Seat No. ______ Major/Class ______ School ______ Student ID ______ Name ______

(DO NOT WRITE YOUR ANSWER IN THIS AREA)

Seal line .... Seal line .... Seal line ....

WARNING: MISBEHAVIOR AT EXAM TIME WILL LEAD TO SERIOUS CONSEQUENCE.

SCUT Final Exam

Probability and Statistics Exam paper B (2021–2022-2)

Notice:
1. **Make sure that you have filled the form on the left side of seal line.**
2. **Write your answers on the exam paper.**
3. **This is a close-book exam.**
4. **The exam with full score of 100 points lasts 120 minutes.**

Question No.	I	II	III	IV	V	VI	VII	VIII	IX	X	Sum
Score											

I. (10 points) We roll a die n times. Let A_{ij} be the event that the i th and j th rolls produce the same number. Show that the events $\{A_{ij} : 1 \leq i < j \leq n\}$ are pairwise independent but not independent.

Score

Suppose $i < j$ and $m < n$. If $j < m$, then A_{ij} and A_{mn} are determined by distinct independent rolls, and are therefore independent. For the case $j = m$ we have that $\mathbb{P}\left(A_{ij} \cap A_{jn}\right) = \mathbb{P}($$i$ th, j th, and nth rolls show same number)

$$= \sum_{r=1}^{6} \frac{1}{6} \mathbb{P}(j \text{ th and } n\text{th rolls both show } r \mid i \text{ th shows } r) = \frac{1}{36} = \mathbb{P}\left(A_{ij}\right) \mathbb{P}\left(A_{jn}\right)$$

as required. However, if $i \neq j \neq k$,

$$\mathbb{P}\left(A_{ij} \cap A_{jk} \cap A_{ik}\right) = \frac{1}{36} \neq \frac{1}{216} = \mathbb{P}\left(A_{ij}\right) \mathbb{P}\left(A_{jk}\right) \mathbb{P}\left(A_{ik}\right)$$

II. (10 points) Two fair dice are rolled. Show that the event that their sum is 7 is independent of the score shown by the first die.

Score

$$\mathbb{P}(1 \text{ st shows } r \text{ and sum is } 7) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6} = \mathbb{P}(1 \text{ st shows } r) \mathbb{P}(\text{ sum is } 7)$$

III. (10 points) Individuals **A** and **B** begin to play a sequence of chess games. Let $S = \{\text{A wins a game}\}$, and suppose that outcomes of successive games are independent with $P(S) = p$ and $P(F) = 1 - p$ (they never draw). They will play until one of them wins ten games. Let $X =$ the numher of games played (with possible values $10, 11, \ldots, 19$).

Score

a. For $x = 10, 11, \ldots, 19$, obtain an expression for $p(x) = P(X = x)$

b. If a draw is possible, with $p = P(S), q = P(F)$, $1 - p - q = P(\text{draw})$, what is $P(20 \leq X)$?

a. $P(X = x) = P(\text{ A wins in } x \text{ games }) + P(\text{ B wins in } x \text{ games })$

win the last

$= P\left(9S'\text{s in } 1^{\text{st}}\ x-1 \cap S \text{ on the } x^{\text{th}}\right) + P\left(9F'\text{ s in } 1^{\text{st}}\ x-1 \cap F \text{ on the } x^{\text{th}}\right)$

$$= \binom{x-1}{9} p^9 (1-p)^{(x-1)-9} \cdot p + \binom{x-1}{9} (1-p)^9 p^{(x-1)-9} \cdot (1-p)$$

$$= \binom{x-1}{9} \left[p^{10}(1-p)^{x-10} + (1-p)^{10} p^{x-10}\right].$$

b. Possible values of X are now all positive integers $\geq 10 : 10, 11, 12, \ldots.$

Similar to case a),

$P(X = x) = P(\text{ A wins in } x \text{ games }) + P(\text{ B wins in } x \text{ games })$

$= P\left(9S's \text{ in } 1^{\text{st}}\ x-1 \cap S \text{ on the } x^{\text{th}}\right) + P\left(9F'\text{ s in } 1^{\text{st}}\ x-1 \cap F \text{ on the } x^{\text{th}}\right)$

$$= \binom{x-1}{9} p^9 (1-p)^{(x-1)-9} \cdot p + \binom{x-1}{9} q^9 (1-q)^{(x-1)-9} \cdot q$$

$$= \binom{x-1}{9} \left[p^{10}(1-p)^{x-10} + q^{10}(1-q)^{x-10}\right]. \text{ Finally,}$$

$$P(X \geq 20) = 1 - P(X < 20) = \sum_{x=10}^{19} \binom{x-1}{9} \left[p^{10}(1-p)^{x-10} + q^{10}(1-q)^{x-10}\right].$$

IV. (10 points) A 12-in. bar that is clamped at both ends is to be subjected to an increasing amount of stress until it snaps. Let $Y =$ the distance from the left end at which the break occurs. Suppose Y has pdf

Score

$$f(y) = \begin{cases} \left(\frac{1}{24}\right) y \left(1 - \frac{y}{12}\right) & 0 \leq y \leq 12 \\ 0 & \text{otherwise} \end{cases}$$

Compute the following:

a. The cdf of Y.

b. The expected length of the shorter segment $\min(Y, 12 - Y)$ when the break occurs.

a. For $0 \le y \le 12$, $F(y) = \frac{1}{24}\int_0^y \left(u - \frac{u^2}{12}\right) du = \frac{1}{24}\left(\frac{u^2}{2} - \frac{u^3}{36}\right)\Big]_0^y = \frac{y^2}{48} - \frac{y^3}{864}$.

b. The shorter segment has length equal to $\min(Y, 12 - Y)$, and

$$E[\min(Y, 12 - Y)] = \int_0^{12} \min(y, 12 - y) \cdot f(y)dy = \int_0^{6} \min(y, 12 - y) \cdot f(y)dy$$

$$+ \int_6^{12} \min(y, 12 - y) \cdot f(y)dy = \int_0^{6} y \cdot f(y)dy + \int_6^{12} (12 - y) \cdot f(y)dy = \frac{90}{24} = 3.75 \text{ inches.}$$

V. (10 points) Let X and Y be two continuous rv's with joint pdf

Score

$$f(x, y) = \begin{cases} K\left(x + y^2\right) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

1. What is the value of K ?
2. What is the conditional probability that X is at most 0.5 given that $Y = 0.5$?
3. Compute the covariance between X and Y.

1. $1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = \int_0^1\int_0^1 K\left(x + y^2\right)dxdy = K\int_0^1\int_0^1 xdxdy + K\int_0^1\int_0^1 y^2dxdy = \frac{K}{2} + \frac{K}{3} = \frac{5K}{6} \Rightarrow K = \frac{6}{5}$
2. The marginal pdf of Y at 0.5 is $f_Y(0.5) = \int_{-\infty}^{\infty} f(x, 0.5)dx = \frac{6}{5}\int_0^1 (x + \frac{1}{4})dx = \frac{9}{10}$ and the conditional pdf of X given that $Y = 0.5$ is
$$f_{X|Y}(x|0.5) = \frac{f(x, 0.5)}{f_Y(0.5)} = \begin{cases} \frac{4}{3}\left(x + \frac{1}{4}\right), & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$
And thus $P(X \le 0.5|Y = 0.5) = \int_0^{0.5} \frac{4}{3}\left(x + \frac{1}{4}\right) dx = \frac{1}{3}$.
3. $$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)dxdy = \frac{7}{20};$$
$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x, y)dxdy = \frac{3}{5};$$
$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x, y)dxdy = \frac{3}{5};$$
$$Cov(X, Y) = E(XY) - E(X)E(Y) = \frac{7}{20} - \frac{9}{25} = -0.01.$$

VI. (10 points) Suppose that X and Y are two independent rv's, both of which has uniform distribution in $(0, 2)$.

Score

1. Determine the joint pdf of X and Y.
2. Compute the probability $P(X + Y \leq 1)$
3. Compute the probability $P(X \leq Y)$
4. Compute $V(X - Y)$.

1. Since X and Y are independent, their joint pdf is

$$f(x,y) = \begin{cases} f_X(x)f_Y(y) = \frac{1}{4}, & 0 \leq x \leq 2,\ 0 \leq y \leq 2; \\ 0, & \text{otherwise.} \end{cases}$$

2. $P(X + Y \leq 1) = \int_0^1 \left[\int_0^{1-x} \frac{1}{4} dy\right] dx = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$.
3. $P(X \leq Y) = \int_0^2 \left[\int_x^2 \frac{1}{4} dy\right] dx = \frac{1}{4} \times 2 = \frac{1}{2}$.
4. $V(X - Y) = V(X) + V(Y) = 2V(X) = 2 \times \frac{2^2}{12} = \frac{2}{3}$.

VII. (10 points) Suppose the expected tensile strength of type-A steel is 100ksi and the standard deviation of tensile strength is 8ksi. For type-B steel, suppose the expected tensile strength is 95ksi and the standard deviation of tensile strength is 7ksi, respectively. Let $\bar{X}$ = the sample average tensile strength of a random sample of 40 type-A specimens, and let $\bar{Y}$ = the sample average tensile strength of a random sample of 35 type-B specimen. (Using the Central Limit Theorem to answer the following questions) $(\Phi(1.65) = 0.95, \Phi(2.89) = 0.998, \Phi(1.96) = 0.975)$ (a) What is the approximate distribution of $\bar{X}$? of $\bar{Y}$? (b) Calculate $P(\bar{X} - \bar{Y} \geq 10)$.

Score

a. According to the CLT , $\bar{X}$ has approximately a normal distribution $N\left(100, \frac{8^2}{40}\right)$, i.e. $N(100, 1.6)$, $\bar{Y}$ has approximately a normal distribution $N\left(95, \frac{7^2}{35}\right)$,i.e.$N(95, 1.4)$.

b. According to the CLT, $\bar{X} - \bar{Y}$ has approximately a normal distribution $N(5, 3)$. $P(\bar{X} - \bar{Y} \geq 10) = 1 - P(\bar{X} - \bar{Y} < 10) \approx 1 - \Phi\left(\frac{10-5}{\sqrt{3}}\right) = 1 - \Phi\left(\frac{5}{\sqrt{3}}\right) = 1 - \Phi(2.886) = 1 - 0.998 = 0.002$.

VIII. (10 points) Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from the pdf

Score

$$f(x;\theta) = \begin{cases} \frac{x}{\theta}e^{-\frac{x^2}{2\theta}}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

(a) Use the method of moments to find an estimator for θ. (b) Find the maximum likelihood estimator for θ.

(a)

$$\begin{aligned} E(x) &= \int_{-\infty}^{+\infty} x f(x,\theta)dx = \int_0^{+\infty} \frac{x^2}{\theta}e^{-\frac{x^2}{2\theta}}dx \\ &= \int_0^{+\infty} xe^{-\frac{x^2}{2\theta}}d\frac{x^2}{2\theta} = x\left(-e^{-\frac{x^2}{2\theta}}\right)\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\frac{x^2}{2\theta}}dx \\ &= 0 + \frac{1}{2}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\theta}}dx = \sqrt{\frac{\theta \cdot \pi}{2}} \\ \therefore \theta &= \frac{2E^2(x)}{\pi} \end{aligned}$$

The moment estimator is $\hat{\theta}_M = \frac{2}{\pi}\bar{X}^2$ (b) The likelihood function is

$$f(x_1, \cdots, x_n, \theta) = \prod_{i=1}^n \frac{x_i}{\theta} \cdot e^{-\sum_{i=1}^n \frac{x_i^2}{2\theta}} \quad (x_i > 0, i = 1, \cdots, n)$$

The ln(likelihood) is

$$\begin{aligned} \ln f(x_1, \cdots, x_n; \theta) &= \sum_{i=1}^n \ln\left(\frac{x_i}{\theta}\right) - \sum_{i=1}^n \frac{x_i^2}{2\theta} \\ &= \sum_{i=1}^n \ln x_i - n\ln\theta - \sum_{i=1}^n \frac{x_i^2}{2\theta}. \\ \frac{d\ln f(x_1, \cdots, x_{n,}\theta)}{d\theta} &= \frac{-n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n \frac{x_i^2}{2} = 0 \\ \Rightarrow \theta &= \frac{1}{2n}\sum_{i=1}^n x_i^2 \end{aligned}$$

So the maximum likelihood estimator of θ is

$$\hat{\theta}_l = \frac{1}{2n}\sum_{i=1}^n X_i^2.$$

IX. (10 points) It is reported that for a sample of 49 kitchens with gas cooking appliances monitored during a one-week period, the sample mean CO_2 level (ppm) was 654. Suppose that the population of CO_2 level of all homes is normal.

Score

(a) Calculate a 95% confidence interval for true average CO_2 level with the sample standard deviation $s = 168$. ($t_{0.05,48} = 1.68, t_{0.025,48} = 2.0$.)

(b) Suppose that $\sigma = 175$. What sample size would be necessary to obtain an interval width of at most 50ppm for a confidence level of 95% ? ($z_{0.05} = 1.65, z_{0.025} = 1.96$.)

(a) With $n = 49, \overline{x} = 654$ and $t_{\alpha/2,n-1} = t_{0.025,48} = 2.0$, the 95% confidence interval for μ is

$$\overline{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = 654 \pm 2.0\frac{168}{7} = 654 \pm 48 = (606, 702).$$

(b) With $\sigma = 175$ and $\mu \in \overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, the width of CI is $w = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. So

$$n = \left(\frac{2z_{\alpha/2}\sigma}{w}\right)^2 = \left(\frac{2(1.96)(175)}{50}\right)^2 = (13.72)^2 = 188.24,$$

which rounds up to 189.

X. (10 points) The desired percentage of SiO_2 in a certain type of aluminous cement is 5.5. In a test 16 independently obtained samples are analyzed. Suppose that the percentage of SiO_2 is normally distributed with $\sigma = 0.3$ and that $\bar{x} = 5.25$.

Score

(a) Does this indicate conclusively that the true average percentage less than 5.5? Consider a significance level of $\alpha = 0.01$. ($z_{0.01} = 2.33, z_{0.005} = 2.58$.)

(b) If the true average percentage is $\mu = 5.3$ and a level $\alpha = 0.01$ test based on $n = 16$ is used, what is the probability of rejecting H_0? ($\Phi(0.34) = 0.63, \Phi(0.64) = 0.74$.)

(c) What value of n is required to satisfy $\alpha = 0.01$ and $\beta(5.3) = 0.01$? Note that $\beta(5.3)$ is the probability of making a type II error when $\mu = 5.3$.

The hypotheses are $H_0 : \mu = 5.5$ vs $H_a : \mu < 5.5$. The sample mean is $\bar{x} = 5.25$.

(a) $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{5.25-5.5}{0.3/\sqrt{16}} = -3.33 \leq -z_\alpha = -z_{0.01} = -2.33$. Reject H_0.

(b) The probability of making a type II error when $\mu = 5.3$ is

$$\beta(5.3) = 1 - \Phi\left(-z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) = 1 - \Phi\left(-2.33 + \frac{5.5-5.3}{0.3/\sqrt{16}}\right) = 1 - \Phi(0.34).$$

So the probability of rejecting H_0 is $1 - \beta(5.3) = \Phi(0.34) = 0.63$.

(c) Since $z_{0.01} = 2.33$, the requirement that the level 0.01 test also have $\beta(5.3) = 0.01$ necessitates

$$1-\Phi\left(-2.33 + \frac{5.5-5.3}{0.3/\sqrt{n}}\right) = 0.01,\ -2.33+\frac{5.5-5.3}{0.3/\sqrt{n}} = 2.33,\ n = \left[\frac{0.3(2.33+2.33)}{5.5-5.3}\right]^2 = 48.9.$$

So $n = 49$ should be used.