編號:____________________

問題	1: 10分	2: 10分	3: 10分	4: 10分	5: 10分	總分: 100分
得分						
問題	6: 10分	7: 10分	8: 10分	9: 10分	10: 10分	
得分						

1. Let X and Y be random variables with finite means. Find a function $g^*(X)$ such that
$$\min_{g(X)} E\left[(Y-g(X))^2\right] = E\left[(Y-g^*(X))^2\right],$$
where $g(x)$ ranges over all functions.

解答:

$$\begin{aligned} E[Y-g(X)]^2 &= E[(Y-E[Y|X])+(E[Y|X]-g(X))]^2 \\ &= E[Y-E[Y|X]]^2 + E[E[Y|X]-g(X)]^2 \\ &\quad + 2E[(Y-E[Y|X])(E[Y|X]-g(X))]. \end{aligned}$$

The cross term can be shown to be zero by iterating the expectation. Thus

$$\begin{aligned} E[Y-g(X)]^2 &= E[Y-E[Y|X]]^2 + E[E[Y|X]-g(X)]^2 \\ &\ge E[Y-E[Y|X]]^2, \quad \text{for all } g(\cdot). \end{aligned}$$

The choice $g^*(X) = E[Y|X]$ will give equality.

2. Suppose X and Y are independent $N(0,1)$ random variables. Find $P(X^2+Y^2<1)$.

解答: Note that $X^2+Y^2 \sim \chi^2_2$. Thus $P(X^2+Y^2<1) = \int_0^1 \frac{e^{-x/2}}{2}\,\mathrm{d}x = 1-\frac{1}{\sqrt{e}} = 0.3935$.

3. Suppose $X_1, X_2, \ldots$ are jointly continuous and independent, each distributed with marginal pdf $f(x)$, where each X_i represents annual rainfall at a given location.

   (a) Find the distribution of the number of years until the first year's rainfall, X_1, is exceeded for the first time.

   (b) Show that the mean number of years until X_1 is exceeded for the first is infinite.

解答:

(a) Let N denote the number of years until the first year's rainfall, X_1, is exceeded for the first time. Then, for $n \geq 2$,

$$\begin{aligned}
P(N = n-1) &= \int_0^\infty P(X_1 = x, X_2 \leq x, \ldots, X_{n-1} \leq x, X_n > x | X_1 = x)\, f(x)\, \mathrm{d}x \\
&= \int_0^\infty F^{n-2}(x)(1-F(x))f(x)\, \mathrm{d}x \qquad (u = F(x)) \\
&= \int_0^1 u^{n-2}(1-u)\, \mathrm{d}u \\
&= \left(\frac{1}{n-1} - \frac{1}{n}\right)
\end{aligned}$$

or equivalently,

$$\begin{aligned}
P(N \geq n-1) &= \int_0^\infty P(X_1 = x, X_2 \leq x, \ldots, X_{n-1} \leq x | X_1 = x)\, f(x)\, \mathrm{d}x \\
&= \int_0^\infty F^{n-2}(x)f(x)\, \mathrm{d}x \qquad (u = F(x)) \\
&= \int_0^1 u^{n-2}\, \mathrm{d}u \\
&= \frac{1}{n-1}
\end{aligned}$$

(b) $E[N] = \sum_{n=2}^\infty P(N \geq n-1) = \sum_{n=2}^\infty \frac{1}{n-1} = \infty$

4. What is the probability that the larger of two continuous iid random variables will exceed the population median? Generalize this result to sample of size n.

解答: Let m denote the median. Then, for general n we have

$$\begin{aligned}
P(\max(X_1, \ldots, X_n) > m) &= 1 - P(X_i \leq m \text{ for } i = 1, 2, \ldots, n) \\
&= 1 - P(X_1 \leq m)^n = 1 - \left(\frac{1}{2}\right)^n.
\end{aligned}$$

5. Let $U_i, i = 1, 2, \ldots,$ be independent uniform$(0,1)$ random variables, and let X have distribution

$$P(X = x) = \frac{c}{x!}, \quad x = 1, 2, 3, \ldots,$$

where $c = 1/(e-1)$. Find the distribution of

$$Z = \min\{U_1, \ldots, U_X\}.$$

解答:

$$
\begin{aligned}
P(Z > z) &= \sum_{x=1}^{\infty} P(Z > z|x)\, P(X = x) = \sum_{x=1}^{\infty} P(U_1 > z, \ldots, U_x > z|x)\, P(X = x) \\
&= \sum_{x=1}^{\infty} \prod_{i=1}^{x} P(U_i > z)\, P(X = x) \quad \text{(by independence of the } U_i\text{'s)} \\
&= \sum_{x=1}^{\infty} P(U_i > z)^x\, P(X = x) = \sum_{x=1}^{\infty} (1 - z)^x \frac{1}{(e - 1)x!} \\
&= \frac{1}{(e-1)} \sum_{x=1}^{\infty} \frac{(1-z)^x}{x!} = \frac{e^{1-z} - 1}{e - 1}, \quad 0 < z < 1.
\end{aligned}
$$

6. Let $X_1, \ldots, X_n$ be independent random variables with densities

$$
f_{X_i}(x|\sigma) = \begin{cases} e^{i\theta - x} & \text{if } x \geq i\theta, \\ 0 & \text{if } x < i\theta. \end{cases}
$$

Prove that $T = \min_i(X_i/i)$ is a sufficient statistic for θ.

解答: By the Factorization Theorem, $T(X) = \min_i(X_i/i)$ is sufficient because we can write the joint pdf of $X_1, \ldots, X_n$ as

$$
f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} e^{i\theta - x_i} I_{(i\theta, +\infty)}(x_i) = \underbrace{e^{\theta \sum_{i=1}^{n} i} I_{(\theta, +\infty)}(T(x))}_{g(T(x)|\theta)} \cdot \underbrace{e^{-\sum x_i}}_{h(x)}.
$$

Notice, we use the fact that $i > 0$, and the fact that all x_is $> i\theta$ if and only if $\min(X_i/i) > \theta$.

7. Let $X_1, \ldots, X_n$ be a sample from the *inverse Gaussian* pdf,

$$
f(x|\mu, \lambda) = \left(\frac{\lambda}{2\pi x^3} \right)^{1/2} \exp\{-\lambda(x - \mu)^2/(2\mu^2 x)\}, \quad x > 0.
$$

Find the MLEs of μ and λ.

解答: The likelihood is

$$
L(\mu, \lambda|x) = \frac{\lambda^{n/2}}{(2\pi)^n \prod_i x_i} \exp\left\{ -\frac{\lambda}{2} \sum_i \frac{(x_i - \mu)^2}{\mu^2 x_i} \right\}.
$$

For fixed λ, maximizing with respect to μ is equivalent to minimizing the sum in the exponential.

$$
\frac{d}{d\mu} \sum_i \frac{(x_i - \mu)^2}{\mu^2 x_i} = \frac{d}{d\mu} \sum_i \frac{((x_i/\mu) - 1)^2}{x_i} = -\sum_i \frac{2((x_i/\mu) - 1)}{x_i} \frac{x_i}{\mu^2}.
$$

Setting this equal to zero is equivalent to setting

$$\sum_i \left(\frac{x_i}{\mu} - 1\right) = 0,$$

and solving for μ yields $\hat{\mu}_n = \bar{x}$. Plugging in this $\hat{\mu}_n$ and maximizing with respect to λ amounts to maximizing an expression of the form $\lambda^{n/2}e^{-\lambda b}$. Simple calculus yields

$$\hat{\lambda}_n = \frac{n}{2b} \quad \text{where} \quad b = \sum_i \frac{(x_i - \bar{x})^2}{2\bar{x}^2 x_i}.$$

Finally,

$$2b = \sum_i \frac{x_i}{\bar{x}^2} - 2\sum_i \frac{1}{x_i} + \sum_i \frac{1}{x_i} = -\frac{n}{\bar{x}} + \sum_i \frac{1}{x_i} = \sum_i \left(\frac{1}{x_i} - \frac{1}{\bar{x}}\right).$$

8. Let $X_1, \ldots, X_n$ be a random sample from a population with pdf

$$f(x|\theta) = \frac{1}{2\theta}, \quad -\theta < x < \theta,\ \theta > 0.$$

Find, if one exists, a best unbiased estimator of θ.

解答: To find a best unbiased estimator of θ, first find a complete sufficient statistic. The joint pdf is

$$f(x|\theta) = \left(\frac{1}{2\theta}\right)^n \prod_i I_{(-\theta,\theta)}(x_i) = \left(\frac{1}{2\theta}\right)^n I_{[0,\theta)}(\max_i |x_i|).$$

By the Factorization theorem, $\max_i |X_i|$ is a sufficient statistic. To check that it is a complete sufficient statistic, let $Y = \max_i |X_i|$. Note that the pdf of Y is $f_Y(y) = ny^{n-1}/\theta^n$, $0 < y < \theta$. Suppose $g(y)$ is a function such that

$$E[g(Y)] = \int_0^\theta \frac{ny^{n-1}}{\theta^n} g(y)\,\mathrm{d}y = 0, \quad \text{for all } \theta.$$

Talking derivatives shows that $\theta^{n-1}g(\theta) = 0$, for all θ. So $g(\theta) = 0$, for all θ, and $Y = \max_i |X_i|$ is a complete sufficient statistic. Now

$$E[Y] = \int_0^\theta y\frac{ny^{n-1}}{\theta^n}\,\mathrm{d}y = \frac{n}{n+1}\theta \quad \Rightarrow \quad E\left[\frac{n+1}{n}Y\right] = \theta.$$

Therefore $\frac{n+1}{n}\max_i |X_i|$ is a best unbiased estimator for θ because it is a function of a complete sufficient statistic.

9. Show that for a random sample $X_1, \ldots, X_n$ from a $N(0, \sigma^2)$ population, the most

powerful test of $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma = \sigma_1$, is given by

$$\phi\left(\sum X_i^2\right) = \begin{cases} 1 & \text{if } \sum X_i^2 > c, \\ 0 & \text{if } \sum X_i^2 \le c. \end{cases}$$

For a given value of α, the size of the Type I Error, show how the value of c is explicitly determined.

解答: From the Neyman-Pearson lemma the UMP test rejects H_0 if

$$\frac{f(x|\sigma_1)}{f(x|\sigma_0)} = \frac{(2\pi\sigma_1^2)^{-n/2} e^{-\sum_i x_i^2/(2\sigma_1^2)}}{(2\pi\sigma_0^2)^{-n/2} e^{-\sum_i x_i^2/(2\sigma_0^2)}} = \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left\{\frac{1}{2}\sum_i x_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\right\} > k$$

for some $k \ge 0$. After some algebra, this is equivalent to rejecting if

$$\sum_i x_i^2 > \frac{2\log(k(\sigma_1/\sigma_0)^n)}{\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)} = c \quad \left(\text{because } \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} > 0\right).$$

This is the UMP test of size α, where $\alpha = P_{\sigma_0}(\sum_i X_i^2 > c)$. To determine c to obtain a specified α, use the fact that $\sum_i X_i^2/\sigma_0^2 \sim \chi_n^2$. Thus

$$\alpha = P_{\sigma_0}\left(\sum_i X_i^2/\sigma_0^2 > c/\sigma_0^2\right) = P(\chi_n^2 > c/\sigma_0^2),$$

so we must have $c/\sigma_0^2 = \chi^2_{n,\alpha}$, which means $c = \sigma_0^2\chi^2_{n,\alpha}$.

10. If $X_1, \ldots, X_n$ are iid from a location pdf $f(x - \theta)$, show that the confidence set

$$C(x_1, \ldots, x_n) = \{\theta : \bar{x} - k_1 \le \theta \le \bar{x} + k_2\},$$

where k_1 and k_2 are constants, has constant coverage probability.

解答:

$$\begin{aligned} P_\theta(\theta \in C(X_1, \ldots, X_n)) &= P_\theta(\bar{X} - k_1 \le \theta \le \bar{X} + k_2) \\ &= P_\theta(-k_2 \le \bar{X} - \theta \le k_1) \\ &= P_\theta\left(-k_2 \le \sum Z_i/n \le k_1\right), \end{aligned}$$

where $Z_i = X_i - \theta$, $i = 1, \ldots, n$. Since this is a location family, for any θ, $Z_1, \ldots, Z_n$ are iid with pdf $f(z)$, i. e., the Z_is are pivots. So the last probability does not depend on θ.