Name:

UTeid:

M378K Introduction to Mathematical Statistics
Fall 2024
University of Texas at Austin
In-Term Exam II
Instructor: Milica Čudina

Notes: This is a closed book and closed notes exam. The maximal score on the real exam will be 100 points.
There are many ways in which any single problem can be solved. The solutions herein are just one possible way to tackle the given problems.
Time: 50 minutes

All written work handed in by the student is considered to be
their own work, prepared without unauthorized assistance.

The University Code of Conduct

"The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community. As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity."

"I agree that I have complied with the UT Honor Code during my completion of this exam."

Signature:

2.1. **Formulas.** If Y has the binomial distribution with parameters n and p, then $p_Y(k) = \mathbb{P}[Y = k] = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, \ldots, n$, $\mathbb{E}[Y] = np$, $\mathrm{Var}[Y] = np(1-p)$. The binomial coefficients are defined as follows for integers $0 \le k \le n$, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. The moment generating function of Y is given by $m_Y(t) = (pe^t + q)^n$.

If Y has a geometric distribution with parameter p, then $p_Y(k) = p(1-p)^k$ for $k = 0, 1, \ldots$, $\mathbb{E}[Y] = \frac{1-p}{p}$, $\mathrm{Var}[Y] = \frac{1-p}{p^2}$. Its mgf is $m_Y(t) = \frac{p}{1-qe^t}$ for t such that $qe^t < 1$.

If Y has a Poisson distribution with parameter λ, then $p_Y(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k = 0, 1, \ldots$, $\mathbb{E}[Y] = \mathrm{Var}[Y] = \lambda$. Its mgf is $m_Y(t) = e^{\lambda(e^t - 1)}$.

If Y has a uniform distribution on $[l, r]$, its density is

$$f_Y(y) = \frac{1}{r-l} \mathbf{1}_{(l,r)}(y),$$

its mean is $\frac{l+r}{2}$, and its variance is $\frac{(r-l)^2}{12}$. Let $U \sim U(0,1)$. The mgf of U is $m_U(t) = \frac{1}{t}(e^t - 1)$.

If Y has the standard normal distribution, then its mean is zero, its variance is one, and its density equals

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad x \in \mathbb{R}.$$

Its mgf is $m_Y(t) = e^{\frac{t^2}{2}}$.

If Y has the exponential distribution with parameter τ, then its cumulative distribution function is $F_Y(y) = 1 - e^{-\frac{y}{\tau}}$ for $y \ge 0$, its probability density function is $f_Y(y) = \frac{1}{\tau} e^{-y/\tau}$ for $y \ge 0$. Also, $\mathbb{E}[Y] = SD[Y] = \tau$. Its mgf is $m_Y(t) = \frac{1}{1-\tau t}$.

The mgf of $Y \sim \Gamma(k, \tau)$ is

$$m_Y(t) = \frac{1}{(1-\tau t)^k} \text{ for } t < 1/\tau.$$

Its expectation is $k\tau$ and its variance is $k\tau^2$. The χ^2−distribution with n degrees of freedom is the special case $\Gamma\left(\frac{n}{2}, 2\right)$

2.2. DEFINITIONS.

Problem 2.1. (10 points) Write down the definition of the **moment generating function** of a random variable Y.

Solution:

$$m_Y(t) = \mathbb{E}[e^{tY}]$$

for all $t \in \mathbb{R}$ such that the above expectation exists. In fact, we say that the moment-generating function **exists** if there exists a positive number b such that $m_Y(t)$ is finite for all t such that $|t| \leq b$.

Problem 2.2. (10 points) Write down the definition of the **random sample** of size n from a distribution D.

Solution: A *random sample* of size n from a distribution D is a random vector

$$(Y_1, Y_2, \ldots, Y_n)$$

such that:

1. $Y_1, Y_2, \ldots, Y_n$ are **independent**, and
2. Y_i has the distribution D for every $i = 1, 2, \ldots, n$.

2.3. TRUE/FALSE QUESTIONS.

Problem 2.3. (5 points) Let $Y_1, Y_2, \dots, Y_n$ be a random sample. Then,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(0,\infty)}(Y_i)$$

is a well-defined statistic. *True or false?*

Solution: TRUE

Problem 2.4. (5 points) Let $Y_1, Y_2, \dots, Y_n$ be a random sample from $N(\mu, \sigma)$ with μ **known** and σ **unknown**. Then,

$$\frac{1}{n}\sum_{i=1}^{n} (Y_i - \mu)^2$$

is a well-defined estimator for σ^2. *True or false?*

Solution: TRUE

2.4. Free-response problems.

Please, explain carefully all your statements and assumptions. Numerical results or single-word answers without an explanation (even if they're correct) are worth 0 points.

Problem 2.5. (15 points)Let (Y_1, Y_2) be a random vector with the joint pdf

$$f_{Y_1,Y_2}(y_1, y_2) = \tfrac{1}{4}\mathbf{1}_{\{-1 \le y_1 \le 1\}}\mathbf{1}_{\{-1 \le y_2 \le 1\}}.$$

Find $\mathbb{P}\Big[\,|Y_1| + |Y_2| \le 1/2\Big]$.

Solution: The pair (Y_1, Y_2) is uniformly distributed over the square $[-1,1] \times [-1,1]$, while the region $\{(y_1, y_2) \in [-1,1]^2 : y_1^2 \le y_2^2\}$ corresponds to the square with vertices $(\frac{1}{2}, 0)$, $(0, \frac{1}{2})$, $(-\frac{1}{2}, 0)$ and $(0, -\frac{1}{2})$. The side length of this square is $1/\sqrt{2}$, so its total area is $\frac{1}{2}$. The total area of the square $[-1,1]$ is 4, and, since we are dealing with a geometric-probability problem, the answer is $\frac{1}{2}/4 = \frac{1}{8}$.

Problem 2.6. (10 points) Let $Y \sim U(l, r)$ What is the moment generating function of Y?

Solution: See **Example 6.1.5** from the lecture notes.

Problem 2.7. (20 points) In Croatia, if you go to the chocolate-factory store, you can buy broken off chunks of rice-puff chocolate. From past experience, we know that the weight of the individual chunks has mean of 40 grams and standard deviation of 5 grams. Assume that the weights of individual pieces of chocolate are independent.

You buy 400 chocolate chunks. What is the probability that the total weight exceeds 16128 grams?

Solution: Let $n = 400$ denote the total number of chocolate chunks. Let $Y_i, i = 1, \dots, n$ be the random variables which stand for the weights of individual chunks. Then, their total weight can be expressed as

$$S = Y_1 + \cdots + Y_n$$

We can use the Central Limit Theorem (CLT) here since $n = 400$. We have that S is approximately normal with mean $40(400) = 16000$ and standard deviation $5\sqrt{400} = 100$.

The probability we are asked to calculate is

$$\mathbb{P}[S > 16128] = \mathbb{P}\left[\frac{S - 16000}{100} > \frac{16128 - 16000}{100}\right] \approx \mathbb{P}\left[Z > 1.28\right]$$

where $Z \sim N(0,1)$. We get

$$\mathbb{P}[S > 16128] \approx 1 - \Phi(1.28).$$

If we consult the standard normal tables, we get our answer as $1 - 0.8997 = 0.1003$.

2.5. MULTIPLE CHOICE QUESTIONS.

Problem 2.8. (5 points) Let Y be a uniform random variable on $[0,1]$, and let $W = Y^2$. The pdf of W is

(a) $\frac{1}{2\sqrt{|w|}}\mathbf{1}_{\{-1<w<1\}}$

(b) $\frac{1}{\sqrt{w}}\mathbf{1}_{\{0<w<1\}}$

(c) $\frac{1}{2\sqrt{w}}\mathbf{1}_{\{0<w<1\}}$

(d) $2w\mathbf{1}_{\{0<w<1\}}$

(e) none of the above

Solution: The correct answer is **(c)**.
We use the h-method. $g(y) = y^2$ and $h(w) = \sqrt{w}$. Therefore

$$f_W(w) = f_Y(h(w))h'(w) = \tfrac{1}{2\sqrt{w}}\mathbf{1}_{\{0<w\leq 1\}}.$$

Problem 2.9. (5 points) Let $Y_1, Y_2, \dots, Y_n$ be independent, identically distributed normal random variables with mean μ and standard deviation σ. What is the distribution of the random variable Y defined as

$$Y = \left(\frac{Y_1 - \mu}{\sigma}\right)^2 + \left(\frac{Y_2 - \mu}{\sigma}\right)^2 + \dots + \left(\frac{Y_n - \mu}{\sigma}\right)^2 ?$$

(a) $N(0, \sqrt{n})$
(b) $\chi^2(n)$
(c) $\chi^2(n-1)$
(d) $N(0, n^2)$
(e) **None of the above.**

(*Note:* In our notation $N(\mu, \sigma)$ means normal with mean μ and *standard deviation* σ.)

Solution: The correct answer is **(b)**.

Problem 2.10. (5 points) You are monitoring a cash register at a store. A seemingly endless queue of customers is waiting. The times it takes for Cassie the Cashier to check out a single customer is exponential with mean 5 minutes for each customer. Moreover, the customer service times are independent. Cassie the cashier can go on a break after every batch of 10 customers leave. She just came in from a break. What is the distribution of her waiting time until the next break?

(a) $\Gamma(10, 5)$, i.e., $k = 10$, $\tau = 5$
(b) $\Gamma(50, 1)$, i.e., $k = 50$, $\tau = 1$
(c) $\chi^2(25)$
(d) $E(50)$
(e) $E(1/50)$

Solution: The correct answer is **(a)**.
We need the distribution of the sum of 10 independent exponential random variables $E(\tau)$ with parameter $\tau = 5$. Each $E(5)$ is a special case of the gamma distribution with parameters $k = 1$ and $\tau = 5$. Gammas are additive in the shape parameter, so the result is $\Gamma(10, 5)$.

Problem 2.11. (5 points) A math graduate student basically survives on espresso and chocolate. Their daily chocolate consumption is normally distributed with mean 16 oz and standard deviation 4 oz. Their daily espresso consumption is normally distributed with mean 20 oz and standard deviation 3 oz. Assume that espresso consumption and chocolate consumption are independent.

What is the probability that chocolate consumption exceeds coffee consumption in a single day?

(a) About 0.2119
(b) About 0.2839
(c) About 0.4207
(d) About 0.4432
(e) **None of the above.**

Solution: The correct answer is **(a)**.
Let Y_1 be the chocolate consumption and let Y_2 be the espresso consumption. We are given that Y_1 and Y_2 are independent. Also,

$$Y_1 \sim N(\mu_1 = 16, \sigma_1 = 4) \quad \text{and} \quad Y_2 \sim N(\mu_2 = 20, \sigma_2 = 3).$$

We need to calculate $\mathbb{P}[Y_1 > Y_2] = \mathbb{P}[Y_1 - Y_2 > 0]$. From the given information, we can conclude that

$$Y_1 - Y_2 \sim N(\mu = -4, \sigma = \sqrt{3^2 + 4^2} = 5).$$

So,

$$\mathbb{P}[Y_1 - Y_2 > 0] = \mathbb{P}\left[\frac{Y_1 - Y_2 - \mu}{\sigma} > \frac{0 - \mu}{\sigma}\right] = \mathbb{P}[Z > 0.8]$$

where $Z \sim N(0,1)$. From the standard normal tables, we get $1 - 0.7881 = 0.2119$.

Problem 2.12. (5 points) Let $Y_1, \ldots, Y_{100}$ be independent random variables with the Bernoulli $B(p)$ distribution, with $p = 0.2$ The best approximation to $\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$ (among the offered answers) is

(a) $N(0,1)$
(b) $N(100, 20)$
(c) $N(0.2, 0.04)$
(d) $N(20, 4)$
(e) $N(20, 20)$

(*Note:* In our notation $N(\mu, \sigma)$ means normal with mean μ and *standard deviation* σ.)

Solution: The correct answer is **(c)**.
The sum $W = Y_1 + \cdots + Y_n$ is binomially distributed with mean $np = 20$ and variance $np(1-p) = 16$, i.e., standard deviation 4. It is well approximated by a normal $N(20, 4)$. Since $\bar{Y} = \frac{1}{n}W$, its best approximation will a normal with mean $\frac{1}{100}20 = 0.2$ and standard deviation $\sigma = \frac{1}{100}4 = 0.04$.