, 0.025)s/\sqrt{n} = 6 \pm 3.18(2)/2 = (2.82, 9.18).$$

26. **TRUE/FALSE**
    Type I error may occur only when we reject the null hypothesis.

    - TRUE
    - FALSE

    SOLUTION
    TRUE

27. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**
    Suppose that A and B are two events such that $P(A) = 0.5$, $P(A \cap B) = 0.3$. Then $P(A|B) =$?

    (a) 0.4
    (b) 0.3
    (c) Not sufficient information to compute
    (d) None of the given options

    SOLUTION
    (c)

28. **FILL IN THE BLANK**
    The following are the average weekly losses of worker-hours due to accidents in three industrial plants before and after a certain safety program was put into operation.

	Industrial Plant 1	2	3
before	45	124	33
after	36	119	29

Assume that the data are normally distributed.

What is the **absolute value of the computed test statistic** in testing the alternative hypothesis that the safety program reduces the average weekly losses of worker-hours due to accidents at a 5% significance level?

Answer: ________

(Provide your answer in decimal form and round it to two decimal places if necessary.)

SOLUTION
3.928.

Take the difference $D_1 = 9, D_2 = 5, D_3 = 4$; therefore $\bar{D} = 6$, $s_D^2 = 7$, and

$$t = \frac{6-0}{\sqrt{7}/\sqrt{3}} = 3.927922.$$

29. **MULTIPLE RESPONSE: CHOOSE ALL ANSWERS THAT APPLY**
    Suppose X and Y are independent random variables. Which of the following must be **CORRECT**?

(a) X^2 and Y^{1000} are independent.

(b) $X+Y$ and $X-Y$ are independent.

(c) The events $\{X \leq 10\}$ and $\{Y > 10\}$ are independent.

(d) The events $\{X \leq 10\}$ and $\{Y > 10\}$ are mutually exclusive.

SOLUTION

(a), (c).

30. **MULTIPLE CHOICE: CHOOSE ONE ANSWER ONLY**

Assume that the survival time (in years) of a patient who has a certain cancer follows the exponential distribution with average survival time equal to 4 (years). If such a patient has survived for 4 years, what is the probability that this patient can survive for another 4 years?

(a) e^{-1}

(b) $1-e^{-1}$

(c) e^{-2}

(d) $1-e^{-2}$

SOLUTION

(a).

$X \sim \text{Exp}(1/4)$. $P(X > 4+4 | X > 4) = P(X > 4) = e^{-4/4} = e^{-1}$.

END OF PAPER